



A Layer Model of Sound Quality

Jens Blauert¹, Ute Jekosch²

¹ Institute of Communication Acoustics, Ruhr-University Bochum, Germany

² Chair for Communication Acoustics, Technical University Dresden, Germany

jens.blauert@rub.de

Abstract

Sound-quality is a complex and multi-layered phenomenon [1]. When analysing or modelling the formation process of sound-quality judgements, a variety of quality elements and quality features [2] have to be taken into account, whereby the actual relevance and salience of each of them is situation dependent. In this paper, we present some ideas with the aim of structuring the quality-formation process into different layers according to the degree of abstraction involved. Depending on this degree, different sets of references and evaluation and assessment methods have to be employed.

Index Terms: sound-quality assessment, abstraction level

1. Introduction

To explain our ideas in a figurative way, we start with some epistemological reflections: A common way of looking at reality and perception is as follows: There exists a physical world independent of observers that may stimulate the sensory organs of individuals and thus give rise to the appearance of sensory objects in the perceptual worlds of these individuals. This way of looking at the world is called *scientific realism* in philosophy and, although it is understood that scientific realism is inherently transcendental [3], most engineers think that way because this perspective carries a long way as a conceptual model for many engineering tasks.

An alternative way of looking at the world is known as *constructivism* [4,5,6]. This approach to reality postulates that the world and everything in it is our “brainchild”, that is, appears and exists in the course of biological brain functions – thus forming a relationship in the sense of bijective mapping.

We certainly do not aim at entering into philosophical arguments in this paper but we think that the constructivists’ approach can serve as a useful model for structuring the process of quality formation on strict perceptual grounds. Since we do have to understand perceptual processes when dealing with quality formation we go for the constructivists’ perspective for the sake of investigation. The epistemological question of whether there is existence without perception at all is irrelevant for our approach and will thus not be discussed here.

Those of our readers who have problems to accept constructivism as a general model of reality are kindly asked to take our argumentation as a thought experiment and, as such, as a basis for a conceptual model just for the purpose as laid out above.

2. The nature of percepts

Under the constructivist’s assumption that any– and everything that exists in our world is essentially a percept linked to brain function, we can now ask the question of which kinds of percepts exist at all.

There is indeed a wide agreement that – in congruence with the classical triadism of soul, body and mind – there are three basic kinds of percepts that our perceptual world is essentially composed of. Following [7], they may be denoted *feelings*, *things* and *concepts*.

- *Feelings* (e.g., hunger, fear, pain, sadness, joy) are linked to processes inside the human body (incl. subcutane ones) and they usually are also perceptually localized just there.
- *Things (sensations)* are percepts which are related to the sensory organs. In audition we usually call them auditory events, yet, there are visual, olfactory, tactile, and gustatory events as well. Things are usually localized outside the body or upon its surface, yet, particularly in audition, inside-the-body-locatedness also happens (e.g., tinnitus, borborygmus.)
- *Concepts (thoughts, ideas, notions)* are associated to feelings and things but are not supposed to result directly from input by body-internal sensors or sensory organs. Concepts actually represent remembrances of feelings and/or things. This makes them a very special kind of percepts for the following reason: On the one hand, concepts are experienced as perceptual items of their own kind. On the other hand, however, they stand for something else, namely, for feelings and/or things. In other words, concepts are descriptors of feelings and/or things, relating to co-ordinative and characteristic features of these.

Concepts can be grouped according to the degree of abstraction by which they refer to the features of their source percepts. We talk of *individual*, *collective* and *final concepts*. For example, the remembrance of the face of a specific person is an individual concept. A generic notion like “horses” would be a collective concept. Collective concepts may incorporate different degrees of abstraction, e.g., the notion “fauna” has a higher one than the notion “horse”, the notion “life” even more so. Final concepts show the highest possible level of abstraction, for instance, “existence vs. “non-existence”.

Abstraction, in this context, is understood as a process of selecting and grouping relevant and of leaving out irrelevant and/or redundant distinct particulars of the items of consideration, resulting in some sort of generalization but preserving the essence.

3. Abstraction levels involved

Sound-quality evaluation and assessment encompass processes where concepts are inherently involved. Consequently, our approach of making use of the degree of involved abstraction as a structuring property is obviously justified. Based on this

approach four different quality layers have been identified [8] and put into order according to increasing degree of abstraction, namely,^{*)}

- *Auditive quality*
- *Aural-scene quality*
- *Acoustic quality*
- *Aural-communication quality*

Both these labels and this ordering of the different layers relate to the constructivist's model of reality as set forward above. They may come out differently when the scientific-realism model is applied. The reason for this lies in the different notion of objectivity in the two models. In scientific realism the "reality" as postulated behind perception is considered objective because it is assumed to be existing independently of being observed. In constructivism, objectivity denotes the degree of inter-individual agreement, that is, the degree as to which an observation is independent from a specific observer.

We shall now discuss the four levels in more detail, thereby taking our examples predominantly from audio technology, room acoustics and product-sound design. This is no general restriction of our arguments, but readers should be aware that sound quality is highly situation specific. This means that any argument has to be weighted with respect to, for instance, the specific product and the specific use case, always taking the individuality of the observers into account.

We like to recall at this point of our discussion that, according to an accepted definition, "quality" marks the perceptual distance between a set of sound-character features and a set of reference features [1, 9]. The distance is a measure of the degree to which expectations are fulfilled – or even exceeded. The references can be external or internal to the observer who judges on the quality, they can represent good or bad quality. Quality is finally assessed on a scale good vs. bad or high vs. low, respectively.

3.1 Auditive quality

Hearing is a biological process in the course of which things appear in the world as being heard. These aural things (usually called auditory events) are modality-specific, that is, essentially different from sensory events of other senses. Like all percepts, auditory events exist at a specific time at a specific location in space, distinct by their characteristic features (attributes). Examples for such characteristic features are loudness, pitch, timbre, roughness, tonal balance, spaciousness. Please note that these features are not things in themselves (loudness is not an auditory event, but every auditory event has a specific loudness). Thus, the characteristic features only come to the fore in the course of description by observers, in other words, conceptually.

Psycho-acoustics attempts at describing the temporal and spatial positions and the characteristic features of auditory events quantitatively, that is, as result of measurement. To this end, thresholds, difference thresholds and points of perceptual equality are determined in listening tests, whereby, in classical psycho-acoustics, the listeners are supposed to attend selectively to the features under investigation, that is, in a discretic (analytic) way. Scales of physical quantities may serve as references for the quantitative description (indirect scaling).

*) In earlier talks and in [8] and we have also used other denominators for the same four levels of abstraction, e.g., sound-quality-as-such for sensation quality, transmission quality for acoustic quality, presentation quality for aural-scene quality, and product-sound quality or content quality for aural-communication quality. Further, we have now interchanged the order of acoustic quality and aural-scene quality.

Also, the magnitudes and intervals of perceptual features may be estimated directly (direct scaling) [e.g., 10].

Classical psycho-acoustic further tries to arrive at judgments on the perceptual features which are not influenced by the experimental context and/or any cognitive interpretation – such judgments sometimes being called "unbiased". The assumption behind this is that unbiased judgments selectively reflect functions of the auditory periphery. To come close to this goal, simple acoustic signals are preferably used in the listening tests, such as impulses, tone bursts, sinusoids and noises of different bandwidths.

Although it is well questionable whether unbiased auditory judgment exists at all [e.g., 11, 12] – mind that already the test conditions may impose some bias, e.g., headphone vs. loudspeaker presentation, light vs. darkness – they are nevertheless commonly postulated.

From the abstraction-level point of view, classical psycho-acoustics judges on features of auditory events directly and concretely, deliberately avoiding any abstraction. Sound-quality judgments on the basis of psycho-acoustic features must thus be taken as judgments on the sound quality "as such", that is, on a kind of quality where no cognitive interpretation takes place during perception. Yet, the formation and articulation of quality judgments regarding such features requires conceptual involvement.

Since quality judgments require references, the question arises of what these references may actually be in the case of auditive quality. We would like to offer an explanation here for which we do not yet have direct empirical evidence, but which makes some intuitive sense: Everybody knows that auditory events annotate to specific feelings, that is, they have a feeling-related component to them. There are, for example, hungry, fearful, painful, sad and joyful sounds. For instance, a shrill sound is "painish", a hollow one "hungryish" or "fearish". This does not mean that actual feelings must be apparent when listening to these sounds. Yet, if this happens, that is, feelings are actually present in association with auditory events, then we speak of affect or emotion.

Feelings, as we know, can be classified as positive or negative. Our hypothesis now is that the feelings which the sounds symbolize determine the judgment on sound-quality-as-such, that is, the judgement on auditive quality. If a sound stands for positive (good) feeling, its quality is rated high, if it stands for negative (bad) feeling, its quality is rated low. Consequently, the references that listeners apply when performing this basic kind of sound-quality judgment must be looked for in a region of the mind where those concepts are situated which reflect feelings. Unfortunately, the specific correlations between auditory events and feelings are not yet well investigated into, both in psychological and physiological terms. However, some relevant research is on its way [13, 14].

3.2 Aural-scene quality

Under the aspect of auditive quality (see Sect. 3.1.) we have considered judgments of listeners when attending to selected features of auditory events in a discretistic (analytic) way. This is, however, a rather artificial way of listening to things. Usually, the discretistic mode becomes active only when we are confronted with unexpected events. Usually, we use to be in a more syncretistic (integrative, holistic) listening state-of-mind, that is, we do not attend to discretistic features of auditory events but rather to aural objects as perceptual entities and to scenarios composed of such entities (aural scenes).

The formation of aural objects and their grouping in the course of compositions of aural scenes are based on processes in the auditory system and in the brain which involve a substantial degree of abstraction. Mind that objects are not things but constructs in the sense of individual concepts!

There is agreement among experts that part of the object formation and grouping proceeds more or less automatically, according to relatively fixed rules which are assumed to develop in the course of early experience [15]. Examples of such so-called primitive rules are the well known *Gestalt* rules [16]. They are considered to be backbones of object and scene formation – for instance, the following ones:

- *Proximity* elements are close together
- *Similarity* elements appear to be alike
- *Common fate* elements move according to a pattern, e.g., a wave
- *Good continuation* elements are aligned in the same direction
- *Closure* elements are connected to each other although they form an incomplete structure
- *Simplicity* elements are “simple”
- *Habit* elements form familiar and regular shapes
- *Persistence* element grouping recurs over time such as initially grouped

At the abstraction level addressed in aural-scene-quality assessment, we see auditory effects like the precedence effect, auditory-stream segregation (e.g., the cocktail-party effect), attention-controlled effects, rudimentary phone and melody recognition but no speech comprehension yet. In other words, object formation is included and, consequently, a significant amount of abstraction. Yet, higher intellectual functions like symbolic representation are not yet addressed. To express it in a placative way, we talk of conceptual processes as also animals like birds can perform. Obviously, already at this level, cross-modal information is taken into account, for example, visual cues may lead to direction-finding head movements.

The implications of all this in terms of sound-quality assessment can, for example, be illustrated by considering the tasks of tonmeisters (audio-recording experts), who actually typically focus on effects at the aural-scene level. Firstly, tonmeisters pick-up sound signals that faithfully mirror the relevant characteristics of the acoustic signals as emitted by sound sources, e.g., speakers or musical instruments. Then, secondly, they process and mix these signals in such a way that the consumers, that is, listeners in a play-back situation, experience aural scenes at a desired level of aural-scene quality (quality of presentation). The desired quality features may, for instance, include issues like a good balance of timbre, aural transparency, clear identification and localization of sound sources in a mixture, high speech intelligibility, clear audio perspective (incl. proper distance cues).

In order to realize such features, tonmeisters need methods to assess them. Since they are experienced listeners and know many of the quality elements and features well, they often use their own “golden” ears for this purpose. Yet, since most of them are trained in audio technology and psycho-acoustics as well, they are also capable of making use of the “classical” psycho-acoustical and acoustical measuring methods as mentioned in Sections 3.1. & 3.3 – whenever helpful.

When producing musical recordings or arranging play-backs for musical performances, tonmeisters actually design aural scenes such as to achieve a perceptually coherent product, thereby certainly taking basic psycho-acoustic and physical criteria into account – but always drawing upon their rich professional experience. In other words, they listen and interpret what they hear, and then they design and realize aural scenes as perceptually coherent entities.

Perceptual judgments are, of course, employed in this context, however, these are no longer only basic psycho-

acoustic ones, because, among other reasons, the program material (the content) is now taken into account as relevant context. Quality-building features which are typically discussed at this level are, for example, spatial impression, immersion, presence, ease of listening, perceptual plausibility.

These features, to be sure, represent notions of considerable abstraction. As we have mentioned, they are usually also very context-dependent. Thus, there is indeed much interpretation going on in the course of the realization of an aural product of convincing aural-scene quality – among other things by considering emotions, actions and prior knowledge of the prospective listeners. However, the program material itself – for example, a composition – is not part of the tonmeisters’ job. In other words, tonmeisters do their best to convey and present program material (content) effectively. They do not create this content themselves but they put them effectively into scene. This is indeed, why we talk of aural-scene quality at this level of sound-quality assessment.

It is also the aural-scene quality which, for instance, is the issue when the aural-output quality of audio codecs, loudspeakers or surround-sound systems has to be assessed. Here, again, it is the perceptual form of the aural objects and scenes rendered by such equipment which is evaluated in these cases, but not the content – e.g., a musical composition.

The methods for perceptual assessment of abstract perceptual constructs vary significantly from those used for scaling discretistic features as in auditive-quality assessment. There are actually two classes of measuring tasks which have to be distinguished at this point, namely, (i) perceptual analysis of aural objects and scenes with the aim of identifying and scaling the perceptual components that they are composed of, and (ii) assessment of the presentation as a whole, e.g., with the aim at arriving at single-number criteria such as the mean opinion score (MOS) [e.g., 17]. By the way, the common reductionists’ assumption that a (weighted) sum of discretistic components can render a valid syncretistic judgment should not be made without a stringent validity check. A syncretistic judgment may well encompass more than the sum of the discretistic components.

For tasks of class (i), methods like the semantic differential and/or multi-dimensional scaling are in use [17, 18] – also in combination. Various variations of these methods exist. For class-(ii) tasks, methods that scale preference and/or appropriateness (suitability) are usually chosen.

For the assessment of aural-scene quality the measured values have finally to be compared to appropriate reference values. Unfortunately, establishing valid sets of references – which, to be sure, have to be specific for each specific product and use case and for individual observers – requires a lot of time and effort in listening tests.

To minimize the necessary efforts, very often a specific product which is considered to offer a high aural-scene quality is chosen as a reference (e.g., a particular loudspeaker, codec, or even a musical record). Other products are then “benchmarked” against it. In this case, just quality differences have to be evaluated and not absolute quality. This eases the task considerably. Further, by restricting the evaluation to quality differences, question of the following kind can be tackled: “Which one from a selection of available surround-sound systems is best for a specific purpose and a specific user group in terms of aural-scene quality?”

Attempts have been made or are currently followed up [e.g., 19] to also predict more abstract perceptual features instrumentally, such as speech intelligibility, audio sound quality and/or quality of experience (QoE). They are built around elaborate models of the auditory system. Some of the proposed methods have even been standardized, for instance, PESQ (perceptual evaluation of speech quality) and PEAQ (perceptual evaluation of audio quality) [see, e.g., 17]. Yet,

most of these procedures do not yet sufficiently include processing steps beyond the periphery of the auditory system and are thus limited in their applicability.

As we have pointed out more than once, the context obviously plays an important role for aural-scene quality. This requires model algorithms that include explicit knowledge, namely, cognition (e.g., implemented as top-down processes).

3.3 Acoustic quality

Acoustic quality concerns the physical aspects of audition, that is, the *acoustics* of it. By definition, acoustics (as a branch of physics) is the science of vibration and waves in elastic media. The vibrations and waves are, consequently, physical objects. The relevant issue for our discussion here is that elasto-dynamic vibrations and waves are correlated with auditory events. For instance, if acoustic waves in a frequency range from about 16 Hz to 16 kHz enter the ear canals of human beings, it is highly probably that auditory events pop up in the worlds of these persons – provided that the listeners are awake and their auditory systems are intact.

Assuming that the relation between the wave properties and the features of auditory events were known to such detail that they could be mapped bijectively upon each other, it were possible to predict auditory events solely from the results of physical measurements and theory-based calculations. This would be advantageous for the following reasons: Physical measurement methods are engineered to render results with a high degree of objectivity, that is, independent from specific laboratories and/or specific experimenters. Also, instrumental methods are readily available that measure acoustical quantities with high precision. Further, physical theories are based on strictly logical grounds, experimentally verified and, thus, by definition, highly objective as well.

By the use of wave quantities for predicting features of auditory events it is hoped that the precision and objectivity of physics can be extrapolated to the features of the associated auditory events. Also, the high prediction power of physical measurements and calculations might thus become applicable to and exploitable for technological tasks in audition as well.

This approach is indeed widely used and fairly successful in more general terms. Nevertheless, it is necessary, particularly in the context of our current article, to analyze it more closely regarding quality assessment. The constructivists' point of view is helpful again since it is strictly perception-based. Actually, there are two aspects of acoustics as a branch of physics that we want to consider at this point.

- *Physical objects are conceptual constructs*
Obviously, physical objects and the physical quantities used to describe them are abstracted from the actually observed things (sensory sensations) from which they are derived. In fact, they are not the things themselves but collective concepts of these.
- *Acoustics is mainly based on visual conception*
Acoustical vibrations and waves are co-ordinative alteration of/in elastic media like liquids and solids, that is, they are fluctuations in space and time. These fluctuations are predominantly conceived and measured visually – only in rare cases tactilely (e.g., length differences) or auditorily (e.g., time differences). In other words, the concept of acoustic vibrations and waves is perceptually most closely related to visual events. In any case, the science of (physical) acoustics is hardly based on audition.

It goes without saying that the approach of mapping auditory events on acoustic (physical) quantities which are (i) concep-

tual constructs and (ii) originate from observations through other sensory modalities than audition, embodies a significant degree of abstraction. Hence, when applying (physical) acoustics to sound-quality assessment, one has to be aware of a number of restrictive implications. The most relevant ones are the following.

Since the physical model of the world represents an abstraction from sensory reality – which, by the way, actually constitutes its generalizability and predictive power – the (abstract) model predictions never exactly meet the (concrete) individual sensations of listeners. Further, as we have stated above, acoustics is to a great extent based on our conception of the visual world. To relate the physical model to the auditory world is thus a cross-modal endeavor. However, cross-modal relations have only recently become subject of systematic research [e.g., 20, 21, 22, 23].

Physical acoustics deals solely with vibrations and waves, that is, with fluctuations of matter in space in time – expressly, with co-ordinative features only. Auditory events, however, are specific in terms of their co-ordinative and characteristic features. This touches upon a problem known as *qualia* in philosophy [e.g., 24]. In vision, for instance, a spectral representation of electro-magnetic waves cannot de facto explain what a colour essentially means perceptually. The same holds for timbre (tone colour) in audition.

All these restrictive implications boil down to one essential item, namely, the problem of *validity* of physical (acoustic) considerations with regard to their perceptual (sensory) counterparts. In other words, we have to carefully consider the question of the extent as to which acoustic quantities actually represent what we finally want to assess, namely, auditory events with regard to their quality. Having outlined this and being aware of the reservations following from it, we think that acoustic data and physical calculations are, nevertheless, quite useful for sound-quality assessment. This also holds for non-acoustic data (e.g., electric signals) that represent acoustic ones, and for the methods of signal-processing going with them.

The “physicalistic” approach to audition is particularly useful when dealing with aspects of auditory perception that can be well represented by physical conditions. For example, when electrical signals which mirror acoustical signals are sent through an electro-acoustical transmission chain, distortions may occur that harm the acoustic signals at the output of the chain system and, eventually, lead to audible impairment of the related auditory events – thus reducing their quality.

Please mind also that many purely acoustic systems can be conceptualized as transmission chains, for example performance spaces like theaters and concert halls. In fact, the physicalistic approach to sound quality is thus quite popular among consultants for architectural acoustics. Actually, the problem of validity is less pronounced when the methods are only used to detect and track quality differences between systems or different versions of systems.

The set of reference for acoustic quality is usually given by a specification sheet of instrumentally measurable values and their tolerance ranges and/or by explicit target signals. In room-acoustics, for example, the acoustical consultant of a concert hall may have to guarantee a certain reverberation time, and could actually be made liable if this is not achieved – although the reverberation time, if not excessively inadequate, is not a very relevant predictor for the “quality of the acoustics” of a hall from the concert-goers' point of view (problem of validity!).

As physical parameters for sound-quality evaluation and assessment, such acoustic quantities are most popular which can directly be measured instrumentally, for example, sound-pressure level, impulse responses, transmission functions, reverberation times, sound-source positions or lateral

energy fractions. However, more sophisticated parameters as rendered by signal-processing algorithms which partially mimic the auditory system are also available. There are, for example, methods to predict elementary auditive features like loudness, pitch, roughness and sharpness instrumentally.

We finish this section by stressing once more that the assessment of acoustic quality requires conceptual processes of high abstraction, that is, the application of intelligence. The reason for it is obvious: Acoustic-quality assessment requires a physical conception of the world, that is, an abstract, multi-modal world model. The formation of such a model is evidence of “theoretical” thinking.

3.4 Aural-communication quality

The generic purpose of audition is to gather up and process information from and about the environment, that is, to identify sound sources with respect to their positions and states of movement in time and space and to their characteristic features. Furthermore, particularly in human beings, inter-individual communication is predominantly performed via the auditory pathway. This makes audition the prominent social sense of human beings.

Looked at it from a broader perspective, each sound signal and, hence, each auditory event can be conceived as a *sign carrier* through which information about the world is communicated to the end of an assignment of “meaning” being elucidated in the listeners’ minds. Sound quality at its most abstract level is thus *aural-communication quality*. Ergo, sound designers are, essentially, engineers of communication processes. To elaborate further on this statement, we take as an example an issue of pronounced current interest, namely, the notion of *product-sound quality*. This is indeed a very figurative example since the role of sounds as signs comes clearly to the fore [25].

Actually, in product-sound design, it is not the quality of the sound-as-such which is relevant. Rather, product sounds are considered to be of high quality when they clearly and convincingly signal to the potential or actual product user that the product of concern is of high quality itself. In other words, in these cases it is not the quality of the sound but the *sound of quality* which is the issue of interest. The focus lies on the requirement that the product sound must support the quality of the product by signifying the actual product quality.

Signs, in the sense used here, are mental units that are conceived as standing for something other than themselves, that is, as pointing at something else. This certainly implies that they are concepts and not things or feelings. The degree of their conceptual abstraction varies according to the things, feelings, or individual and/or concepts (individual or collective) which are actually signified by them. In the light of these facts, it is obvious that not every sound and/or every auditory event, respectively, is qualified to act as a sign carrier. The suitability of specific sounds for communication purposes is, hence, an element of communication quality.

As taught by semiotics [26], assignment of meaning to an auditory event requires three components – in our case an auditory event conceived as sign carrier, a listener, and the concept of a product as a cognitive referent. With these three determinants being present, an assignment of meaning may take place – such that the sign is understood. Treating items of perception as sign carrier leads to associations, denotations, connotations and interpretations, and, finally, to the assignment of meaning – a process which is called *semiosis* in the field.

The assignment of meaning is actually a complex, ruled and self-organized process, implemented in the form of

so-called schemata. These schemata are understood as having been evolved during prior experience and subsequent learning.

For example, to understand speech sounds, one has to be able to identify phones and, subsequently, phonemes, has to know the syntactic and grammatical rules and, last not least, know the vocabulary. In other cases, understanding of signs may be more intuitively based, for example when conceiving the meaning of the sound of a thunder. Yet, in any case, cognition and, thus, a high amount of abstraction are always involved. It is indeed not far from the mark to hypothesise that the perceptual organization of our world at large is based on the formation of signs, their conception, interpretation and the subsequent assignment of meaning.

This view is of paramount importance for sound-quality assessment for the following reason. As we all know very well, listeners only rarely react to what they hear – that is, in a reflective way, directly responding to their auditory events. Rather, their responses are based on what the auditory events mean to them in their specific life situation.

When evaluating and assessing communication quality, the measuring method as used in psychophysics and physics are of little help. Methods as employed in the evaluation of presentation quality, such as multi-dimensional scaling or preference scaling may be useful, but the scenarios in which the measurements take place have to encompass real or simulated communication processes. An example is the assessment of spoken-dialogue systems [27] where actual dialogues had to be experimentally realized. When the measurement of meaning is the issues, methods as used in cognitive psychology may be applied. Actually, the methods to measure these effects and to transform the results into profiles to be included into references for quality assessment draw heavily upon cognition. It certainly takes extensive expertise in cognitive psychology (incl. semiotics) to manage these tasks.

By the way, although the term product-sound is mainly used for industrial products, the ideas behind it can be applied in a more general way, for example, to musical performances. There is no doubt that musical sounds are considered a medium for communication, that is, as sign carriers. The signs comprise the artistic form, the content and, finally, the function that the composer had in mind when writing his/her composition. Musical communication quality depends on elements like the melody, the timing and the rhythm, but also on the musical stile, the instruments selected, the particular players of these instruments and, last but not least, on the spaces where the performances take place [28].

Aural communication quality as a subcategory of sound quality, namely, as that quality sector which encompasses the highest amount of abstraction, will gain in importance in the context of sound-quality evaluation. To deal with it properly, requires serious efforts across scientific disciplines.

4. Summary & conclusions

In this article, we presented an attempt to structure the broad field of sound-quality evaluation and assessment on strictly perceptual grounds, whereby the degree of abstraction involved is taken as the structuring property. To this end we took the constructionists’ view as to which the world is actually a brainchild. In other words, we assume that the world is composed of percepts, with each percept being bijectively associated to a specific functional brain state.

Percepts can be grouped into feelings, things and concepts. Concepts, although they are percepts themselves, are descriptors of attributes of other percepts, that is, of

feelings, things and/or of other concepts. Their descriptive function must be seen as more or less abstracting. The degree of abstraction, namely, the extent as to which a concept is distanced from and generalized with respect to its referents, is used here to rank-order different levels of sound-quality evaluation.

The degree of abstraction at our lowest layer of quality, *auditive quality*, is given due to the description of coordinative and characteristic features of auditory events. The next higher degree of abstraction is given by what we have called *aural-scene quality*. Here, primitive perceptual grouping is involved, for example, through *Gestalt*-based schematic processes, including cross-modal ones. The intellectual level of *Gestalt* processes is relatively low. It is assumed that many animals are able to perform them. A significantly higher degree of abstraction is required by *acoustic quality*. This quality layer employs the comparisons of features of auditory events with physical quantities and thus requires a physical (conceptual) model of the world, surely an intellectual achievement. Physical quantities are constructs originating from a predominantly visual, objectivistic view on the world. The highest degree of abstraction, though, is established by *aural-communication quality*. At this quality layer, auditory events are conceived as signs, that is, as mental entities that point to referents. These referents are feelings, things or concepts – the latter at different degrees of abstraction. The involved processes are schemata at a high cognitive level.

We are well aware of the fact that our proposal of how to conceptually categorize the field of sound-quality evaluation and assessment, namely, according to the degree of abstraction involved, does not by itself offer any instructions for proper technical action. Yet, we expect that our way of thinking may serve as a means of structuring the problem at large and to deal with it in a more adequate way. It is a prerequisite for valid and precise argumentation to be aware of the abstraction level at which one is operating in each case. For example, when having to select methods for quantitative assessment of sound quality, it is of paramount importance to be aware of the degree of abstraction being involved in every specific assessment task. Only a method which takes account of the specific level of abstraction employed will provide valid results. For example, mistaking physical quantities as concrete percepts may lead to serious systematic errors in the interpretation of experimental results.

Please be reminded that sound-quality judgments are essentially situation specific and individual. Abstraction may provide useful generalization, but any generalization certainly has its limits and, thus, requires strict validity checks.

5. References

- [1] Blauert, J. & Jekosch, U. (1996) Sound-quality evaluation – a multi-layered problem. *ACUSTICA* united with *ACTA ACUSTICA* **83**, 747–753
- [2] Jekosch, U. (2005) Voice- and speech-quality perception. Springer, Berlin–Heidelberg–New York
- [3] Kant, I. (1781) *Kritik der reinen Vernunft* (Critique of pure reason). Johann-Friedrich Hartknoch, LV–Riga
- [4] Berkeley, G. (1710) *Treatise on the principles of human knowledge*, Jeremy Peypat, EIR–Dublin
- [5] Maturana H.U. (1978) “Biology of language: The epistemology of reality”. In: Miller, G.A. & Lenneberg, E. (eds.), *Psychology and Biology of Language and Thought: Essays in Honor of Eric Lenneberg*, 27–63, Academic Press, New York NY,
- [6] Dominicus, R.D. (2009) *Die Subjekttheorie des radikalen Konstruktivismus* (The subject theory of radical constructivism). Diplomica-Verlag, D–Hamburg
- [7] Lungwitz, H., (1947) *Die Entdeckung der Seele* (The discovery of the psyche). 5th ed., Walter de Gruyter, D–Berlin
- [8] Blauert, J. & Jekosch, U. (2007) Auditory quality of concert halls – the problem of references. *Proc. 19th Int. Congr. Acoust., ICA 2007 Madrid*, paper RBA 06–004, *Revista de Acústica* **38**, ES–Madrid
- [9] Blauert, J. & Jekosch, U. (2003) Concepts of sound quality – some basic consideration. *Proc. Internoise*, 72–79, ROK–Jeju Island
- [10] Guilford, J.P. (1954) *Psychometric methods*. McGraw Hill, New York NY
- [11] Blauert, J. & Gusk, R. (2009) Critique of “pure” psychoacoustics. *Fortschr. Akustik, NAG/DAGA’09*, 1550–1551, Dtsch. Ges. Akust., D–Berlin
- [12] Gusk, R. & Blauert, J. (2009) Psychoacoustics without psychology? *Fortschr. Akust. NAG/DAGA’09*, 1518–1519, Dtsch. Ges. Akust., D–Berlin
- [13] Vjästfäll, D. (2003) Affects as a component of perceived sound and vibration quality. *Doct. diss, Chalmers Univ. of Technology*, S–Gothenburg
- [14] Bergman, P., Sköld, A., Vjästfäll, D. & Fransson, N. (2009) Perceptual and emotional categorization of sound. *J. Acoust. Soc. Amer.* **126**, 3156–3167
- [15] Bregman, A. S. (1991) *Auditory Scene Analysis*. The MIT press, Cambridge MA.
- [16] Avant, L. & Helson, H. (1973), “Theories of perception”. In: Wolmann, B.B. (ed.) *Handbook of General Psychology*, 419–448, Prentice Hall, Englewood Cliffs NJ
- [17] Bech, S. & Zacharov, N. (2006) *Perceptual audio evaluation: theory, method and application*. Wiley, GB–Chichester
- [18] Osgood, C.E., Suci, G., & Tannenbaum, P. (1957) *The measurement of meaning*. Univ. Illinois Press, Urbana, IL
- [19] Blauert, J., Braasch, J., Buchholz, J., Colburn, H.S., Jekosch, U., Kohlrausch, A., Mourjopoulos, J., Pulkki, V. & Raake, A. (2009), Aural assessment by means of binaural algorithms – the *AABBA* project. *Proc. Int. Symp. Auditory Audiolog. Res. – ISAAR’09, DANA VOX Jubilee Foundation*, DK–2750 Ballerup (in press)
- [20] Kohlrausch, A. & van der Paar, St (2005) “Audio-visual interaction in the context of multi-media applications”. In: Blauert, J. (ed.) *Communication Acoustics*, Springer, Berlin–Heidelberg–New York
- [21] Haverkamp, M. (2009) Look at that sound! – Visual aspects of auditory perception. *Proc. 3rd Int. Congr. on Synaesthesia, Science and Art*, E- Granada:
- [22] Haverkamp, M. (2007): Essentials for description of cross-sensory interaction during perception of a complex environment. *Proc. Inter-Noise 2007*.: Turkish Acoust. Soc. TR- Istanbul
- [23] Altinsoy, E. (2005), Audio-tactile interaction in virtual environments. *Doct. diss, Ruhr-Universität Bochum*, Shaker Verlag, D-Aachen
- [24] Levine, J. (1993) “On leaving out what it's like”. In: G. Humphreys, G. & Davies, M. (eds.): *Consciousness. Psychological and Philosophical Essays*, 121–136. Basil Blackwell, GB–Oxford
- [25] Jekosch, U. (1999), Meaning in the context of sound-quality assessment. *ACUSTICA/acta acustica* **85**, 681–684
- [26] Jekosch, U. (2005) “Assigning meaning to sounds – semiotics in the context of product-sound design”. In: Blauert, J. (ed.) *Communication Acoustics*, 193–219, Springer, Berlin–Heidelberg–New York
- [27] Möller, S. (2005) *Quality of telephone-based spoken-dialog systems*, Springer, New York NY
- [28] Blesser, B. & Salter, L.-R. (2007) *Spaces speak, are you listening? Experiencing aural architecture*, MIT Press, Cambridge MA