Sound Quality Assessment: Concepts and Criteria

2825 (D-8)

Tomasz Letowski
The Pennsylvania State University
University Park, PA

Presented at the 87th Convention 1989 October 18-21 New York

AUDIO ES ®

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

SOUND QUALITY ASSESSMENT: CONCEPTS AND CRITERIA

Tomasz Letowski
Department of Communication Disorders
The Pennsylvania State University
University Park, PA 16802

ABSTRACT

Auditory assessment of sound quality plays an important role in our work, social life, and recreation and is of professional interest to audio professionals, acousticians, audiologists, musicians, and psychologists. Although the concept of sound quality is widely used, the term itself is not clear and does not have a precise meaning. Such situation causes various conceptual and practical problems. In addition, despite a large number of terms describing sound character these terms do not form a system of well-defined and clearly linked perceptual parameters. Such a system and several related definitions are discussed here.

0. INTRODUCTION

A formalized auditory assessment may be of heuristic or diagnostic character. In former case, the purpose of assessment is to collect normative data while in later case the purpose is to evaluate if and to what extend a specific property or an object differs from the accepted standard. The aim of such auditory assessment is either to gather information about the external world (objects) or about the listeners themselves (subjects). These two dichotomies create four basic domains of auditory judgments as shown in Tab. 1.

	subject-oriented tests	object-oriented tests
heuristic judgments	psychoacoustic research	sound quality assessment
diagnostic judgments	audiological evaluation	diagnostic listening tests

Tab. 1. Basic applications of auditory assessment

Physical sounds stimulating our hearing result in auditory images located in our auditory (perceptual) space defined by a set of auditory sensations [1]. McAdams [2] defined an auditory image as a "psychological representation of a sound exhibiting an internal coherence in its acoustical behavior." The stress on the aspect of coherence is important since different acoustical components can be either combined in the auditory space into a single percept representing a physically meaningful entity or can create separate percepts interfering with one another. A single auditory image can be perceptually analyzed by the listener by focusing his attention on the individual sensations or details of the image, and/or by making changes to his actual frame-of-reference [3]. Several coexisting images also can be merged together into a more generalized picture of the acoustical environment surrounding the listener.

Auditory judgments intended to determine characteristic features and subjective utility of sound signals and/or their sources are called sound quality assessments or applied psychoacoustics, to differentiate them from traditional psychoacoustics which is a behavior-oriented science [4]. Sound quality assessment is a necessary tool in music and audio research as well as in the development of new systems and services. The object of such assessment is perceived sound quality (PSQ) [5], or shorter, sound quality, of auditory images created by external stimuli. Such auditory images are the ultimate criterion of value for all sounds and their sources which existence is justified by our ability to hear and listen.

One of the most challenging tasks facing audio professionals and hearing scientists is the understanding of the capacity of various psychological dimensions used to express our reaction to sound. The large number of terms used in sound description is a blessing for artistic freedom but it is a problem when it comes to meaningful communication between people. As Perkins [6] observed "terminology for [sound] quality has proliferated over the years until we now are mired in a terminological swamp, with terms whose linkage is physiological, anatomical, acoustical, and psychological, all milling around together as referents for what is exclusively a psychological phenomenon: [sound] quality." The task of this paper is to build a road through the swamp.

1. AUDITORY ASSESSMENT: OBJECTS AND CLASSES

Auditory assessment may have either direct or indirect character. Direct assessment takes place when the objects of assessment are sound signal while indirect assessment concerns sound sources and transmission systems. In addition, indirect assessment can be either of a sound source itself (e.g., violin), a person actuating the sound source (e.g., violinist), or both of these elements together (e.g., violin concert performed on the stage). It is important to realize that individual sensations of the auditory image may change their relative importance depending on the object of assessment.

Auditory images are commonly described in terms of loudness, pitch, (apparent) duration, spatial character (spaciousness), and timbre. The first three of these sensations are unidimensional while both spaciousness and timbre are multidimensional. In addition, both spaciousness and timbre carry information about a sound source and its location while the three other sensations are not, in general, sound source specific. Therefore, loudness, pitch, and duration are usually excluded from sound judgments concerning music instruments and sound transmission systems. However, some or all of these sensations should be considered when the object of assessment is the sound itself (e.g., a warning signal) or a person actuating the sound source. There also are many cases when loudness, pitch, and/or duration are judged on the basis of their own merit without simultaneous judgments of spaciousness and/or timbre, e.g., classical psychoacoustics. This leads to three general classes of sound assessment:

- class A - assessment of loudness, pitch, and/or duration,
- class B - assessment of timbre and/or spaciousness, and
- class C - assessment of some or all sensations from both above groups.

For example, auditory assessment of sound recordings belongs to class B since recordings can be listened to at any loudness level. Both pitch and duration (rhythm, tempo) are elements of an "excitation function" of the recording and are not objects of consideration unless both musical and acoustical value of recordings are judged together. In latter case, the assessment would belong to class C. Psychophysical scaling involves tasks class A, while tasks during concert hall assessment can be either class B (long signals) or C (impulses). Quality assessment of telephone transmission belongs to class C since one of the assessment criteria is "loudness loss."

2. TIMBRE AND SPACIOUSNESS

Timbre has been defined by the American National Standards Institute (ANSI) [7] as "that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar." This definition is supplemented by a note stating that "timbre depends primarily upon the spectrum of the stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus." The phrase "similarly presented" refers foremost to sound duration and spatial presentation. In a similar definition by Plomp [8], loudness and pitch are supplemented by duration.

Authors writing about psychoacoustics and music perception usually describe loudness, pitch, and timbre as three orthogonal dimensions of an auditory space. Such understanding of timbre, disguised sometimes as sound quality or tone color, can be found in several dictionaries, lexicons, books, and scientific publications [e.g. 9-10]. However, since the ANSI definition of timbre is not precise, the other, more global, interpretation of timbre also is possible. According to this interpretation both loudness and pitch are two dominating dimensions of timbre. Interestingly, such an interpretation of timbre is directly expressed, or can be inferred from positions taken toward timbre by Schönberg [11], Clark and Milner [12], Winckel [13], Tiffany and Carrell [14], Gabrielsson and Sjögren [15], Bladon and Lindblom [16], and others. The advantage of such interpretation of timbre is that the conceptual possibility exist for talking about the timbre of a single sound and for comparison of timbre of sounds which differ in pitch and/or loudness.

From the practical point of view, timbre is the perceived sound spectrum carrying information about the source of sound. According to Roederer [17] "[timbre] perception is the mechanism by means of which information is extracted from the auditory signal in such a way as to make it suitable for: (1) storage in the memory with an adequate label of identification, and (2) comparison with previously stored and identified information." Therefore, to make the definition of timbre more practical and univocal, the author suggests redefining timbre to the form given below.

Timbre is that attribute of auditory image in terms of which the listener judges the spectral character of sound. Timbre enables the listener to judge that two sounds which have, but do not have to have, the same spaciousness, loudness, pitch, and duration are dissimilar.

The definition of spaciousness, complementary to the definition of timbre, is proposed below.

Spaciousness is that attribute of auditory image in terms of which the listener judges the distribution of sound sources and the size of acoustical space. Spaciousness enables the the listener to judge that two sounds which have, but do not have to have, the same pitch, loudness, duration, and timbre are arriving from different locations.

Both of the above definitions sanction comparison of timbre and/or spatial character of auditory images even if other sensations of those images are not equal. Such situations are very common in practice and include, among others, comparison of atonal sounds and comparison of speech samples produced by different talkers.

Auditory processing has limited capabilities and we are unable to compare effectively several sensations at the same time. Moreover, differences along more sensitive sensations tend to mask differences along other sensations. It can be said that the fewer the number of sensations differentiating one auditory image from another, the greater the reliability of assessment. Therefore, it is practical to compare only those auditory images which are identical, or very similar, in respect to all judgment-non-specific sensations. This is especially important in the case when the judgment-non-specific sensations are such sensitive sensations as loudness and pitch. Removing variability along these sensations enables one "to bring other dimensions into focus" [18].

3. SOUND QUALITY

Timbre, sound quality, and tone color are commonly regarded as synonymous terms [19-22]. The Random House Dictionary [23] refers to [sound] quality as "the tonal color, or timbre, that characterizes a particular [...] sound." At the same time these three terms are applied by various authors to a large number of different concepts which hardly can fit under the meaning of a single term. There is also an enormous variety of other terms, such as sound color, timbral color, spectral color and spectral timbre, which are used interchangeably or in conjunction with one or another of the above terms. And although each of the discussed terms appears to have its own shade of meaning there is little consistency in their usage and different sets of them are treated by different authors as synonyms [24].

In general, the term quality refers to a character of an object or a merit of its superiority. In the domain of psychoacoustics, the first of these connotations is parallel to the connotations of timbre and/or spaciousness while the second has a unique meaning. Therefore, sound quality should imply a rating or assessment rather than be viewed as an equivalent to timbre which is emotionally neutral and simply represents differences between auditory images. One timbre is no better than another unless we set up a target (reference) point and judge the goodness of fit or superiority of one fit over another. Therefore, it seems logical to differentiate between timbre (perceptual attribute) and quality (emotional attribute) of sound. Such a concept has been proposed by the author elsewhere [25]. This concept enables timbre to reflect sound categories while sound quality reflects superiority, fit, and the listener's level of appreciation. Such a differentiation reflects well in the types of scales used for multidimensional (parametric) assessment. Timbre constituents are those which are assessed on the quantitative (dominance or similarity) scales, while quality components are assessed on the preference scales [26].

There is one more practical reason to differentiate between timbre and sound quality. Sound quality extends beyond timbre and incorporates an impression of spaciousness. In such cases one may talk about timbral as well as spatial quality of sound. These two qualities merge together in the overall quality of sound. Sound quality, therefore, should be defined generally enough to embrace all of these meanings. Such a definition is proposed below.

> Sound quality is that assessment of auditory image in terms of which the listener can express satisfaction or dissatisfaction with that image. Sound quality can be judged by comparing images produced by several external stimuli or by referencing a perceived image to the concept residing in the listener's memory.

A simple model of sound quality assessment is shown in Fig. 1. In this model the sound quality consists of two separate qualities: spatial quality and timbral quality. In the domain of perceptual assessment the above terms are replaced respectively by sound character, spaciousness, and timbre. The separate concepts of the character and quality are the consequences of the definition of sound quality adopted above.

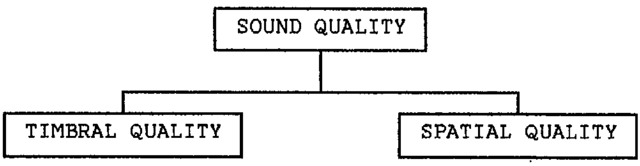

Fig. 1. Basic domains of sound quality assessment.

The model of sound quality assessment presented above is limited to the two multidimensional sensations of timbre and spaciousness what constitutes a framework for sound quality assessment of class B. However, the quality of loudness (comfort of listening or naturalness of loudness), quality of pitch (perceived intonation), and/or quality of duration can easily be added to the model.

4. GLOBAL ASSESSMENT OF SOUND QUALITY

Sound quality is a multidimensional entity and as such may be expressed either in terms of its overall value or in terms of partial values measured for several constituent sensations. These two types of assessment are called respectively (A) global assessment and (B) parametric assessment.

Global assessment of sound quality can be conceived as a weighted function of the auditory image along different perceptual sensations which constitute this quality. Projection of an auditory image on such individual dimensions can be performed by the parametric assessment of sound. The quality of an auditory image can also be assessed globally as a whole. Such a global assessment can be made according to more than one specific criterion of sound quality. The three basic criteria are:

(a) fidelity (accuracy), which relates one auditory image to another,

(b) naturalness, which relates an auditory image to a specific internal standard, and

(c) pleasantness, which relates an auditory image to a set of various internal standards.

Fidelity, or accuracy, is the perceptual measure of the degree of similarity between auditory images produced by two sounds: a compared sound (variable) and a reference sound (standard). Technically speaking, fidelity assessment refers either to (a) similarity between the input and the output signal of a specific transmission system or (b) similarity between signals at the outputs of two or more compared devices (e.g., musical instruments). Perceptual comparison of sounds produced by two hearing aids, violins, signal horns, etc., becomes fidelity judgment if one is regarded as a standard.

Naturalness refers to perceptual similarity between an auditory image produced by a given sound and a generalized conceptual image residing in the memory of the listener and used as the point of reference. Since a reference criterion is an engram residing in a long-term memory such assessment is generally less reliable than fidelity judgment which is based on the direct comparison of two auditory images.

Judgment of pleasantness is a manifestation of the subjective satisfaction resulting from listening to a given sound. This criterion is also used in its negative form as "unpleasantness" or "annoyance" of sound [27-28]. It is important to note that listeners are usually more sensitive and reliable when asked to use the negative form of the scale. Such a relation could be an evidence of the nonlinear character of various rating (categorical) scales used in the assessment of sound pleasantness and, perhaps, other sensations. The nonlinear character of such scales, especially the standard CCIR/ISO/OIRT quality scale ("excellent-good-fair-poor-bad"), has already been suggested [29] and even experimentally confirmed [30]. This calls for extreme caution when equal-interval rating scales are assumed in tasks related to the assessment of sound pleasantness.

Signals, which are unfamiliar to the listener, should be assessed regarding their pleasantness rather than their naturalness. Naturalness of such signals usually does not have clear meaning to the listeners. For example, hearing aid users give quite sensible answers when asked about pleasantness of the amplified sound but become quite confused when asked about its naturalness. It should also be noted that there are situations (synthetic signals, hearing impaired listeners, etc.) where less natural sounds can be regarded as more pleasing.

5. PARAMETRIC ASSESSMENT OF SOUND QUALITY: MURAL

Sound quality cannot be sufficiently well described by a global assessment alone. Global assessment enables identification of the degree to which a particular object satisfies our perceptual needs but is too general to answer any other question. Multidimensional character of sound quality requires multidimensional (parametric) assessments.

An ideal system of parameters for the multidimensional assessment of sound should include a set of well-defined and mutually linked sensations (parameters), a hierarchically organized structure, and a system of weights applied to constituent sensations. Several authors have compared sets of different terms used in sound quality assessment and concluded that many existing terms are related to each other and only a small number of them denote relatively independent auditory sensations. The number of independent dimensions is, however, not clear and varies from 3 to 11 dimensions depending on the author.

There have been several systems of assessment parameters proposed to date. However, these systems are not compatible with each other and they typically mix sound character with sound quality. In addition, those systems provide no relative weights of assessment parameters and do not include any internal linkages between more general and more detailed parameters. Some of these problems are eliminated in the hierarchical system of auditory sensations proposed by the author. This system - called MURAL, which stands for MUltilevel auditoRy Assessment Language - is shown in Fig. 2. Parameters (sensations) which share the same circle of the system are treated as fairly independent and complimentary while parameters related radially are regarded as hierarchically dependent. The system can be used for sound quality assessment as well as for sound character description. This change from one to another type of assessment requires only the change from preference to dominance criteria in the process of assessment.

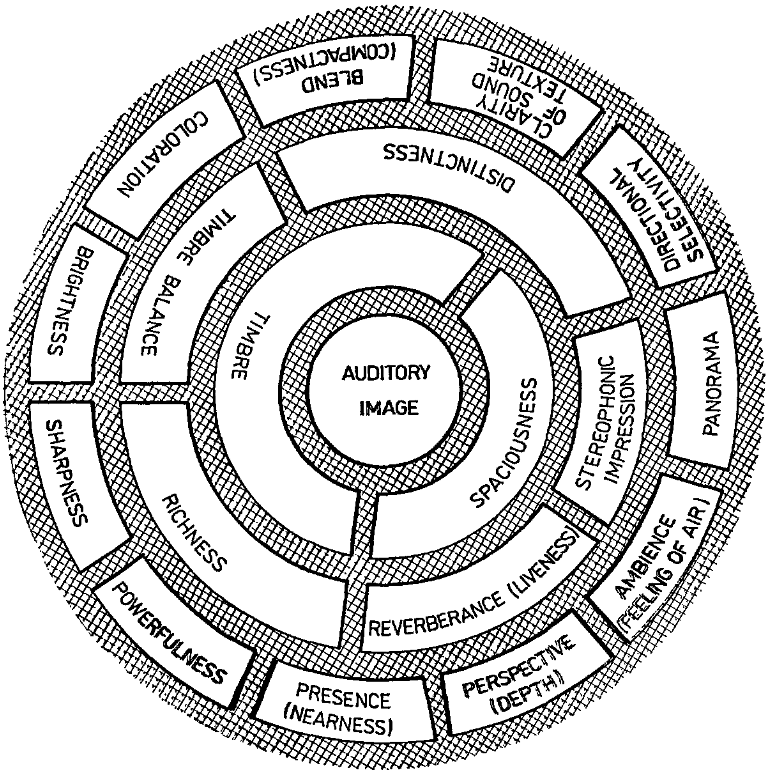
AUDITORY
IMAGE
TIMBRE
SPACIOUSNESS
TIMBRE BALANCE
DISTINCTINESS
STEREOPHONIC IMPRESSION
REVERBERANCE (LIVENESS)
RICHNESS
COLORATION
BLEND (COMPACTNESS)
CLARITY OF SOUND TEXTURE
DIRECTIONAL SELECTIVITY
PANORAMA
AMBIENCE (FEELING OF AIR)
PERSPECTIVE (DEPTH)
PRESENCE (NEARNESS)
POWERFULNESS
SHARPNESS
BRIGHTNESS

Fig. 2. MURAL

The MURAL system resulted from author's earlier research on sound quality assessment [31-33]. The current version of the system, as shown in Fig. 2, is by no means a final product and is open for discussion and critique. The system is still missing relative weights of various parameters. Being known, they can be used as a criterion in the listener selection process or as a powerful tool for identification of major deficiencies in the objects of assessment. Such weights, however, depend on the interaction between the object of assessment and the listener and cannot be generalized too far. This is the main factor limiting the development of a comprehensive system of general purpose parameters. A much easier task is to develop several independent weighting systems which can be used for particular applications. Such systems of weights which will be applied to several practical cases are currently under development.

6. SUMMARY

The purpose of this paper was to bring into focus some critical issues of timbre and sound quality judgments. The concepts, definitions, and discussions presented above were intended to make existing terminology less ambiguous and more practical. The foundation of the proposed system of concepts are the separate domains of sound character and sound quality. Both domains are described by the same set of constituent sensations but they are measured on different assessment scales. The object of assessment is an auditory image which in many practical cases is sufficiently described by multidimensional sensations of spaciousness and timbre. All basic components of both these sensations are combined together in a parametric system of sound assessment - MURAL. The future success of this system depends on the development of clear and univocal definitions of all component sensations and on the identification of reliable perceptual weights for selected applications.

7. REFERENCES

[1] Letowski, T. and Makowski, W. Properties of auditory images. Technika Radia TV, 1, pp. 16-21, 1977 (in Polish).

[2] McAdams, S. Spectral fusion, spectral parsing, and the formation of auditory images. (Stanford, CA: Stanford University, Department of Music, Report STAN-M-22, 1984).

[3] Heyser, R. C. Geometrical considerations of subjective audio. J. Audio Eng. Soc., 22, pp. 674-682, 1974.

[4] Letowski, T. Auditory assessment of signals and their sources. (Warsaw: Chopin Academy of Music, 1984) (in Polish).

[5] Gabrielsson, A. and Sjögren, H. Perceived sound quality of sound-reproducing systems. J. Acoust. Soc. Am., 65, pp. 1919-1933, 1979.

[6] Perkins, W. H. Vocal function. In: Handbook of speech pathology and audiology, L. E. Travis (ed.). (Englewood Cliffs, NJ: Prentice-Hall, 1971). p. 484.

[7] American National Standards Institute (ANSI), American national standard acoustical terminology. S1.1-1960 (1976).
[8] Plomp, R. Timbre as a multidimensional attribute of complex tones. In: Frequency analysis and periodicity detection in hearing, R. Plomp and G. Smoorenburg (eds.). (Leiden: Suithoff, 1970).
[9] Random House Dictionary of English Language (2nd edition). (New York, NY: Random House, 1988).
[10] White, D. W. The audio dictionary.(Seattle, WA: University of Washington Press, 1987).
[11] Schönberg, A. Harmonielehre. (Vienna: Universal Edition, 1922).
[12] Clark, M. and Milner, P. Dependence of timbre on the total loudness produced by musical instrument. J. Audio Eng. Soc., 12, pp. 29-33, 1964.
[13] Winckel, F. Music, sound, and sensation. (New York, NY: Dover, 1967).
[14] Tiffany, W. R. and Carrell, J. Phonetics: theory and application. (New York, NY: McGraw-Hill, 1977).
[15] Gabrielsson and Sjögren [5]
[16] Bladon, R. and Lindblom, B. Modeling the judgment of vowel quality differences. J. Acoust. Soc. Am., 69, pp. 1414-1422, 1981).
[17] Roederer, J. G. Introduction to the physics and psychophysics of music. (New York, NY: Springer-Verlag, 1974).
[18] Gabrielsson and Sjögren [5]
[19] Helmholtz, H. On the sensation of tone (second English edition). (London: Longmans and Co., 1885).
[20] Josephs, J. J. The physics of musical sound. (Prinston, NJ: Van Nostrand, 1967).
[21] Roederer [17]
[22] Rossing, T. The science of sound. (Reading, MA: Addison-Wesley, 1982).
[23] Random House Dictionary [9]
[24] Slawson, W. Sound color. (Berkeley, CA: University of California Press, 1985).
[25] Letowski [4]
[26] Letowski [4]
[27] Berglund, B., Berglund, U., and Lindvall, T. Scaling loudness, noisiness, and annoyance of community noises. J. Acoust. Soc. Am.,, 60, pp. 1119-1124, 1976.
[28] Stephani, O. and Bluthgen, B. Distortion during copying and mastering - an investigation. Studio Sound, 22, p. 60, 1980.
[29] Letowski [4]
[30] Jones, B. L. and McManus, P. R. Graphic scaling and qualitative terms. SMPTE Journal, 95 (11) 1986.
[31] Letowski and Makowski [1]
[32] Szlifirski, K. and Letowski, T. The system of basic criteria for sound quality assessment of sound recordings. Przeglad Techniki RTV, 1, 1981 (in Polish).
[33] Letowski [4]