Spatial Audio Quality Inventory (SAQI).
Test Manual.

author: Alexander Lindau, audio communication Group, TU Berlin
date: 27/02/2014 14:27:14
release: 1.0

Contents

Introductory notes

This test manual has been created following guidelines in the German standard DIN 33430 [1].

The following text includes specific terminology. Please refer to the "Appendix C: Glossary of Terms" for further explanations/definitions.

1 Type of Test

The Spatial Audio Quality Inventory is intended for a qualitatively differentiated, comparative auditory assessment of real, imagined and simulated acoustic scenes.

2 Basic Test Concept

The perceptual evaluation of virtual acoustic environments may be based on overall criteria such as plausibility [2] and authenticity [3] or on differentiated perceptual qualities. However, only the latter will be suitable to reveal specific shortcomings of a simulation under test and allow for a directed technical improvement. To this end the Spatial Audio Quality Inventory (SAQI) was developed. Its purpose is to allow qualitatively differentiated

assessments of unimodal or supramodal auditory differences between technically generated acoustic environments (VAES) as well as with respect to a presented or imagined acoustic reality.

The SAQI comprises 48 verbal descriptors of perceptual qualities assumed to be of practical relevance when comparing virtual environments to real or imagined references or amongst each other. It was generated by a Focus Group of 20 German experts for virtual acoustics. Five additional experts helped verifying the unambiguity of all descriptors and the related explanations. Moreover, an English translation was generated and verified by seven bilingual experts.

The vocabulary in its entirety (including perceptual descriptors, circumscriptions, scale end label, and - if given - illustrative sound examples, see below) is intended to enable experts in the field to train any laymen to use it for assessments of VAEs.

Rationale and methodology pursued in constructing the SAQI vocabulary are described in more detail in [5], [6].

3 Test Materials

3.1 Provided Materials

The current document is available from the SAQI website [4], more specifically, there, a link to the TU Berlin's Research Data Repository *DepositOnce* is given, where all data is hosted.

Additional resources are provided in a zip-container. After unpacking, you should obtain the following folder structure:

```
\1 references
      References [3], [6], [7] as *.pdf
\2 audio files
      "comb filter like - 7 examples.mp3"
      "compressor effects - 2 examples.mp3"
      "roughness - 3 examples.mp3"
\3 mfiles
      \1 tools
      \2 data
      \3 examplePlots
      \4 csvExports
      "plot_saqi_results.m"
      "saqi2csv.m"
```

3.2 Description of Provided Materials

The test materials provided in these folders include this test manual (folder '1 manual') which contains for both English (cf. Appendix A: SAQI-EN) and German language (cf. Appendix B: SAQI-GER),

(a) a list of descriptors for auditory qualities,
(b) clarifying circumscriptions for each descriptor, and
(c) scale end label required for constructing rating scales of a semantic differential,
(d) a system of modifications of auditory qualities with respect to
    a. temporal variability and

b. interactivity, and

(e) a collection of assessment entities typical for the context of Virtual Acoustic Environments (VAEs) and

(f) a glossary of used specific terms (cf. Appendix C: Glossary of Terms).

Further, the materials comprise illustrative audio examples for selected auditory qualities in order to increase the understandability of the respective descriptor terms (cf. Appendix D: Sound Examples, (folder '2 audiofiles').

For the evaluation of SAQI test results two Matlab© scripts are provided (folder '3 mfiles'). The first one, saqi2csv.m, converts whisPER results files ('TSD.mat') into *.csv-files, that may conveniently be imported into a statistics software such as SPSS. The second one, plot_saqi_results.m, provides means for a fast visualization of individual and inter-individual results obtained from SAQI assessments. Usage instructions are given in the respective headers of these m-files.

If freely available, references cited in this text have been included in the folder '0 references' as *.pdfs.

3.3 Additional Materials

Further, a Matlab© software for listening tests - whisPER v1.8.0 - can be obtained from the website of the TU Berlin [7], which implements the complete SAQI as a semantic differential test. The included whisPER software's "User Documentation" may be consulted for detailed assistance.

Paper versions of the SAQI test may be constructed form the tables presented in the appendices of this document.

4 Test Organization

4.1 Defining Reference Stimuli

The SAQI is intended for assessing *auditory differences* related to some kind of 'test' or comparison or reference stimulus. This comparison stimulus might, however, exist either physically or only mentally. In the first case, the reference stimulus is typically denoted as an *external*, or *outer*, or *explicitly given reference*, whereas in the second case it is referred to as

an *imagined* or *memorized auditory representation,* being a sum of a subject's (a) prior experiences with, or (b) inferences towards respective stimuli (see [8] for a more detailed discussion).

Hence, at the beginning the researcher has to decide whether the comparison task should involve an explicitly given or an imagined reference stimulus. When using whisPER, in the first case some kind of acoustic stimulus has to be defined and all questionnaire items will be formulated as referring to the difference of test stimulus and reference stimulus, both being accessible for listening throughout the listening test. In the second case, subjects have to be instructed to imagine a suitable auditory reference, and consequently, in this case whisPER software will not present a comparison stimulus. Instead subjects will have to envision a certain cognitive representation from the auditory long-term memory, as, e.g., the impression of attending a (typical) classical concert or a conducting a (typical) narrowband phone call.

Of course, such an inner reference will vary across subjects. However, there may exist research questions which demand experience-based assessments. Whether the variability of this inner reference may be considered to be tolerable or not is another question. It will to a wide extend depend on the (concept-) representativeness of the selected sample of subjects (i.e. their degree of familiarity with the stimuli under test) and of the kind of instruction being given by the experimenter.

4.2 Qualitative Descriptors and Circumscriptions

The SAQI (cf. Appendix A: SAQI-EN, and Appendix B: SAQI-GER) comprises 48 qualitative descriptors sorted into 8 categories (timbre, tonalness, geometry, room, time behavior, dynamics, artifacts, and general impressions) which are to be considered as describing 'perceived differences with respect to [insert descriptor name]'.
Further, when using the whisPER software for SAQI tests, each qualitative descriptor will be accompanied by its written circumscription when presented to the subject. However, this will not render obsolete an adequate semantic training of test subjects (see section 5.4).

It has to be emphasized here again, that qualifier circumscriptions were not primarily intended as instructions for naïve test subjects, but for conveying the meaning of the perceptual qualifiers *to the expert user* (i.e., the experimenter, the researcher). Hence, although, by default whisPER presents circumscriptions together with quality names, circumscriptions may not contain ideal formulations for instructing/reminding test subjects. Instead, an experimenter might construct more intuitive formulations being easier under-

standable to a non-expert audience. In whisPER, circumscriptions may be edited by changing a singular Matlab file (refer to the WhisPER User Documentation and its section about adding new languages).

During the test, the presentation order of the qualitative descriptors may be randomized (e.g. from within whisPER). However, for economic reasons the descriptor 'Difference' should always be assessed in the first place (automatically done with whisPER, SAQI test is stopped when no overall difference is perceived). Additionally, being intended for handling possibly overlooked or newly emerging aspects of VAEs, the descriptor 'Other' included in the vocabulary should always be rated as last item (automatically done with whisPER, in case, subjects will be asked to enter a suitable name for their perception and then being presented a rating scale). Researchers are cordially invited to share their experience with this item with the author (alexander.lindau@tu-berlin.de).

4.3 Rating Scales

Each descriptor is completed by scale end label required for constructing rating scales of a semantic differential. WhisPER provides such rating scales for each auditive quality. Scale types may vary between

(a) bipolar
(b) unipolar, or
(c) dichotomous

ones, depending on the respective quality. For the majority of SAQI items the usage of bipolar continuous scales is preferable, allowing indicating amount AND qualitative direction of a perceived difference (e.g., perceived difference in *pitch,* scale ends: higher – lower). However, there are also some qualities that may be perceived to vary only with respect to their amount, thus requiring unipolar scales (e.g., perceived difference with respect to *distortion*, scale ends: less intense – more intense). A singular quality has been defined to be dichotomous: perceived difference with respect to *front-back position*, scale: [not confused], [confused].

All scale labels have been constructed in order to semantically express an increase of the quality under test when reading the scale from the left hand to the right hand label (see sections 11 and 12). In whisPER, rating scales are oriented vertically. Hence, in order to allow for intuitive perceptual rating, the right hand label (encoding an increase) is always displayed on

top of the scale. See section 9 for further details on the kind of encoding of rating results used by whisPER for saving results.

4.4 Assessment Entities

Perceptual assessments may further be addressed to a selection of reference objects typical for VAEs. Hence, five basic assessment entities were defined, providing an ideal-type ontology of the presented scene such as:

(a) foreground sources,
(b) background sources,
(c) the simulated room acoustical environment,
(d) the reproduction system (e.g. loudspeaker artifacts, amplifier noise), and
(e) the laboratory environment (HVAC noise, environmental sounds).

In combination, these five entities are thought to incorporate all possible objects of interest.

However, as the need for defining such assessment entities may depend on the actual research question, the whisPER program provides means for choosing/re-defining/omitting the reference objects. Subjects may be asked to indicate the suitable assessment entity/entities using multiple-choice radio buttons. Further, for completeness and when using whisPER, two more answering categories “don’t know” and “other” will automatically be presented.

4.5 Further Modifications of Perceptual Qualities

Finally, and typical for VAEs, perceptual qualities may be further differentiated with respect to time-variance or their behavior with respect to interaction. Thus, perceived differences might be either constant or time-varying. The time-variance might be periodically or otherwise rule-based or non-regular and it might be continuous or discontinuous. Additionally, perceived differences may depend either on user interaction, on scene events (referring to the actual audio content, too) or on none of them (i.e., be independent). Again, the whisPER package provides means for extending the SAQI assessments in this respect. Subject may be asked to indicate the perceived kind of modification (if any) via a hierarchical selection using radio buttons.

5 Test Execution

5.1 Physical Requirements of Test Subjects

There is no limitation of applying the SAQI with respect to the age of subjects. It might be advisable to screen subjects for visual or auditory impairments and for adequate reading

comprehension. Additionally, when presenting audio stimuli, care should be taken to not exceed tolerable sound pressure levels.

5.2 Test Administration

The SAQI might be administered to singular individuals as well as to groups of subjects. Additionally, the SAQI is suitable for either Computer Assisted Personal (CAPI) or Paper and Pencil (PAPI) interviewing techniques. In the generic and complete case a SAQI test should be executed as follows (as implemented, e.g., in the whisPER software):

(1) At first, the actual detectability of any (global) auditive difference will be asked for, and, if there is one, its intensity will be rated (if there is none, the SAQI test should be stopped here).
(2) Then, all qualitative descriptors and their accompanying written circumscriptions will be presented (potentially in random order) and the perceived amount (if any) will be rated using 2-7 step rating scales.
(3) If an auditory difference quality was perceived (i.e. rated), subjects will be asked to further differentiate their perception with respect to time variance and interactivity.
(4) Concurrently, subjects will be asked to assign their perception to a certain assessment entity (i.e. a scene element or similar).
(5) After all qualities have been presented, subjects will be asked whether they perceived any remaining differences not being included in the test so far. In case, they may enter a suitable name and rate the perceived difference using a provided intensity (i.e. unipolar) scale.

The above list describes the complete (i.e. maximum) extend of a SAQI test. However, when using the whisPER software, the test may conveniently be customized (i.e. mostly reduced) in many respects (see next section).

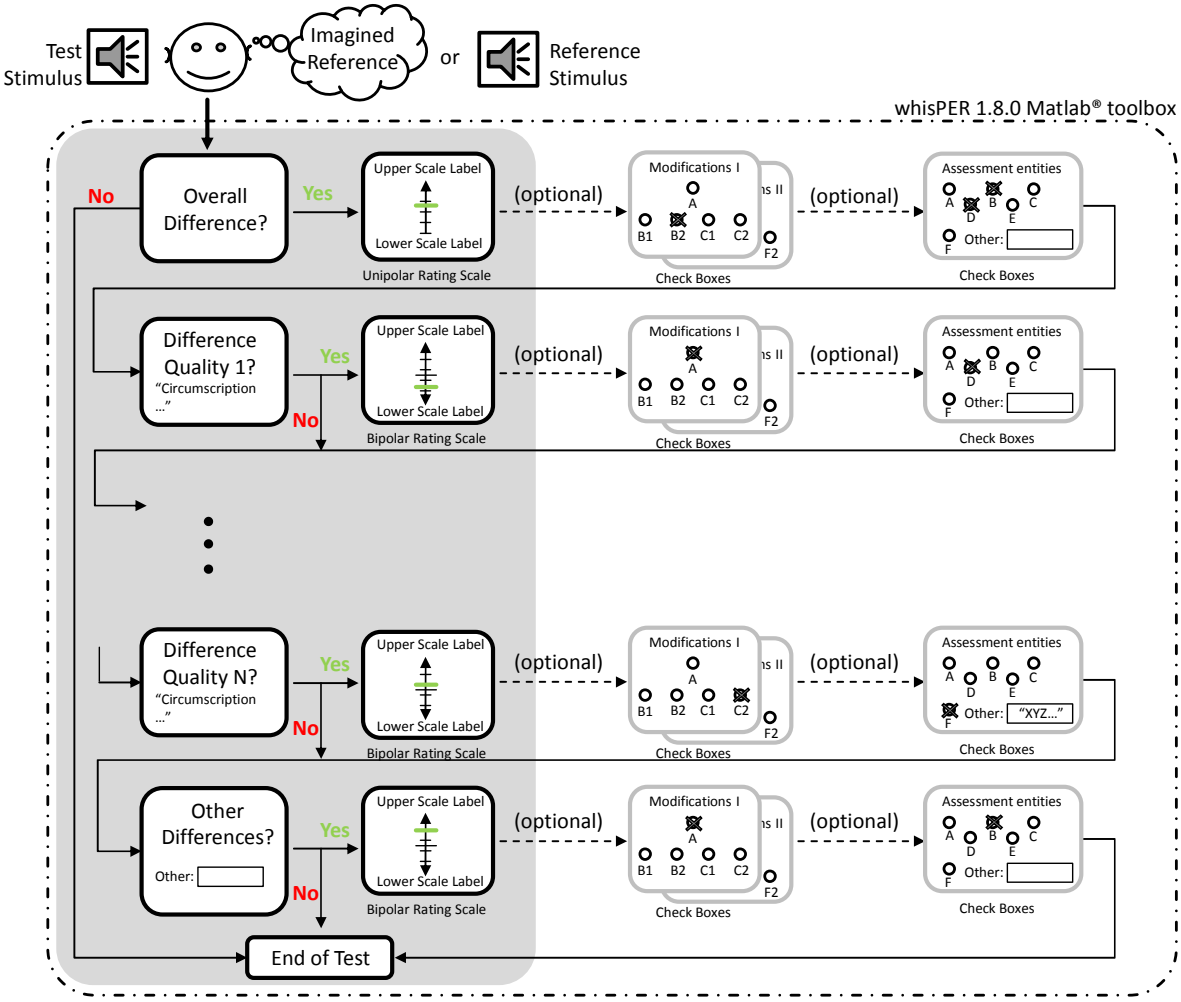

Figure 5-1. Illustration of a how the (complete) SAQI may be administered.

5.3 Customizing SAQI Tests

From the preceding sections it should have become clear that the SAQI is a potentially rather extensive test instrument. However, and, e.g., for studies pursuing a more confirmatory than exploratory approach one may conveniently customize/reduce it to specific needs. Suitable approaches to such a reduction will be discussed in the following.

First, and depending on the individual research question, it might be that not all qualities from the SAQI catalogue need to be assessed. Hence, singular perceptual qualities may be omitted. It might also be that one wants to omit a complete category of perceptions (e.g., the complete geometry section). *Vice versa*, it might also be that one is interested only in qualities from a singular category (e.g., the artifacts section).

Additionally, it might appear helpful to integrate certain perceptual qualities into larger questionnaire items, i.e. by logical conjunction or nondisjunction (e.g., "Please rate per-

ceived differences in width OR height OR depth, i.e., rate any difference in volumetric extent", or "Please rate perceived differences with respect to noise-like, AND pitched, AND impulsive artifacts"). The whisPER program provides means for convenient pre-selecting qualities on an individual or category base. In case you want to alter the predefined questionnaire items, please contact the current maintainers of the whisPER project.

Further, it might be that the differentiation of perceived auditory qualities with respect to time and interactivity is not needed, or is needed in certain respect only. Before discussing ways for reducing, Table 5-1 presents a hierarchically ordered overview of *all* modifications that may be expressed additionally per perceptual quality as provided with the SAQI.

Table 5-1 Potential additional modifications of perceived auditory qualities, hierarchically organized with respect to temporal variability, and interactivity (selection referring to checking whisPER options 1, 2, and 3).

<table>
<tr><td colspan="3">The perceived difference is …</td></tr>
<tr><td rowspan="2">… constant</td><td>… varying periodically or otherwise rule-based with time</td><td>… varying non-regularly with time</td></tr>
<tr><td colspan="2">… in a continuous / discontinuous manner</td></tr>
<tr><td colspan="3">… and depending on scene events / user interaction / independent.</td></tr>
</table>

When using the whisPER program, assessable modifications may be reduced in different ways. The following tables (Table 5-2 to Table 5-5) illustrate the reduction options which may be chosen.

Table 5-2 Proposal for a reduced assessment of modifications (presented selection refers to checking whisPER option 1).

<table>
<tr><td colspan="3">The perceived difference is …</td></tr>
<tr><td>… constant.</td><td>… varying periodically or otherwise rule-based with time.</td><td>… varying non-regularly with time.</td></tr>
</table>

Table 5-3 Proposal for a reduced assessment of modifications (presented selection refers to checking whisPER options 1 and 2).

<table>
<tr><td colspan="3">The perceived difference is …</td></tr>
<tr><td rowspan="2">… constant.</td><td>… varying periodically or otherwise rule-based with time</td><td>… varying non-regularly with time</td></tr>
<tr><td colspan="2">… in a continuous / discontinuous manner.</td></tr>
</table>

Table 5-4 Proposal for a reduced assessment of modifications (presented selection refers to checking whisPER option 3).

<table>
<tr><td>The perceived difference is …</td></tr>
<tr><td>… depending on scene events / user interaction / independent.</td></tr>
</table>

Table 5-5 Proposal for a reduced assessment of modifications (presented selection refers to checking whisPER options 1 and 3).

<table>
<tr><td colspan="3">The perceived difference is …</td></tr>
<tr><td>… constant</td><td>… varying periodically or otherwise rule-based with time</td><td>… varying non-regularly with time</td></tr>
</table>

… and depending on scene events / user interaction / independent.

When resigning to ask for any further modification of perceptual qualities (in whisPER: unchecking options 1, 2 and 3) subjects should be instructed to rate their sensations of auditory qualities in a suitably integrated manner (see section 5.4).

Further, and often depending on the presented content, it might be advisable to customize the types of scenic or environmental assessment entities (or reference objects) a subject may ascribe its perceived difference perception to. When using whisPER, the following options may be chosen from:

- Do not assign perceived differences to specific entities,
- assign perceived differences to one or more from up to five pre-proposed or self-defined assessment entities, and two additional answering categories “don’t know” and “other” (the latter being automatically presented when using whisPER).

Finally and as already mentioned in section 4.2, circumscriptions may be edited to be more easily understandable to an audience of non-expert users. Refer to section 4.2 for details and implementation.

5.4 Test Subject Training

The SAQI is not intended to be used with naïve, untrained test subjects. Instead, it was designed to be – in its entirety – self-explanatory to a specific expert user group. Typically, this targeted expert user will be a developer and/or researcher engaged with the development and evaluation of virtual acoustic environments. It is further assumed, that – by his in-depth knowledge – the expert will be capable of training naïve subjects to a valid and reliable use of the SAQI. Thereby, it should be at the discretion of the expert of how to achieve this ability in his test subjects. However, in the following, examples are outlined of how such training might be conducted.

Basically, the experimenter has to ensure that all auditory qualities are clearly understood by the test subjects. As a first step, subjects could be invited to a personal interview.

At first it should be emphasized that all assessments to be made relate to *auditive differences relative to some sort of reference* (inner or outer). Whereas in the latter case further explanations simplify towards explaining the proper use of the scale in order to correctly encode the perceived amount and direction of a perceptual difference (in whisPER scales are accordingly labelled: “Stimulus A is more/less … than stimulus B’), in the former case the

inner reference has to be suitably evoked. In referring to [2] the following phrase may be proposed:

"Please, compare the auditory impression of the presented stimulus to your expectations of a [corresponding (real) event]. For each of the following auditive qualities rate the amount and direction of a perceived deviation (if any)."

As 'corresponding (real) event' a suitable experience may be referred to, e.g., visiting a symphonic concert, going to a movie theater, listening to a lecture, etc.
Afterwards, subjects would be presented with the descriptor terms and asked to explain the meaning of each of them in their own words. As we think that perceptions as psychological states may not be defined *per se*, subjects may be advised to give respective:

(a) further explanations, or paraphrases
(b) synonyms,
(c) examples from daily live, where the described percept may typically be encountered,
(d) onomatopoetic transcriptions (e.g. 'Noise-like artifact: It sounds like sssssh…') or
(e) refer to typical physical causes.

When suspecting a misunderstanding the experimenter should inquire accordingly. To this end he might make use of the provided circumscriptions, scale label and audio examples. Training should be finished only if the experimenter himself is convinced that the subject has a clear understanding of the perceptual qualities.

As mentioned before, for three perceptual qualities illustrative audio examples have been provided. This was done as in these cases it was felt that both descriptor terms and written circumscriptions will not suffice to – even between experts – clearly convey the targeted sensation. For all other qualities though, it is assumed, that while being an audio expert the experimenter himself may also be able to construct suitable sound examples, this way enhancing the understandability of potentially problematic sensation to naïve subjects. In the future it might be that a collection of audio examples will be made available illustrating not only three auditory qualities but the complete sensory catalogue covered by the SAQI.

Subsequent (or concurrently) to explaining perceptual qualities the usage of scales and the meanings of scale label should be explained. Mostly, scales should be rather self-explanatory (standard uni- and bipolar scales) and presentation of some selected examples should suffice. However, additional explanations could be needed for scales of horizontal and vertical

direction, as here, difference ratings have to be given directly in degrees, and for the singular dichotomous scale for 'perceived difference in front back position'.

In the next step – and, if being used in the SAQI test – subjects should be made known to the meaning and application of time-variant and interactivity-related modifications to qualities. Also in this case, subjects should demonstrate their proper understanding, e.g., by giving explanations in their own words or by illustrating suitable examples. As mentioned in section 5.3 it might appear useful to refrain from asking for further differentiation of perceived qualities and, instead, to instruct subjects to give their difference ratings somehow integrated for all sub-aspects of time-variability and interactivity. Whereas this might be difficult to explain to subjects exemplification may be helpful. Hence, one could explain, that difference ratings, e.g., for loudness should then cover continuous loudness variations as well as sudden loudness jumps, or loudness variations observed with user interactions as well as such occurring only with certain scene events.

Similar advices refer to assessments entities being assignable to perceived qualities. They should be made known to the subjects and – most importantly – be clearly distinguished from each other. Moreover, it should be explained that a perceived quality may be assigned to multiple entities at once.

In any case, the complete questionnaire should be known to the subjects in advance, as knowledge of other qualities is assumed mandatory for enabling differentiated and reliable judgments.

To make subjects familiar with the usage of the questionnaire (especially in case of the computer-aided versions, as the one provided with whisPER) test runs with illustrative example stimuli could be conducted.

In order to save training time the installation of a permanent sensory panel used to the SAQI and its administration might be helpful.

6 Test Languages

So far, the SAQI is available in German and English language (see Appendix B: SAQI-GER and Appendix A: SAQI-EN). Both versions have also been implemented in whisPER software. The German and English versions are assumed to be semantically compatible (see [6] for ra-

tionale and method). However, empirical assessments of inter-language compatibility are planned for the future.

Of course, the SAQI may be translated into other languages. If you plan such a translation, we propose using the English version as an initial point and translating the SAQI in panel discussions of bilingual experts. The English version was especially provided to facilitate future translations: As we assume the typical expert-in-the-field to be more or less 'bilingual' in at least a 'scientific community English' (see also [6]), suitable panels of specialists may conveniently be formed on a national level.

Intents towards creating a French version have been reported to the author already. All scientists in the field are cordially invited to contribute additional national versions. The author will be glad in providing methodological counselling.

Future translations may also easily be integrated into the whisPER software. Please, contact the maintainers of the whisPER project in case you plan a translation.

7 Test Quality Criteria

7.1 Objectivity

Objectivity when conducting SAQI test will be increased by obeying known rules of good scientific practice (see, e.g., [10]). Confounding influences should be controlled by randomization or standardization. Test preparation, instructions, procedures and approaches to evaluation of results should be standardized and well documented (e.g., use written instructions and interview guidelines, use trained interviewers/test operators, use double blind and computer aided test administration, document statistical sampling, analysis procedures, and raw test data). Special care has to be taken in ensuring clear understanding of SAQI items in test subjects.

7.2 Reliability

Classical Test Theory (CTT) assumes a questionnaire to be an integrative instrument measuring a one or more latent (not directly observable) psychological variable or constructs. To increase reliability of the measurement a high number of semantically related questionnaire items are used. A reliability analysis is then applied in order to assess in how far all items are actually measuring the same latent construct. The SAQI comprises items for both basic auditory perceptions (as, e.g., loudness, direction, and spectral coloration) and higher constructs

(naturalness, clarity, degree-of-liking, presence) which in sum might thought to contribute a (rather simple) 'construct' as, e.g., "perceived overall difference".

Only little empirical data have been collected with the SAQI, so far. Hence, statements with respect to reliability are still informal. In [8], nine subjects compared two different sound field simulations (individual and non-individual dynamic binaural synthesis) to acoustic reality (using 45 out of the 48 SAQI qualifiers). For standardized ratings Cronbach's α was found to be 0.564 (individual simulation), or 0.55 (non-individual simulation), respectively. If Cronbach's α was calculated for standardized *absolute* ratings values increase to 0.876, or 0.793, respectively. Hereby, it might depend on the targeted conclusions (assessment of *systematic* deviations, assessment of *any* perceptible deviations) whether it is advisable to assess raw or absolute SAQI ratings.

7.3 Validity

Content validity of the SAQI questionnaire is thought to be ensured to a wide extend by the expert-based approach that was chosen for its development (see [6]).

Regarding construct validity it can be stated that it is at least assumed, that the SAQI will be able to fulfill its major goals: revealing differences between VAEs and both in respect to their overall performance and in a qualitatively differentiated manner (for empirical evidence see [9]). Hereby, performance is understood as the degree of perceptual accuracy, which can also be thought of as the perceived degree of agreement with a given reference. In the future it might possible to interpret a SAQI overall score as a rating of simulation accuracy. However, until then it remains to be decided how individual ratings can suitably be aggregated (see also section 7.2) and what the obtained overall score will stand for.

With respect to criterion validity no external criteria have been defined so far to be predicted by SAQI scores.

For test dimensionality see section 7.4.

7.4 Dimensionality

So far, no assessment of latent factor structures underlying the SAQI has been conducted. However, appropriate data are currently being collected and factor analytic results will be presented in the future. However, in accordance to the creation process of the SAQI (see [6])

is assumed that – while being semantically not perfectly orthogonal in any case – each quality descriptor is of practical relevance.

7.5 Standardization

So far, no standardization of either the SAQI overall score or with respect to scorings on individual perceptual qualities is planned. However, inferences on the comparative performance of an assessed system might be drawn from the results of a planned Round Robin on Auralisation once available. This Round Robin will involve SAQI-tests of a number of different state-of-the-art implementations for Virtual Acoustic Environments. Further, SAQI-based assessment results for state-of-the-art implementations of (a) non-individual, and (b) individual data-based dynamic binaural synthesis have been presented in [9].

8 The WhisPER Matlab Toolbox v1.8.0

As mentioned before, the complete SAQI test has been implemented in German and English language in the whisPER Matlab® toolbox for listening tests v1.8.0. The toolbox maybe downloaded from http://www.ak.tu-berlin.de/whisper. Usage instructions can be found in the provided User's Manual of whisPER. The Users' Manual also gives details on how to integrate a new language into whisPER.

9 Evaluating SAQI Test Results from WhisPER

The whisPER User's Manual (see section 8) gives details on the format in which SAQI test results are being saved (order of test subjects and rated qualities, assigned items, entities, range and values of raw ratings). However, in order to allow for an intuitive understanding of test results some information on the data format is repeated here, too.

When saving rating results, these are always encoded as if being assigned to the stimulus under test when compared to the reference stimulus (the latter being an inner or an outer reference). This direction of encoding is also retained in saved results if test and references stimuli are chosen to vary randomly for each assessed perceptual quality (whisPER option). Further, ratings are encoded to reflect perceived differences according to a logical increase of the quality under test (in direction of left hand to right hand scale labels, see sect. 4.3). Thus, raw SAQI rating as obtained from whisPER may intuitively be interpreted: positive difference ratings in terms of perceived high frequency coloration, sharpness, distance, or clarity etc., refer to a perception of increased distance, emphasized high frequencies and a

sharper, more distant and more clear sound of the test stimulus as compared to the reference.

As mentioned already in section 3 there are two Matlab® functions provided to allow a fast evaluation of obtained SAQI ratings.

The first one, `saqi2csv.m`, converts whisPER results files ('TSD.mat') into *.csv-files, that may conveniently be imported into a statistics software such as SPSS. The script only exports the raw ratings into the *.csv-file. In the *.csv data is organized displaying subjects/cases as rows and qualities as columns, and the first column containing the subject ID (see Figure 9-1). As usage is simple and instructions are given in the respective file further explanations are not assumed necessary.

	A	B	C	D	E
1	subject	Unterschied	Klangfarbe hell-dunkel	Klangfarbliche Ausprägung im Höhenbereich	Klangfarbliche Ausprägung im Mittenbereich
2	2	0,42466	0,42282	0,27517	-0,26175
3	3	0,80822	0,3557	0,47651	0,28859
4	4	0,64384	0	0,12752	0
5	9	0,87671	0,50336	0,65101	0
6	12	0,79452	-0,77181	-0,62416	-0,27517
7	1	0,50685	0,4094	0,16779	0
8	6	0,72603	0,16779	-0,43624	0
9	8	0,89041	0,19463	0,12752	0
10	5	1	-1	-0,69128	-0,4

Figure 9-1 Screenshot from an exemplary *.csv-file produced with the script `saqi2csv.m`

The second Matlab script, `plot_saqi_results.m`, provides means for a fast visualization of individual and inter-individual results obtained from SAQI assessments. Detailed usage instructions are given in the respective header of the m-file. Exemplary plots produced by the script are shown in the following (Figure 9-2 to Figure 9-5). Plots may be customized to a wide extend (arrangement of subplots, figure and font sizes, choice of language, choice of between given or user defined plot labels, etc.). Note that, up to now, group results are visualized using means and confidence intervals. One should, however, it should check whether data support using these parameters for spread and central tendency. In the future, non-parametric distribution plots (boxplots) will be implemented, too.

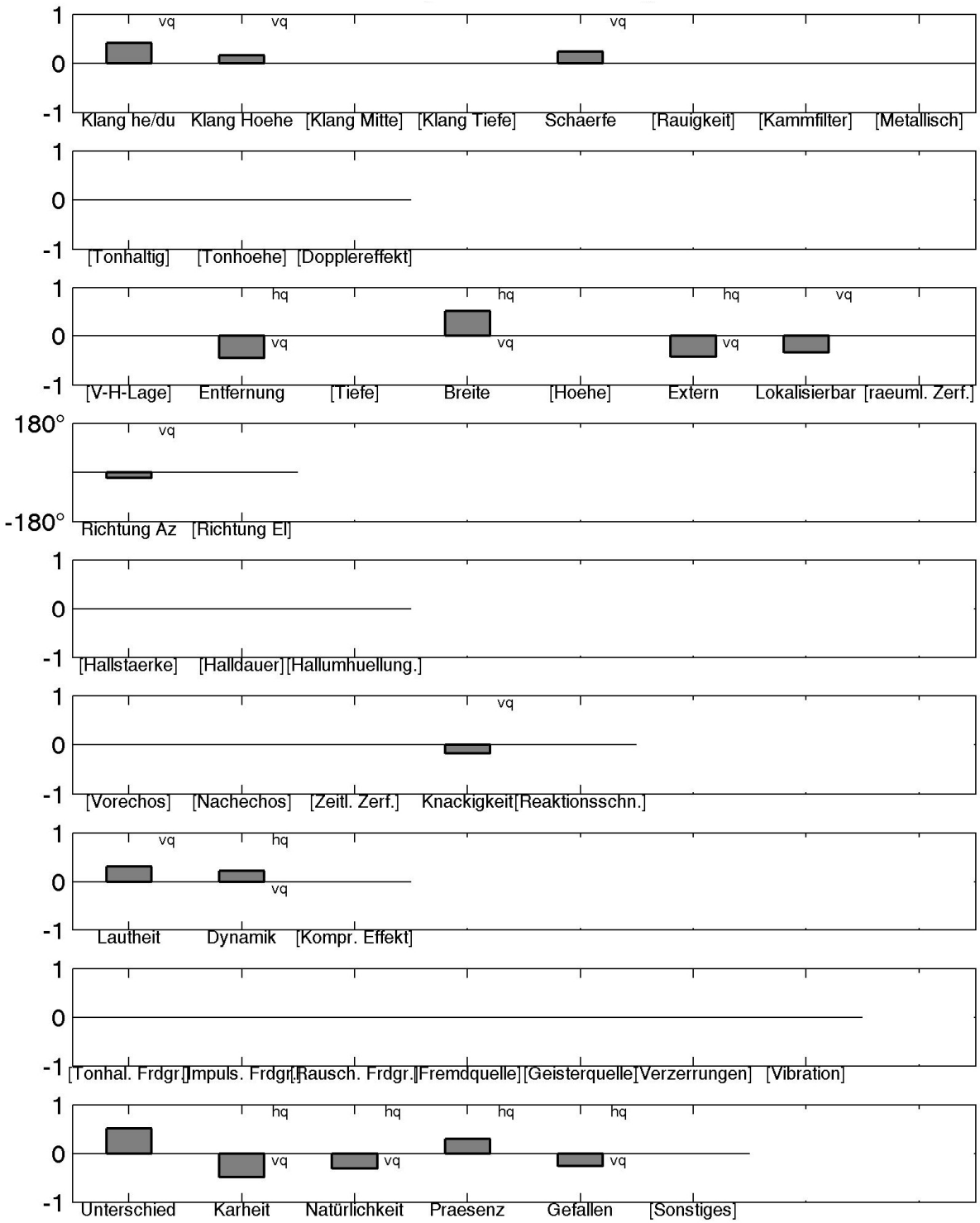

Figure 9-2 Plot of an individual's ratings for a complete (German) SAQI test. Qualities are ordered according to overall categories. Perceptual qualities that have not been rated (i.e., which were not perceived as different) are displayed with brackets around their names. Small letters at the right of the bars are short labels indicating additionally assigned assessment entities. If they would have been assessed (not the case in this example), short labels for modifications might have been displayed, too (left to the bars).

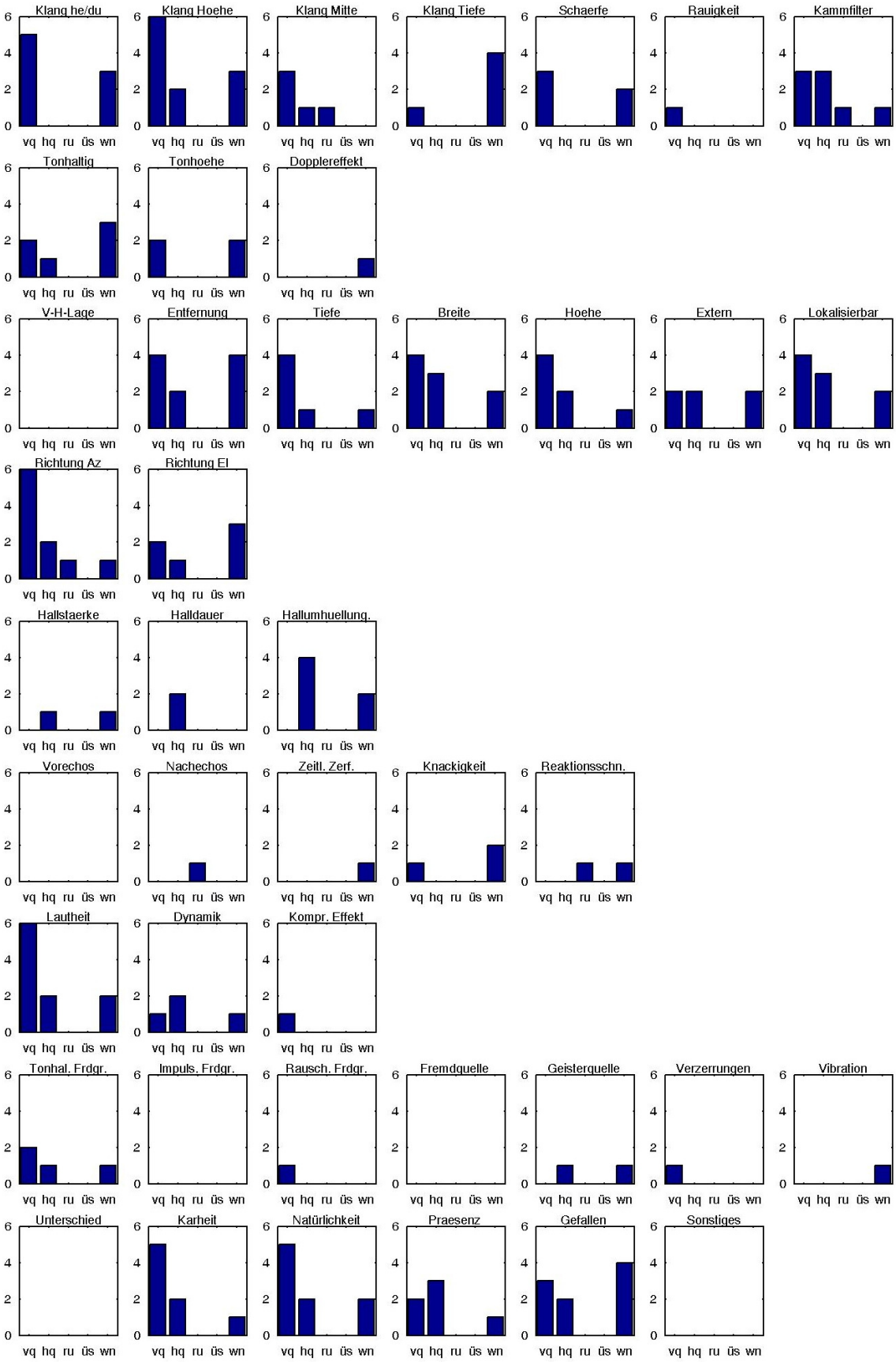

Figure 9-3 Plot of histograms of assessment entities as assigned to individual categories by a group of subjects which conducted the same SAQI test. Qualities are ordered according to overall categories. Similar plots may be produced for as-

signed temporal and interactivity-related modifications of perceptual qualities. For plotting reasons entities are encoded with short labels.

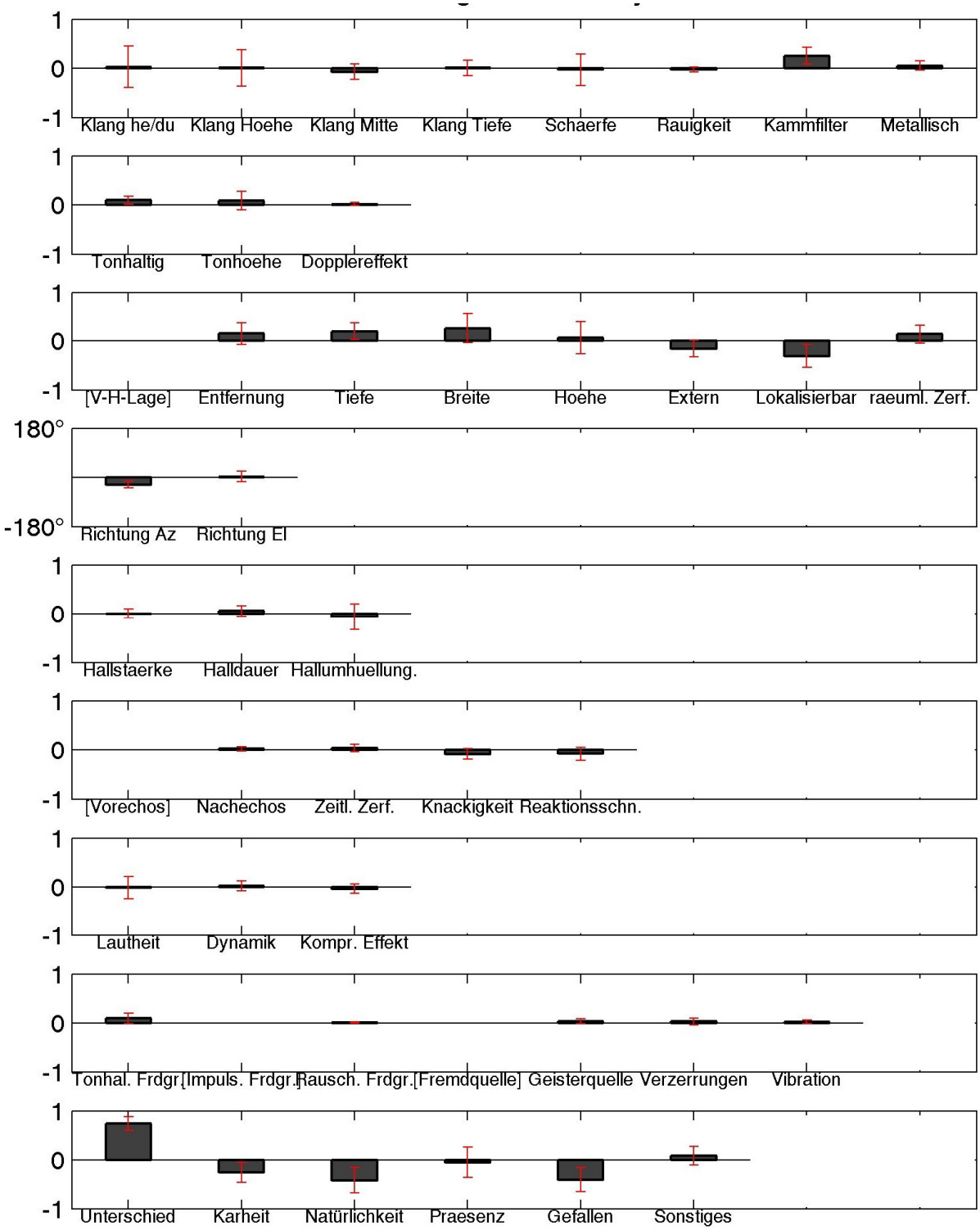

Figure 9-4 Pot of average ratings with 95% confidence intervals from a group of subjects which conducted the same SAQI test. Qualities are ordered according to overall categories. Perceptual qualities that have not been rated by any subject (i.e., which were not perceived as different) are displayed with brackets around their names.

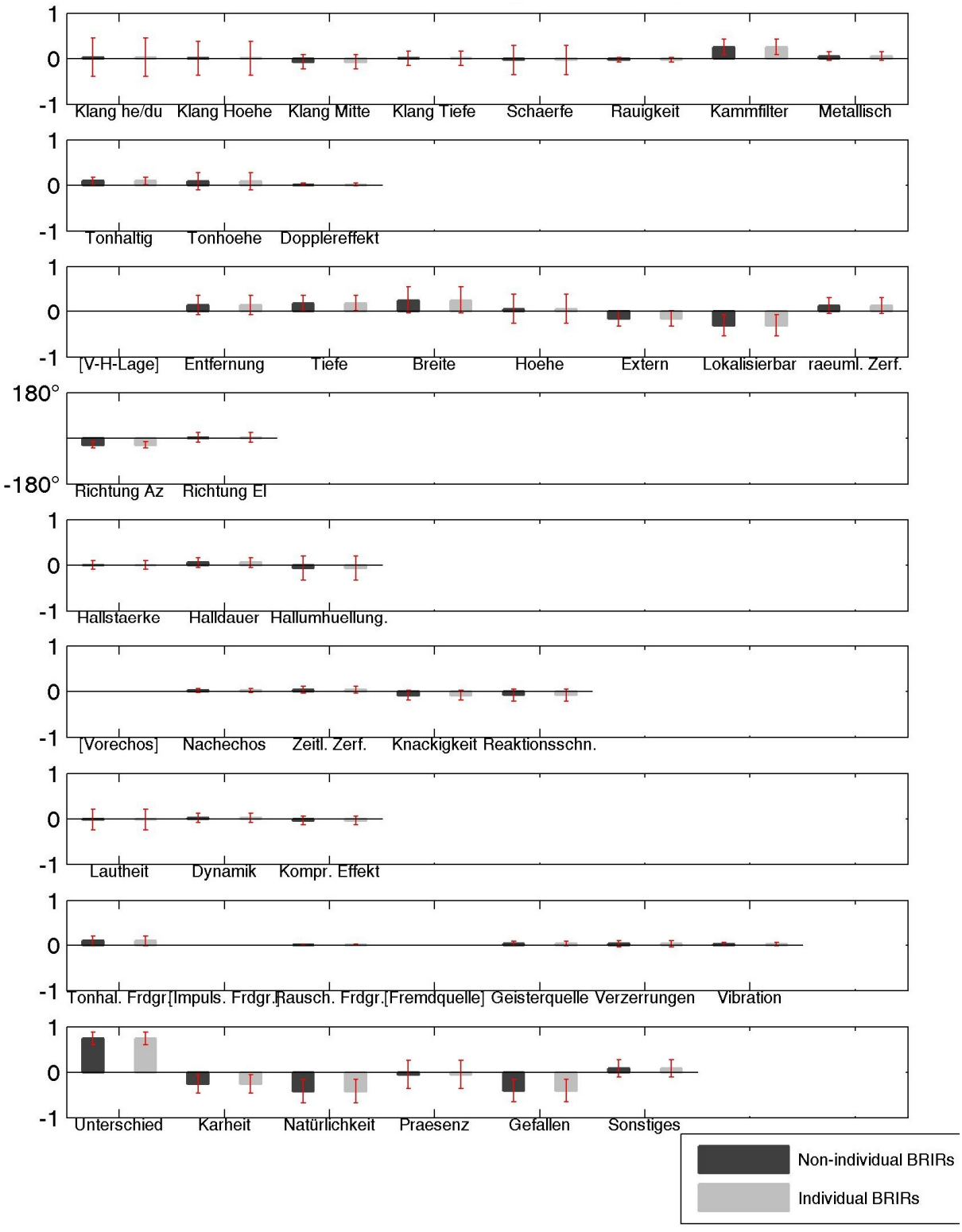

Figure 9-5 Plot displaying a comparison of ratings from a group of subjects which conducted the same SAQI test under two conditions (mean ratings with 95% confidence intervals). Qualities are ordered according to overall categories. Perceptual qualities that have not been rated by any subject (i.e., which were not perceived as different) are displayed with brackets around their names.

10 References

[1] DIN 33430: *Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen*. Berlin: Beuth

[2] Lindau, Alexander; Weinzierl, Stefan (2012): "Assessing the Plausibility of Virtual Acoustic Environments", in: *Acta Acustica united with Acustica*, Vol. 98, No. 5, pp. 804-810

[3] Brinkmann, Fabian; Lindau, Alexander; Vrhovnik, Martina; Weinzierl, Stefan (2014): "Assessing the Authenticity of Individual Dynamic Binaural Synthesis", in: *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*. Berlin

[4] http://www.ak.tu-berlin.de/saqi/

[5] Lindau, Alexander; Erbes, Vera; Lepa, Steffen; Maempel, Hans-Joachim; Brinkmann, Fabian; Weinzierl, Stefan (2014a): "Eine Fokusgruppe für die Entwicklung eines Vokabulars zur sensorischen Beurteilung virtueller akustischer Umgebungen", in: *Fortschritte der Akustik: Tagungsband d. 40. DAGA.* Oldenburg

[6] Lindau, Alexander; Erbes, Vera; Lepa, Steffen; Maempel, Hans-Joachim; Brinkmann, Fabian; Weinzierl, Stefan (2014b): "A Spatial Audio Quality Inventory for Virtual Acoustic Environments (SAQI)", in: *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*. Berlin

[7] Ciba, Simon; Wlodarski, André; Maempel, Hans-Joachim (2009): "WhisPER – A new tool for performing listening tests", in: *Proc. of the 126th AES Convention*. Munich, preprint 7749, http://www.ak.tu-berlin.de/whisper/

[8] Kuhn-Rahloff, Clemens (2011): *Prozesse der Plausibilitätsbeurteilung am Beispiel ausgewählter elektroakustischer Wiedergabesituationen. Ein Beitrag zum Verständnis der „inneren Referenz" perzeptiver Messungen*. Technische Universität Berlin

[9] Lindau, Alexander; Brinkmann, Fabian; Weinzierl, Stefan (2014): "Qualitative and Quantitative Deviations of Individual and Non-individual Dynamic Binaural Synthesis from Acoustic Reality", in: Proc. of the EAA Forum Acusticum 2014. Krakow (PL)

[10] Bortz, Jürgen; Döring, Nicola (2006): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4th ed. Heidelberg: Springer

11 Appendix A: SAQI-EN

Table 11-1: Spatial Audio Quality Inventory (SAQI) - English version

	perceptual quality	**circumscription**	**scale end label**
	Difference	Existence of a noticeable difference.	none – very large
Timbre	Tone color bright-dark	Timbral impression which is determined by the ratio of high to low frequency components.	darker – brighter
	High-frequency tone color	Timbral change in a limited frequency range.	attenuated – emphasized
	Mid-frequency tone color	Timbral change in a limited frequency range.	attenuated – emphasized
	Low-frequency tone color	Timbral change in a limited frequency range.	attenuated – emphasized
	Sharpness	Timbral impression which e.g., is indicative for the force with which a sound source is excited. Example: Hard/soft beating of percussion instruments, hard/soft plucking of string instruments (class. guitar, harp). Emphasized high frequencies may promote a 'sharp' sound impression.	less sharp – sharper
	Roughness*	Timbral impression of fierce or aggressive modulation/vibration, whereas individual oscillations are hardly distinguishable. Often rated as unpleasant.	less rough – more rough
	Comb filter coloration*	Often perceived as tonal coloration. 'Hollow' sound. Example: speaking through a tube.	less pronounced – more pronounced
	Metallic tone color	Coloration with pronounced narrow-band resonances, often as a result of low density of natural frequencies. Often when exciting metallic objects such as Gongs, bells, rattling tin cans audible. Applicable to room simulations, plate reverb, spring reverb, too.	less pronounced – more pronounced
Tonalness	Tonalness	Perceptibility of a pitch in a sound. Example for tonal sounds: voiced speech, beeps.	more unpitched – more pitched
	Pitch	The perception of pitch allows arranging tonal signals along a scale "higher - lower".	lower – higher
	Doppler effect	Continuous change of pitch (see above). Often perceived as a 'continuous detuning'. Example: 'Detuned' sound of the siren of a fast-moving ambulance.	less pronounced – more pronounced
Geometry	Horizontal direction	Direction of a sound source in the horizontal plane.	shifted anticlockwise - shifted clockwise (up to 180°)
	Vertical direction	Direction of a sound source in the vertical plane.	shifted up – shifted down (up to 180°)
	Front-back position	Refers to the position of a sound source before or behind the listener only. Impression of a position difference of a sound source caused by 'reflecting' its position on the frontal plane going through the listener.	dichotomous scale: not confused / confused
	Distance	Perceived distance of a sound source.	closer – more distant
	Depth	Perceived extent of a sound source in radial direction.	less deep – deeper
	Width	Perceived extent of a sound source in horizontal direction.	less wide – wider

	perceptual quality	circumscription	scale end label
	Height	Perceived extent of a sound source in vertical direction.	less high - higher
	Externalization	Describes the distinctness with which a sound source is perceived within or outside the head regardless of their distance. Terminologically often enclosed between the phenomena of in-head localization and out-of-head localization. Examples: Poorly/not externalized = perceived position of sound sources at diotic sound presentation via headphones, good/strongly externalized = perceived position of a natural source in reverberant environment and when allowing for movements of the listener.	more internalized – more externalized
	Localizability	If localizability is low, spatial extent and location of a sound source are difficult to estimate, or appear diffuse, resp. If localizability is high, a sound source is clearly delimited. Low/high localizability is often associated with high/low perceived extent of a sound source. Examples: sound sources in highly diffuse sound field are poorly localizable.	more difficult – easier
	Spatial disintegration	Sound sources, which - by experience - should have a united spatial shape, appear spatially separated. Possible cause: Parts of the sound source have been synthesized/simulated using separated algorithms/simulation methods and between those exists an unwanted offset in spatial parameters. Examples: fingering noise and playing tones of an instrument appear at different positions; spirant and voiced phonemes of speech are synthesized separately and then reproduced with an unwanted spatial separation.	more coherent – more disjointed
Room	Reverberation level	Perception of a strong reverberant sound field, caused by a high ratio of reflected to direct sound energy. Leads to the impression of high diffusivity in case of stationary excitation (in the sense of a low D/R-ratio). Example: The perceived intensity of reverberation differs significantly between rather small and very large spaces, such as living rooms and churches.	less – more
	Reverberation time	Duration of the reverberant decay. Well audible at the end of signals.	shorter – longer
	Envelopment (by reverberation)	Sensation of being spatially surrounded by the reverberation. With more pronounced envelopment of reverberation, it is increasingly difficult to assign a specific position, a limited extension or a preferred direction to the reverberation. Impressions of either low or high reverberation envelopment arise with either diotic or dichotic (i.e., uncorrelated) presentation of reverberant audio material.	less pronounced – more pronounced
Time behavior	Pre-echoes	Copies of a sound with mostly lower loudness prior to the actually intended the starting point of a sound.	less intense – more intense
	Post-echoes	Copies of a sound with mostly decreasing loudness after the actually intended the starting point of a sound. Example: repetition of one's own voice through reflection on mountain walls.	less intense – more intense
	Temporal disintegration	Sound sources, which - by experience - should have a united temporal shape, appear temporally separated. Causes similar to "Spatial disintegration", however, here: due to timing-offsets in synthesis. Example: fingering noise and playing tones of an instrument appear at different points in time.	more coherent – more disjointed

	perceptual quality	circumscription	scale end label
	Crispness	Characteristic which is affected by the impulse fidelity of systems. Perception of the reproduction of transients. Transients can either be more soft/more smoothed/less precise, or - as opposed - be quicker/more precise/ more exact. Example for 'smoothed' transients: A transmission system that exhibits strong group delay distortions. Counter-example: Result of an equalization aiming at phase linearization.	less pronounced – more pronounced
	Speed	A scene is identical in content and sound, but evolves faster or slower. Does not have to be accompanied by a change in pitch. Examples of technical reasons: rotation speed, sample rate conversion, time stretching, changed duration of pauses between signal starting points; movements proceed at a different speed.	reduced – increased
	Sequence of events	Order or occurrence of scene components. Example: A dog suddenly barks at the end, instead - and as opposed to the reference - at the beginning.	unchanged – changed
	Responsiveness	Characteristic that is affected by latencies in the reproduction system. Distinguishes between more or less delayed reactions of a reproduction system with respect to user interactions.	lower – higher
Dynamics	Loudness	Perceived loudness of a sound source. Disappearance of a sound source can be stated by a loudness equaling zero. Example of a loudness contrast: Whispering vs. Screaming.	quieter – louder
	Dynamic range	Amount of loudness differences between loud and soft passages. In signals with a smaller dynamic range loud and soft passages differ less from the average loudness. Signals with a larger dynamic range contain both very loud and very soft passages.	smaller – larger
	Dynamic compression effects*	Sound changes beyond the long-term loudness. Collective category for a variety of percepts caused by dynamic compression. Examples: More compact sound of sum-compressed music tracks in comparison to the unedited original. 'Compressor pumping': Energy peaks in audio signals (bass drums, speech plosives) lead to a sudden drop in signal loudness which needs a susceptible period of time to recover.	less pronounced – more pronounced
Artifacts	Pitched artifact	Perception of a clearly unintended sound event. For example, a disturbing tone which is clearly not associated with the presented scene, such as an unexpected beep.	less intense – more intense
	Impulsive artifact	Perception of a clearly unintended sound event. For example, a short disturbing sound which is clearly not associated with the presented scene, such as an unexpected click.	less intense – more intense
	Noise-like artifact	Perception of a clearly unintended sound event. For example, a noise which is clearly not associated with the presented scene, such as a background noise from of a fan.	less intense – more intense
	Alien source	Perception of a clearly unintended sound event. Examples: an interfering radio signal, a wrongly unmuted mixing desk channel.	less intense – more intense

	perceptual quality	circumscription	scale end label
	Ghost source	Spatially separated, nearly simultaneous and not necessarily identical image of a sound source. A kind of a spatial copy of a signal: a sound source appears at one or more additional positions in the scene. Examples: two sound sources which are erroneously playing back the same audio content; double images when down-mixing main and spot microphone recordings; spatial aliasing in wave field synthesis (WFS): sound sources are perceived as ambivalent in direction.	less intense – more intense
	Distortion	Percept as a result of non-linear distortions as caused e.g. by clipping. 'Scratchy' or 'broken' sound. Often dependent on signal amplitude. Perceptual quality can vary widely depending on the type of distortion. Example: clipping of digital input stages.	less intense – more intense
	Tactile vibration	Perception at the border between auditory and tactile modality. Vibration caused by a sound source can be felt through mechanical coupling to supporting surfaces. Examples: Live Concert: bass can be 'felt in the stomach', headphone cushions vibrate noticeably on the ear/head.	less intense – more intense
General	Clarity	Clarity/clearness with respect to any characteristic of elements of a sound scene. Impression of how clearly different elements in a scene can be distinguished from each other, how well various properties of individual scene elements can be detected. The term is thus to be understood much broader than the in realm of room acoustics, where Clarity is used to predict the impression of declining transparency with increasing reverberation.	less pronounced – more pronounced
	Speech intelligibility	Impression of how well the words of a speaker can be understood. Typical of low speech intelligibility: station announcements. Typical for high speech intelligibility: Newscaster.	lower – higher
	Naturalness	Impression that a signal is in accordance with the expectation/former experience of an equivalent signal.	lower – higher
	Presence	Perception of 'being-in-the-scene', or 'spatial presence'. Impression of being inside a presented scene or to be spatially integrated into the scene.	lower – higher
	Degree-of-Liking	Difference with respect to pleasantness/unpleasantness. Evaluation of the perceived overall difference with respect to the degree of enjoyment or displeasure. Note that 'preference' might not be used synonymously, as, e.g., there may be situations where something is preferred that is - at the same time - not liked most.	lower – higher
	Other	Another, previously unrecognized difference.	less pronounced – more pronounced

*see Appendix D: Sound Examples

Table 11-2: *Hierarchical description system for modifications of perceptual qualities*

The perceived difference is …		
… constant	… varying periodically or otherwise rule-based with time	… varying non-regularly with time
	… in a continuous / discontinuous **manner**	
… **and** depending on scene events / user interaction / independent.		

Table 11-3: *Hierarchical description system for assessments entities*

All audible events				
Intended audible events (elements of the presented virtual scene)			Unintended audible events	
Foreground sources	Background sources	Room acoustic environment	Reproduction system	Laboratory environment

12 Appendix B: SAQI-GER

Table 12-1: Spatial Audio Quality Inventory (SAQI) - German version („Qualitätsinventar zur Schallfeldvirtualisierung“)

	Wahrnehmungsqualität	**Präzisierende Ergänzung**	**Skalenpole**
	Unterschied	Existenz eines wahrnehmbaren Unterschieds.	gar keiner - sehr großer
Klangfarbe	Klangfarbe hell-dunkel	Klangeindruck der durch das Verhältnis hoher zu tiefer Frequenzanteile bestimmt wird.	dunkler - heller
	Klangfarbliche Ausprägung im Höhenbereich	Klangliche Veränderungen in einem begrenzten Frequenzbereich.	Höhen abgesenkt - Höhen angehoben
	Klangfarbliche Ausprägung im Mittenbereich	Klangliche Veränderungen in einem begrenzten Frequenzbereich.	Mitten abgesenkt - Mitten angehoben
	Klangfarbliche Ausprägung im Tiefenbereich	Klangliche Veränderungen in einem begrenzten Frequenzbereich.	Tiefen abgesenkt - Tiefen angehoben
	Schärfe	Klangeindruck der z.B. auf den Kraftaufwand schließen lässt, mit dem ein Klangquelle angeregt wird. Bsp: Hart/weich angeschlagene Perkussionsinstrumente, hart/weich gezupfte Saiteninstrumente (klass. Gitarre, Harfe). Eine Überbetonung hoher Frequenzen kann einen ‘scharfen’ Klangeindruck fördern.	schwächer ausgeprägt - stärker ausgeprägt
	Rauigkeit	Klangeindruck heftiger oder aggressiver Modulation/Vibration, wobei Einzelschwingungen kaum mehr unterscheidbar sind. Oft als unangenehm bewertet.	schwächer ausgeprägt - stärker ausgeprägt
	Kammfilterartigkeit	Oft tonal wirkende Klangverfärbung. ‘Hohler’ Klang. Beispiel: Sprechen durch ein Rohr.	schwächer ausgeprägt - stärker ausgeprägt
	Metallische Klangfarbe	Klangverfärbung, die von schmalbandig-resonierenden Anteilen geprägt ist, häufig als Resultat einer geringen Eigenfrequenzdichte. Häufig bei Anregung von metallenen Gegenständen wie z.B. Gongs, Glocken, scheppernde Blechdosen hörbar. Anwendbar auch auf Raumsimulationen, Plattenhall, Hallfolie u.ä.	schwächer ausgeprägt - stärker ausgeprägt
Tonalität	Tonhaltigkeit	Wahrnehmbarkeit einer Tonhöhe in einem Klang. Beispiele tonhaltiger Signale: Stimmhafte Sprachanteile, Pieptöne.	weniger tonal - tonaler
	Tonhöhe	Die T.-wahrnehmung erlaubt die Anordnung tonhaltiger Signale entlang einer Skala: "höher - tiefer".	tiefer - höher
	Dopplereffekt	Veränderung der Tonhöhe (s.o.). Oft als ‘kontinuierliche Verstimmung’ wahrgenommen. Beispiel: ‘Verstimmter’ Klang der Sirene eines schnell vorbeifahrenden Krankenwagens.	schwächer ausgeprägt - stärker ausgeprägt
Geometrie	Richtung Azimut	Richtung von Schallquellen in der Horizontalebene.	[entgegen dem Uhrzeigersinn versetzt - im Uhrzeigersinn versetzt] (je bis 180°)
	Richtung Elevation	Richtung von Schallquellen in der Vertikalebene.	[nach oben versetzt - nach unten versetzt] (je bis 180°)
	Vorn-Hinten-Lage	Meint nur die Lage vor bzw. hinter dem *Hörer*. Eindruck des Positionsunterschieds einer Schallquelle, der bei Positionsspiegelung an der durch den Hörer gehend gedachten Frontalebene zustande kommt.	Dichotomes Konstrukt/Kat.-skala: nicht vertauscht-vertauscht
	Entfernung	Wahrgenommene Distanz einer Schallquelle.	näher - ferner
	Tiefenausdehnung	Wahrgenommene Ausdehnung einer Schallquelle in radialer Richtung.	kürzer-tiefer

	Wahrnehmungsqualität	Präzisierende Ergänzung	Skalenpole
	Breitenausdehnung	Wahrgenommene Ausdehnung einer Schallquelle in horizontaler Richtung.	schmaler-breiter
	Höhenausdehnung	Wahrgenommene Ausdehnung einer Schallquelle in vertikaler Richtung.	niedriger-höher
	Externalisierungsgrad	Beschreibt die Deutlichkeit, mit der eine Schallquelle - unabhängig von ihrer Distanz - innerhalb oder außerhalb des Kopfes wahrgenommen wird. Fachlich oft auch zwischen Phänomenen Im-Kopf-Lokalisation und Außer-Kopf-Lokalisation eingegrenzt. Beispiele: Schlecht/nicht externalisiert = wahrgenommener Schallquellenort bei diotischer Schallpräsentation per Kopfhörer; Gut/stark externalisiert = wahrgenommener Schallquellenort beim Hören einer natürlichen Schallquelle in nachhallbehafteter Umgebung unter Zulassen von Bewegungen des Hörers.	internalisierter-externalisierter
	Lokalisierbarkeit	Bei geringer L. sind räumliche Ausdehnung und Ort einer Schallquelle schlecht abschätzbar bzw. erscheinen diffus. Bei hoher L. erscheint eine Schallquelle dagegen klar umgrenzt. Geringe L./große L. gehen oft mit großer bzw. geringer wahrgenommener Ausdehnung einer Schallquelle einher. Beispiele: Schallquellen in stark diffusen Schallfeldern sind schlecht lokalisierbar.	schwieriger lokalisierbar - einfacher lokalisierbar
	Räumliches Zerfallen	Schallquellen, die erfahrungsgemäß eine einheitliche räumliche Gestalt haben sollten, erscheinen räumlich separiert. Mögl. Ursache: Teile der Schallquelle werden verschiedentlich synthetisiert/simuliert und zw. den Syntheseverfahren/-engines besteht eine fälschlicher oder ungewollter Versatz bzgl. räumlicher Parameter. Beispiele: Griffgeräusche und Töne einer Instrumentenquelle kommen nicht vom selben Ort, Frikative und Vokale eines Sprechers werden getrennt synthetisiert und dann fälschlich räumlich versetzt wiedergegeben.	fusionierter - zerfallener
Raum	Nachhallstärke	Wahrnehmung starker Raumanteile, ausgelöst durch ein hohes Verhältnis von reflektierter zu direkter Schallenergie. Führt bei stationärer Anregung zum Eindruck hoher Diffusität (im Sinne eines geringen D/R-Verhältnisses). Beispiel: Die empfundene Nachhallstärke unterscheidet sich wesentlich zw. eher kleinen und sehr großen Räumen, wie z.B. zw. Wohnzimmern und Kirchen.	schwächer ausgeprägt-stärker ausgeprägt
	Nachhalldauer	Dauer des Nachhall-Ausklangvorgangs. Vor allem am Ende von Signalen hörbar.	kürzer- länger
	Nachhallumhüllung	Wahrnehmung des vom-Nachhall-räumlich-umhüllt-Seins. Bei hoher N. kann dem Nachhall nur schwer ein spezifischer Ort, eine begrenzte Ausdehnung oder eine Vorzugsrichtung zugewiesen werden. Eindrücke eher niedriger bzw. eher hoher N. entstehen z.B. bei diotisch vs. dichotisch (z.B. dekorreliert) präsentiertem verhallten Material.	schwächer ausgeprägt-stärker ausgeprägt
Zeitverhalten	Vorechos	Kopien von Schallquellen mit meist geringerer Lautheit bereits vor Beginn des eigentlich intendierten Klangeinsatzes.	schwächer ausgeprägt - stärker ausgeprägt
	Nachechos	Kopien von Schallquellen mit meist abnehmender Lautheit nach Beginn des eigentlich intendierten Klangeinsatzes. Beispiel: Wiederholung der eigenen Stimme durch Reflektion an Gebirgswänden.	schwächer ausgeprägt - stärker ausgeprägt
	Zeitliches Zerfallen	Objekte, die erwartungsgemäß eine einheitliche zeitliche Gestalt haben, erscheinen zeitlich separiert. Ursache analog zu "räumliches Zerfallen" nur: hier zeitl. Versätze bei Synthese. Beispiel: Griffgeräusche und Töne einer Instrumentenquelle kommen nicht zur selben Zeit.	fusionierter - zerfallener

	Wahrnehmungsqualität	Präzisierende Ergänzung	Skalenpole
	Knackigkeit	Eigenschaft, die durch die Impulstreue von Systemen beeinflusst wird. Wahrnehmung des Verlaufs von Einschwingvorgängen, können im Vergleich weicher/verschliffener/weniger präzise, aber auch umgekehrt schneller/präziser/exakter sein. Beispiel für 'verschliffenere' Transienten: Ein Übertragungssystem, das starke Gruppenlaufzeitverzerrungen einfügt. Gegenbeispiel: Ergebnis einer auf Linearphasigkeit abzielenden Phasenentzerrung.	schwächer ausgeprägt - stärker ausgeprägt
	Wiedergabegeschwindigkeit	Eine Szene läuft inhaltlich & klanglich identisch aber offensichtlich schneller oder langsamer ab. Muss nicht mit Tonhöhenänderung einhergehen. Beispiele technischer Ursachen: Umdrehungsgeschwindigkeit, Sample Rate Conversion, Time Stretching, veränderte Pausen zw. Signaleinsätze, Bewegungen laufen mit veränderter Geschwindigkeit ab.	verlangsamt - beschleunigt
	Szenenablauf	Reihenfolge oder Auftreten von Szenenkomponenten. Beispiel: Ein Hund bellt plötzlich am Schluss anstatt - wie in Referenz- zu Beginn.	unverändert - verändert
	Reaktionsschnelligkeit	Eigenschaft, die durch Latenzen im System beeinflusst wird. Zur Unterscheidung on einerseits mehr andererseits weniger verzögerten Reaktionen der Wiedergabeumgebung auf Nutzerinteraktionen.	geringer - höher
Dynamik	Lautheit	Wahrgenommene Lautstärke einer Schallquelle. Verschwinden von Objekten ist durch Lautheit = 0 abbildbar. Beispiel eines Lautheitsgegensatzes: Flüstern vs. Schreien.	leiser - lauter
	Dynamik	Größe der Lautheitsunterschiede zwischen lauten und leisen Passagen. Bei Signalen geringerer Dynamik unterscheiden sich laute und leise Passagen weniger von der durchschnittlichen Lautheit. Dagegen enthalten Signale mit hoher Dynamik sowohl sehr laute als auch sehr leise Passagen.	geringer - höher
	Kompressoreffekte	Klangveränderungen jenseits des langfristigen Lautheitsverlaufs. Sammelkategorie für eine Vielzahl von durch Dynamikkompression hervorgerufenen Perzepten. Beispiele: Kompakterer Klang eines summenkomprimierten Musiktracks gegenüber dem unbearbeiteten Original. 'Kompressorpumpen': Bei Signalenergiespitzen (Bassdrumeinsätze, Plosivlaute) fällt die Signallautheit plötzlich ab und kehrt nach einer spürbaren Zeitspanne wieder auf das vorherige Niveau zurück.	schwächer ausgeprägt - stärker ausgeprägt
Artefakte	Tonhaltiges Fremdgeräusch	Ausbildung einer eigenständigen, in der Szene eindeutig nicht intendierten Wahrnehmungsgestalt. Beispiel: Ein eindeutig nicht zur präsentierten Szene gehöriger Störton, wie z.B. ein unerwarteter Piepton 'aus der Technik'.	schwächer ausgeprägt - stärker ausgeprägt
	Impulshaftes Fremdgeräusch	Ausbildung einer eigenständigen, in der Szene eindeutig nicht intendierten Wahrnehmungsgestalt. Beispiel: Ein eindeutig nicht zur präsentierten Szene gehöriges, kurzes Störgeräusch wie z.B. ein Knacksen 'aus der Technik'.	schwächer ausgeprägt - stärker ausgeprägt
	Rauschhaftes Fremdgeräusch	Ausbildung einer eigenständigen, in der Szene eindeutig nicht intendierten Wahrnehmungsgestalt. Beispiel: Ein eindeutig nicht zur präsentierten Szene gehöriges Rauschen wie z.B. ein Hintergrundrauschen von Lüftern o.ä.	schwächer ausgeprägt - stärker ausgeprägt
	Fremdquelle	Ausbildung einer eigenständigen, in der Szene eindeutig nicht intendierten Wahrnehmungsgestalt. Beispiele: ein eingekoppeltes Radiosignal, ein versehentlich nicht 'stumm' geschalteter Mischpultkanal.	schwächer ausgeprägt - stärker ausgeprägt

	Wahrnehmungsqualität	Präzisierende Ergänzung	Skalenpole
	Geisterquelle	Räumlich getrenntes, annähernd gleichzeitiges nicht unbedingt identisches Abbild einer Schallquelle. Eine Art örtliche Signalkopie: Eine Schallquelle taucht an einem oder mehreren zusätzlichen Orten in der Szene auf. Beispiele: zwei Schallquellen geben fälschlich denselben Audioinhalt wieder, Doppelabbildung bei Mischungen mit Haupt-/Stützmikrofonierung, räuml. Aliasing bei WFS: Schallquellen werden als richtungsmehrdeutig wahrgenommen.	schwächer ausgeprägt - stärker ausgeprägt
	Verzerrungen	Perzept infolge von nichtlinearen Verzerrungen, wie sie z.B. durch Übersteuerungen entstehen. ‚Kratziger' oder 'kaputter' Sound. Oft von Signalamplitude abhängig. Kann seine Qualität je nach Art der Übersteuerung stark ändern. Beispiel: Clipping bei Übersteuerung von digitalen Eingangsstufen.	schwächer ausgeprägt - stärker ausgeprägt
	Vibration	Wahrnehmung am Grenzbereich zwischen auditiver und taktiler Modalität. Spürbarkeit von Vibrationen, die von einer Schallquelle verursacht werden, z.B. durch mechanische Ankopplung an Auflageflächen. Beispiele: Livekonzert: Bass 'geht in den Magen', Kopfhörerauflagen vibrieren spürbar auf Ohren/an Schläfe.	schwächer ausgeprägt - stärker ausgeprägt
Allgemeines	Klarheit	Klarheit/Deutlichkeit beliebiger Szeneninhalte. Eindruck davon, wie klar Szeneninhalte voneinander unterschieden, wie gut verschiedenste Eigenschaften einzelner Szeneninhalte erkannt werden können. Der Begriff ist also weiter gefasst, als der in der Raumakustik durch das Klarheitsmaß prädizierte Eindruck einer mit steigender Nachhallenergie sinkenden Transparenz.	schwächer ausgeprägt - stärker ausgeprägt
	Sprachverständlichkeit	Eindruck davon, wie gut die Worte eines Sprechers verstanden werden können. Typisch für geringe Sprachverständlichkeit: Bahnhofsdurchsagen. Typisch für hohe Sprachverständlichkeit: Nachrichtensprecher.	geringer - höher
	Natürlichkeit	Eindruck, dass ein Signal der Erwartung/Erfahrung an ein solches Signal entspricht.	unnatürlicher - natürlicher
	Präsenz	'In-der-Szene-Sein' im Sinne räumlicher Präsenz. Eindruck in einer präsentierten Szene vor Ort, in die Szene räumlich integriert zu sein.	geringer - höher
	Gefallen	Unterschied bzgl. Angenehmheit/Unangenehmheit.	gefällt weniger - gefällt mehr
	Sonstiges	Weiterer, bisher noch nicht erfasster Unterschied.	schwächer ausgeprägt - stärker ausgeprägt

*see Appendix D: Sound Examples

Table 12-2: *Hierarchical description system for modifications of perceptual qualities (German version)*

Der wahrgenommene Unterschied ist …		
… konstant	… periodisch oder anderweitig regelhaft zeitveränderlich	… nicht regelhaft zeitveränderlich
	… und dabei stetig / unstetig	
… **sowie** szenenabhängig / interaktionsabhängig / unabhängig.		

Table 12-3: *Hierarchical description system for assessments entities (German version)*

Alle Schallereignisse				
Gewollte Schallereignisse (Teile der präsentierten virtuellen Szene)			Ungewollte Schallereignisse	
Vordergrundquellen	Hintergrundquellen	Raumakustische Umgebung	Übertragungssystem	Laborumgebung

13 Appendix C: Glossary of Terms

Table 13-1: Glossary of terms

Notion	**Definition**
scope of the vocabulary	Consensus vocabulary for evaluating apparatus-related perceptual* differences between technically generated acoustic environments (→VAES) as well as with respect to the presented or imagined acoustic reality. *→unimodal or →supramodal **auditory** aspects
intention of the vocabulary	The SAQI vocabulary in total (including →perceptual descriptors, →circumscriptions, →scale end label, and - if given - →illustrative sound examples) is intended to - without further explanation - enable each →expert in the field to train any laymen to its valid and reliable use in the qualitatively differentiated auditive assessment of →VAEs.
elements of the vocabulary:	
Perceptual (or auditory) quality	Designation in SAQI was mandatory. Psychological characteristic. To be formulated as →self-explanatory as possible. Sometimes quality names may have a more 'technical' than actual perceptual appeal (e.g., 'dynamic range', 'reverberation time'). However, descriptors were always thought of as describing "a perception of [quality name]". Perceptual qualities are already sorted into categories.
circumscription	Designation in SAQI was optional. Mostly given if descriptors were not considered to be →self-explanatory. Typical examples would be: further explanations, synonyms, reference to typical physical causes, reference to operationalization, as e.g., by proposal of scale labels, onomatopoetic transcriptions, or illustrative sound examples.
scale end label	Designation in SAQI was mandatory. To serve as scale en label in scales of a semantic differential. Depending on the perceptual quality scales were dichotomous, unipolar or bipolar.
illustrative sound example	Designation in SAQI was optional. Sound examples were given if the descriptor of the → perceptual quality AND its →circumscription were not perceived to be →self-explanatory. Should demonstrate the desired auditory quality in a typical and obvious way.
Modifications (of perceptual qualities):	A major practical interest during creation of the vocabulary was to identify auditory qualities that would permit the finest possible differentiation with respect to various potential technical causes. In this context, it was found that the majority of the qualities can be further differentiated which are typical for the technology of →VAEs. Hence, we identified five typical modifications of *temporal behaviour*. Additionally - and mostly due to spatial discretization in →VAEs - modifications of the temporal behaviour can often be further be distinguished with regard to the perceived *continuity* of the variation (*continuous, discontinuous*). A second way of modification was identified regarding *interactivity*. Hence, constant or time-varying, continuous or discontinuous variations may be relatable to *user interactions*, *to scene events* or *none of the two*. For some of the qualities specified in more detail by those modifications common or standard terms may already exist, e.g.: periodic change of the pitch: *vibrato*; periodic change of loudness: *tremolo*. Additionally, not every modification that can possibly be thought of might be practically useful or relevant. In each case, the experimenter has to decide beforehand whether an assessment is meaningful.

Notion	Definition
temporal variation: constant	Constant difference (permanent during presentation).
temporal variation: varying periodically or otherwise rule-based	Constantly rule-like or "periodic" variation (permanent during presentation).
temporal variation: varying non-regularly	Temporal variation without recognizable regularity (permanent during presentation).
temporal variation: continu-ous/discontinuous	Noticeable discontinuity of qualitative variation (caused e.g., by coarse granularization of simulation parameters).
causality: depending on scene events	Relates to entirety of scene elements (i.e. to audio contents, too). In each case, the experimenter has to decide beforehand whether an assessment is meaningful.
causality: depending on user interaction	Relates to any interaction of the user. In each case, the experimenter has to decide beforehand whether an assessment is meaningful.
causality: independent	Independent from user interaction and scene events. More clearly, means both (a) happening when no scene or interaction is present AND (b) happening independent from user interaction and scene events. In each case, the experimenter has to decide beforehand whether an assessment is meaningful.
assessment entities:	
foreground sources	As part of the presented scene, hence, usually intended. The experimenter has to decide beforehand whether an assessment is meaningful.
background sources	As part of the presented scene, hence, usually intended. The experimenter has to decide beforehand whether an assessment is meaningful.
room acoustic environment	As part of the presented scene, hence, usually intended. The experimenter has to decide beforehand whether an assessment is meaningful.
reproduction system	Includes for example the recording system, signal processing algorithms, the playback system and the reproduction room's acoustics (i.e. for →VAEs usually unintended). The experimenter has to decide beforehand whether an assessment is meaningful.
laboratory environment	Remaining acoustical influences, additionally to already mentioned (e.g. HVAC, environmental noise, i.e. for →VAEs usually unintended). The experimenter has to decide beforehand whether an assessment is meaningful.
Further terms in alphabetical order:	
consensus vocabulary	By means of group discussions consensually agreed set of descriptors for perceptual qualities that characterize as completely as possible the object of study in its entirety.
expert in the field	Intended user group of the SAQI vocabulary. Typically, experts will be developers and / or researchers employed in the development and evaluation of virtual acoustic environments. It is believed that group of experts which created this vocabulary corresponds to a representative sample of this user group.
self-explanatory	The status 'self-explanatory' was assumed to be fulfilled for a descriptor of a perceptual quality, if, in our expert group - serving as a representative for the targeted expert user group - a consensus was reached about that. The self-explanatory perceptual descriptor is preferred over an ordinary 'definition', as giving such definitions for perceptual qualities is - to our understanding - in principle impossible. If definitions are tried to be given (see e.g., German standard DIN 1320, 2009), they often appear to be synonymous, tautological, referring to physical causes or an operational example (e.g., a scale) in the end. However in most cases it was considered helpful to add some clarifying circumscriptions (not definitions!) to the terms.
sound source	As sound sources we understood the entirety natural and technical sound generators or their virtual representations (i.e., for example, speakers, instruments, loudspeakers) and not only their sound-

Notion	Definition
	producing parts (at least as long as the latter are not specific targets of an investigation). This is not thought to be contradictory to a sound source being perceived as a combination of different auditively distinguishable parts (e.g. fret or string noises and tones of sounding strings).
scene	Entirety of acoustic stimuli purposely provided by the →VAE. May comprise everything from a singular →sound source to a complex acoustical environment.
supramodal auditive	Related to the auditory impression of general qualities, not referring to hearing in the first place, examples: degree-of-liking, naturalness, presence, or clarity.
unimodal auditive	Impression exclusively referring to auditory sensory perception, examples: loudness, reverberance, timbre.
Virtual Acoustic Environment (VAE)	We understood VAEs in a wide sense as all possible combinations of algorithms and instrumentation for the simulation, measurement, coding, processing and reproduction of spatial sound fields.

14 Appendix D: Sound Examples

In order to achieve a better understandability for the SAQI perceptual qualities

- roughness,
- comb filter coloration, and
- dynamic compression effects

illustrative audio examples have been prepared (see [6]). The following three files can be found in the folder '2 audiofiles':

- "roughness - 3 examples.mp3"
- "comb filter like - 7 examples.mp3"
- "compressor effects - 2 examples.mp3".

Each file includes a number of short examples sounds. These examples sounds are always presented first as an untreated (original) version which is then immediately followed by one or more versions that have been treated in order to illustrate the targeted auditive quality.