QUALITY ASSESSMENT OF AUDITORY VIRTUAL ENVIRONMENTS

Renato S. Pellegrini

Studer Professional Audio AG
CH-8105 Regensdorf
Institute of Communication Acoustics
44780 Ruhr-Universität Bochum
Renato.pellegrini@studer.ch

ABSTRACT

Practical applications for auditory virtual environments (AVEs) are ever increasing. The achievable quality of AVEs has reached a level, where real-world problems can often be solved in a convenient way by using AVEs, meaning e.g. less expensive or more flexible than a real-world solution. Nowadays, the quality of AVEs is often measured in terms of their technical capacity to approximate the physical behavior of a real environment. As the design goals for AVEs shift from "reproducing the physical behavior of a real environment as accurate as possible" to "stimulating the desired perception directly" this comparative quality measure is no longer feasible. This paper describes parameters influencing the perceived quality of AVEs. Moreover, the dependence of perceived quality on the application is emphasized.

1 QUALITY

1.1 Introduction

This chapter defines a foundation for a theory of quality that seeks to define quality in real and virtual environments. The author likes to point out that quality measures for real and virtual environments are different. Knowing about the differences and naming them in a consistent way helps to prevent errors in the specification and design phase of an AVE. The main reason to use knowledge on human auditory perception in the design process of the AVE is to simplify the simulation in terms of processing cost and memory requirements without degrading the perceived quality. Therefore, meaningful quality measures are necessary in the specification phase already in order to assess the perceived quality for the user within a given implementation. Chapter 1.2 gives a general overview of definitions on quality-related terms. Towards a definition of quality in virtual environments Chapter 1.5 considers mono-modal perception as a special case and a definition of a plausible reproduction is derived in Chapter 1.7. Chapter 2 finally defines static and dynamic aspects of quality in AVEs.

1.2 Definition of Quality

According to [1] product quality should be defined as: "Result of an assessment of the perceived nature of an entity with respect to its desired nature", where *perceived nature* is defined as the "Totality of features of an entity. It signifies the identity of the entity as can be observed/detected by the perceiver". The *desired nature* is defined as the "Totality of features as projected by individual expectations and/or functional requirements and/or social demands". For any real-world or virtual-world product this definition implies that the user's expectation as well as the functional requirements for a specific task and maybe even social demands will influence the assigned quality. Further, only perceptible properties of the environment will be judged by a perception-based quality measure. Therefore, an instrumental quality measure based on physical criteria without using a model of human auditory perception might overestimate (e.g. auditory masking is not accounted for) or underestimate (perceptual relevance of error is higher) the influence of an error on the perceptual quality of the system. According to the German standard DIN55350: "Begriffe der Qualitätssicherung und Statistik" [2] quality is defined as „physical nature of an entity with regards to its ability to fulfill predetermined and fixed requirements[1]". This directly implies two facts: Firstly, the requirements need to be predetermined and fixed. Secondly, we can only define a set of requirements based on a given specific task. Given a specific task, the requirements for an AVE may often not be predetermined directly, since requirements for a perception to occur are often not yet known in a deterministic way in advance. A more practicable definition of quality would therefore include measures of *usability* [3] as outlined in Chapter 1.3. This dissertation focuses on product quality in terms of *usability* from a user point-of-view, while neglecting the influences and aspects of quality on marketing, manufacturing, services and time-to-market.

1.3 Usability

Usability, as described in [4] is a combination of *effectiveness, efficiency, and acceptance. Effectiveness,* according to ISO [3], signifies the accuracy and completeness with which specified users can achieve specified goals in particular environments. Effectiveness relates the goals of using the product to the accuracy and completeness with which these goals can be achieved. Common measures of effectiveness include percent task completion, frequency of errors, frequency of assists to the participant from the testers, and frequency of accesses to help or documentation by the participants during the tasks. It does not take into account how the goals were achieved, but only the extent to which they were achieved. *Efficiency* describes the resources expended in relation to the accuracy and completeness of goals achieved. Efficiency is generally assessed by the mean time taken to achieve the task. Efficiency may also relate to other resources (e.g. total cost of usage). A common measure of efficiency is time on task. Finally, acceptance describes a user's subjective level of satisfaction when using the product. Questionnaires to measure satisfaction and associated attitudes are commonly built using semantic differential scales. A variety of instruments are available for measuring user acceptance of software interactive products, and many companies create their own. Measures, such as satisfaction, usefulness or ease of use can be used to assess the acceptance of the product.

1.4 Presence

In most applications, subjects should experience presence (the sense of being in the virtual environment) to fulfill a specific task intuitively. An exact definition of presence is an issue of current research [5], [6], [7]. A practical definition of presence can be found in [8]. According to Schloerb physical presence designates "the existence of an object in some particular region of space and time". A person is objectively present in a remote environment where the person is not physically present, if there is some type of causal interaction between the person and the environment. This implies that a virtual environment can offer the sense of presence to a subject only if there is interaction involved. Further, physical presence supports subjective presence, consisting of the perception of being located in the same physical space in which a certain event occurs, a certain process takes place, or a certain person stands. Subjective presence is seen as a special case of objective presence, where the specified task is for a person to perceive that he or she is physically present in a given environment. The degree of subjective presence is defined to be the probability that a person perceives that he or she is physically present in the given environment. As of current understanding a similarity of the virtual environment's behavior to a conceivable real environment enhances the sense of presence. In [9] Hendrix and Barfield have shown that the addition of spatialized sound significantly increased the sense of presence. Humans are trained during their whole life in real environments to interpret the available information of all senses to form a single world model without contradictions. The interactive reaction of the environment's behavior to interaction of the listener or third-party input needs to correspond to the listener's expectations, which are based on his life-long experience. This leads to a conceivable world model that is fundamental for subjective presence. It is easily believed that a high degree of subjective presence that corresponds to the user's expectation will have a positive influence on the *usability* of a system, especially on its *effectiveness* and *acceptance* since in a conceivable environment, the user can react intuitively. Experiments comparing the degree of *presence* and *comfort* (a measure of acceptance and satisfaction) have shown that a high degree of presence can contradict a high degree of comfort under certain circumstances [11]. Background noises may be important to the sense of presence as stated in [6], [11], and [16]. In [10] Slater et al. argue that the effectiveness of a virtual display in conveying a sense of presence to the observer might be dependent on the observer's preferred mode of interacting with the environment. Based on the concept of "neurolinguistic programming" employed by some clinical psychologists, they postulate that individuals have a preferred mode of conceptualizing their interactions with the world: visual, auditory, or kinesthetic. According to Ramsdell [11], the auditory modality provides information to the observer on three different levels: the social level (speech and music perception), the warning level and the primitive level. The social level is responsible for all symbolic information like e.g. language and music. The warning level is based on the information carried by sounds with respect to their signaling or warning significance. The primitive level is the least intuitively obvious. On this level, sound serves as neither symbol nor warning, but as the auditory background to everyday life. The acoustic information at this level consists of incidental sounds made by objects in the environment, and by ourselves as we interact with objects in the environment. Ramsdell states that although we do not typically maintain conscious awareness of background sounds, "these incidental noises maintain our feeling of being part of a living world and contribute to our own sense of being live".[1]

1.5 Multi-Modal versus Mono-Modal Perception

[1] This part was written based on the musings of Gilkey [6] on Ramsdell's findings in [11].

Perception-based quality is multi-modal. Perception is always strongly related to selection. Only a small amount of information available at any given moment in time is actually used to form an auditory percept. Which parts of the available information are used strongly depends on the actual state-of-mind of the listener. It is important to keep in mind that perceived auditory quality might strongly depend on cross-modal cues such as e.g. visual input. In 0 Hollier and Voelcker have shown that the visual quality of a TV set strongly depends on it's audio quality and vice versa. Similarly, Beerends and Caluwe [13] prove that there is significant mutual influence between audio and video quality of an audio-visual stimulus. Interestingly, audio quality is affecting perceived video quality to a much lower extent than video quality is influencing audio quality. While in real environments the cross-modal cues are related to each other in a meaningful way, virtual environments have to assure that all modalities relevant to a desired perception are fed consistently. Inconsistencies within different modalities can worsen the overall quality as well as the perceived quality of specific quality features. A good overview on different auditory and non-auditory factors that potentially influence the perception of AVEs can be found in [14].

Layers of quality	***Physical elements*** Accuracy of simulation in technical terms	➢ ***Temporal resolution*** ➢ ***Frequency resolution and bandwidth*** ➢ ***Spatial resolution*** ➢ ***Dynamic behavior*** o *Latency* o *Update rate* o *Temporal and spatial dynamic resolution*
	Psychoacoustic features Quality features which originate from the way the simulated environment is perceived by the user	➢ ***Perceived loudness*** ➢ ***Location accuracy*** ➢ ***Sound timbre*** ➢ ***Auditory spaciousness*** ➢ ***Source size*** ➢ ***Dynamic accuracy*** o *Responsiveness* o *Smoothness* o *Steadiness* ***(depending on the task, others may be relevant)***
	Psychological factors Perceived quality of features which are strongly related to cognition, action and the emotional state of the user.	➢ ***Source-dependent expectation*** o *How should the recording sound like* ➢ ***Task-dependent expectation, expectation on interaction*** o *Guided interactivity (interaction needed for a dynamic task).* o *Unguided interactivity (interaction helps perception of an originally static feature)* ➢ ***Personal expectation*** o *Motivation* o *Personal attitude* o *Taste, aversions*

1.6 Assessing Quality / Layers of Quality

Quality can be measured on different layers. Table 1 shows the different layers and their corresponding quality aspects. While physical elements are accessible through instrumental measurements, psychoacoustical features can only be accessed by measurement techniques based on auditory perception. These may again involve instrumental measurement methods involving computational simulations of auditory perception such as e.g. in PEAQ [15]. Finally, psychological factors may strongly influence perception and bias results between individuals by cognition and emotion. An important part of cognition is depending on interaction and is therefore influenced by the action of the subject.

1.7 Authentic versus Plausible Reproduction

Authenticity describes the property of two entities to be indistinguishable to a human observer. An authentic reproduction of a real environment using an AVE would be indistinguishable from the real environment in its ideal form. The concept of authenticity does not include any restrictions on the action or state-of-mind of the user, and therefore the reasons for the decision, whether a reproduction is authentic or not, are hidden. In any case an AVE will never be able to stimulate an authentic perception as a given real environment, because of the following reasons:

1. The currently available interfacing technology to human senses are bulky and of low quality for most senses, compared to the capabilities of humans' perception.

2. Approximations have to be used in any AVE for a simplified simulation requiring an affordable amount of processing power and memory.

3. Cross-modal interaction is not yet well understood and is known to strongly influence perception.

At least for applications where the auditory part of perception is most prominent, usable interfaces have been developed which cover most of humans' perception-capabilities. The second step, namely approximation, may include knowledge on audible and non-audible parts of the environment using just-noticeable differences (JND) for worst-case situations. Based on this knowledge it is possible to define minimal resolutions in the frequency and the time domain as well as for the dynamic behavior of the system. The resulting copy of the environment can still be seen as an attempt to achieve perceptual authenticity. Nevertheless, without further knowledge on the user's state-of-mind and the action involved within an unspecified application, authenticity can never be guaranteed. Because not all audible features are mutually exclusive, orthogonal, and linear independent and per se inaudible artifacts may still lead to an audible degradation of the overall percept. Therefore a less restrictive definition of the main goal of reproductions is wanted. *Plausibility* is a suitable concept that defines a set of demanded quality features for a given specific application. The reproduction of an environment, e.g. using an AVE, should include all relevant features for a given application. Since the action is defined within the plausibility concept, the relevant quality features for the specific application can be determined in auditory tests and the quality of the specific psychoacoustic features can be assessed independently. Even dependencies of orthogonal quality features can be assessed for the given application, and usability tests can assure a high effectiveness of the AVE. Since the application is known, the AVE can be further simplified, to the extent that only relevant quality features need to be reproduced. Often it can be shown that these additional simplifications are substantial with respect to the savings in processing cost and memory requirements as shown in [16]. The goal for a reproduction using the *plausibility* concept rather than the *authenticity* concept has shifted from "copying an existing environment in all it's physical aspects" to "a suitable reproduction of all required quality features for a given specific application". Wherever we know which quality elements lead to the relevant quality features, we can optimize the simulation of these quality elements, while disregarding irrelevant quality elements. Because the relevant quality features will vary with the application, a specific application needs to be clearly defined before optimizations in terms of a plausible environment for a given action can be found. The resulting AVE will be clearly distinguishable in most cases from its real counterpart, but since the design goal has shifted to a usable environment rather than an indistinguishable copy of the real room, the utility due to efficiency and added value will have been optimized to serve the user's needs for the specified action (see Figure 1).

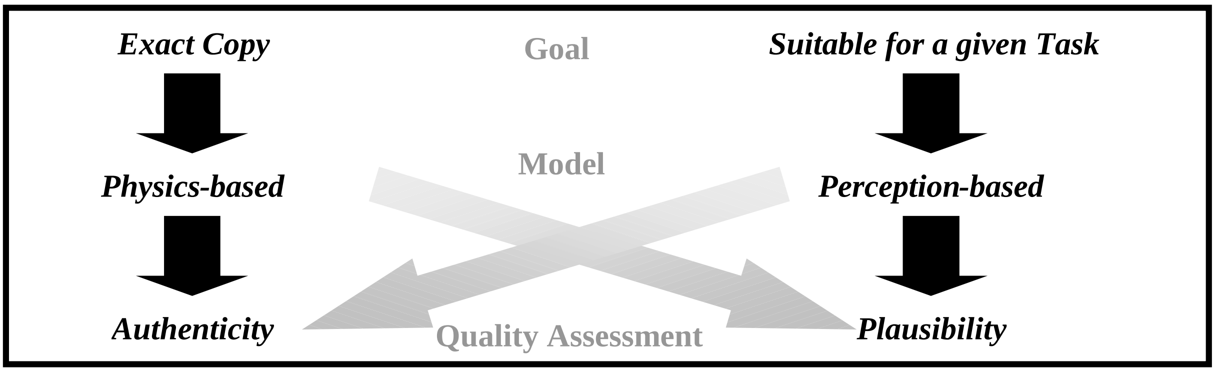

Figure 1: Authentic versus plausible reproduction

2 DEFINING QUALITY FOR AUDITORY VIRTUAL ENVIRONMENTS

Auditory virtual environments aim at simulating an environment for at least the auditory part of perception. AVEs are used to serve real-world applications in an increasing number of different fields.

Defining quality accurately for real and virtual environments is crucial to be able to compare real-world solutions and their virtual replacements. Products based on virtual environments are especially interesting in fields, where real-world-based products are too expensive or not flexible enough to serve a given application. Early implementations of AVEs aimed at reproducing the physical behavior of a desired real environment. Corresponding quality measures assessed the accuracy of the simulation by rating the accuracy of the physical modeling, based for example on the number of reflections accounted for in a room simulation. Since all AVEs are built using simplifications and approximations to the exact physical behavior of sound propagation in a given space, the errors introduced by these approximations may adversely influence human perception. To consider the *added value* (quality with respect to their cost) of using an auditory virtual environment rather than a real-world solution, the AVE's perceived quality needs to be assessed. As has been shown in Chapter 1.7 only quality measures based on plausibility are capable of assessing the influence of these approximations on the utility of the product. According to the *plausibility* concept, auditory virtual environments aim at "stimulating a desired perception" for a given task, that is, human perception rather than the physics of an environment are used as the basis for the design criteria. The main reason to use knowledge on human auditory perception in the design process of the AVE is to simplify the simulation in terms of processing cost and memory requirements without degrading the perceived quality. Therefore, meaningful quality measures are necessary in the specification phase already in order to assess the perceived quality of the user for a given implementation. For an AVE the definition of quality given in Chapter 2.1 implies that the user's expectation as well as the functional requirements for a specific task and maybe even social demands will influence the assigned quality of the AVE. Further, only perceptible properties of the AVE will be judged by a perception-based quality measure, whereas a quality measure based on the accuracy of an AVE without using a model of human auditory perception might overestimate (e.g. auditory masking is not accounted for) or underestimate (perceptual relevance of error is higher) the influence of an error on the perceptual quality of the system. Given a specific task, the requirements for an AVE may often not be predetermined directly, since requirements for a perception to occur are often not known in a deterministic way in advance. A more practicable definition of quality would therefore include measures of usability [3] as given in 1.3.

2.1 Layers of Quality of Auditory Virtual Environments

The definitions of the layers of quality for AVEs are very similar to the general layers of quality defined in Chapter 1.6. But there are in general two complexes of problems, for which special care has to be taken in AVEs. Firstly, if – as in most cases – the virtual environment controls not all modalities, the remaining modalities may influence the percept in quality assessments. Secondly, dynamic aspects of virtual environments may or may not conform to the user's expectation that leads to completely different quality assessments in static and dynamic environments. It is generally believed that in almost all circumstances people feel present in a real environment as long as all senses are stimulated uniformly [17]. However, in [6] Gilkey has reported that auditory cues are a crucial determinant of the sense of presence. The study using adventitiously deafened individuals describe a sense of unconnectedness with their surroundings. In contrast to real environments, it is not imperative for virtual environments to convey a sense of presence. But, as outlined in Chapter 1.4, conveying a sense of presence has a favorable effect on most applications. Therefore, the similarity of the virtual environment's behavior to a conceivable real environment is desirable, since this is believed to enhance the sense of presence. Further, cross-modal influences are known to impact perceived sound quality, as was shown in Chapter 1.5. In AVEs the auditory perception can easily be decoupled from the other modalities. As a consequence, a required additional condition for a mono-modal AVE is that the prominent quality features are to be based on auditory perception only. Additionally, it has to be assured for all other modalities that they do not adversely influence the mono-modal auditory percept. Especially visual cues are known to have a precedent effect on the auditory percept for ambivalent cross-modal information [19]. Therefore, when assessing psychoacoustic features within an AVE, the auditory test setup needs to be designed with care to minimize unwanted cross-modal influences. For dynamic and thus interactive virtual environments, the rules defined in Chapter 2.3 need to be followed to meet the requirements on system latency, on the update-rate and on the resolution. Well-designed AVEs will give similar results in their perceived quality compared to a real counterpart [18].

2.2 Static Aspects of Quality in Auditory Virtual Environments

As was shown in Chapter 1.7 the usability concept is a favorable quality measure, whenever the quality relating to a plausible simulation is to be assessed. To assess the usability of an AVE it is extremely important to control the state-of-mind of the user. Ideally, the state-of-mind of the user should resemble a typical working situation. The most obvious way to do this is to let the user actually act in a similar way as in the real application by letting him perform the task, for which the AVE was designed. By actually performing the task, quality can be tested either directly, by letting the user describe the quality, or indirectly, by measuring the user's performance in executing the given task. In the first case, the user will report on all three levels of quality defined in Chapter 1.6. In the later case *effectiveness* and *efficiency* might be assessable while the *satisfaction* part of *usability* will be mostly hidden. Whenever a real environment corresponding to the virtual environment exists, a comparison between the performance of the user in the real and the virtual environment can be made. This offers very direct and valuable information on the differences between the real and the virtual environment for one individual subject, but is not easily transformable to absolute ratings of the virtual environment. Since in this test setup a real multi-modal environment and a virtual (possibly mono-modal) environment are compared, care has to be taken on what influence the knowledge on the other modalities might have on the test results. For instance, knowing the exact visual representation of the real room may enhance the auditory perception of the virtual room, especially when only the auditory part is simulated (compare 0, [13]). Further, as stated in [20], performance measures are difficult to evaluate in general. The individual differences between subjects are typically much larger than the measurable differences between the environments. Similarly, differences correlated with the learning curve of subjects when repeating the same task in different environments may also be quite large. Therefore, it is difficult to gain statistically significant, objective data. Whenever the task an AVE is designed for is too versatile it is a good idea to directly prompt the user to assess different aspects of the AVE that are known to be important for the task in separate tests.

2.3 Dynamic Aspects of Quality in Auditory Virtual Environments

Auditory virtual environments are considered dynamic, whenever the properties of the source, the environment, or the listener may change over time. These changes may either result from an external input, or from interaction with the user. It is very important for the quality of the AVEs to assure that all dynamic changes are tested according to the requirements of a given application. Whereas a real environment's reaction to an action introduced by the user is always instantaneous[1], a delay between action and corresponding reactions exist in any virtual environment. The application defines the maximal allowable delay for a reaction. Testing the AVE in a slower changing manner won't guarantee its usefulness for the given application, testing it with a higher changeability will most likely affect the perceived quality of the AVE, although, based on the given application, the AVE's quality would not be affected. From a technical point-of-view the latency, the update rate and the resolution of dynamic changes are influencing the dynamic quality of the system. Latency, update rate and resolution granularity all are sources of delays that are not present in any real environment.

In any real environment, changes based on interaction are always instantaneous. This can not be achieved in virtual environments, because the information on what interaction is requested needs to be gathered first, then converted to the digital domain usually involving a "sample-and-hold" stage. Afterwards, the digital input needs to be processed and the rendering stage needs to change the output. The update rate at which new inputs are read defines the absolute accuracy, whereas the update rate at which the output is driven defines the maximum relative accuracy. For instance, head-tracking information could be read at a higher update rate, then lowpass-filtered to enhance the reliability of the movement trajectories. Finally the output update rate would be based on the required accuracy for the affected properties. Defining a "just to be noticeable difference" e.g. a minimum movement angle for spatial attributes can do this. The output update rate must never be higher than the input update rate in any case. A designer of a virtual environment has to make sure that the delays introduced by the virtual-environment system are either below just-noticeable-differences for human perception, or – if not achievable – are below an annoying threshold which may be task dependent. Since the data-flow and processing cannot be done instantaneously, the speed and repetition rate of interactive inputs are becoming a quality factor in dynamic systems. It is extremely important to make sure that the user interacts with the system in a similar way he would interact when performing the specified task. For example, whenever the user is supposed to move his head from time to time, the test setup should allow for head-movements too. Since these head movements take time, the test setup needs to offer enough time per stimulus to allow the user to move the head before assessing the AVE's performance. The speed in which these head movements are performed is task dependent which again implies the importance of specifying the task for quality measurement in virtual environments. As shown in Table 2 perceptual dynamic aspects are closely related to the physical dynamic aspects of virtual environments. From a perceptual point-of-view dynamic aspects can be subdivided into smoothness, responsiveness and steadiness. Smoothness corresponds to the update rate of the system. An AVE with a high smoothness will assure that small transitions between different states of the AVE are inaudible. Responsiveness is directly coupled to the system latency. Latency is either perceptible as synchronization difference between different subparts of the AVE (e.g. visual and auditory information) or as interaction-response time, meaning the time between a change of a property of the AVE and the time where these changes may be perceived. Finally, steadiness strongly relates to resolution. Steady parts of a percept should not change, whenever a dynamic change is introduced to the AVE. Since a dynamic AVE typically is only dynamic in discrete steps (in space and time) the resolution either has to be higher than the least audible change (just-noticeable-difference JND), or interpolation techniques need to be applied. For example, rotating the head should not shift the apparent direction-of-incidence of the auditory percept of a

[1] although it might take some time until the changes in the environment become audible at the listener's end, the starting point of the change is not delayed.

<table>
<tr><th>Physical layer</th><th>Perceptual layer</th></tr>
<tr><td>System latency

The time delay between a change of a property of a source, the environment or the user until all changes are processed and can be perceived at the output.

1. End-to-end system latency.

2. Cross-modal synchronicity.</td><td>Responsiveness

The perceived time delay between an action or an event is initiated and the according change at the output can be perceived. Or the perceived synchronicity of different modalities.</td></tr>
<tr><td>Update rate

The frequency in which input changes are processed.

1. Frequency, at which inputs are processed (absolute accuracy).

2. Frequency, at which the output can be changed (relative accuracy).</td><td>Smoothness

At low speeds a high smoothness of a virtual environment system assures that small transitions at the input will result in a very small or imperceptible output change.

At high speeds a high smoothness assures large changes to be perceived as a smooth transition rather than a train of several steps at the output.</td></tr>
<tr><td>Resolution

The temporal and spatial resolution of possible output changes.

1. Resolution, with which inputs are processed.

2. Interpolation resolution at the output.</td><td>Steadiness

A high steadiness implies that all steady parts with respect to a changed property are not supposed to affect the perception of the output.</td></tr>
</table>

Table 2: Physical and perceptual dynamic properties of virtual environments

fixed sound source in the virtual environment. If the system is not fast enough (meaning either latency or update rate is too low), the apparent direction will shift over time, if the AVE's resolution is too small, the source will jitter around the accurate position.

3 REFERENCES

[1] Jekosch, U. (2000). *"Basic Concepts and Terms of "Quality", Reconsidered in the Context of Product-Sound Quality."*, *t*o be published in Acustica.

[2] DIN55350 (1995b). *„Begriffe der Qualitätssicherung und Statistik"*, Deutsches Institut für Normung.

[3] ISO 9241-11 (1998). *"Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability*", International Standards Organization.

[4] Gleiss, N. (1992). *„Usability – Concepts and Evaluation"*, Tele 2/1992.

[5] Begault D. R. (1994), "*3-D Sound for Virtual Reality and Multimedia"*, Academic Press Inc., London.

[6] Gilkey R., Weisenberger J. (1995). *„The Sense of Presence for the Suddenly Deafened Adult: Implications for Virtual Environments"*, Presence Vol. 4, No. 4, pp. 357-363.

[7] Sheridan, T.B. (1992): *"Defining our Terms"*, Presence No. 1, pages 272-274.

[8] Schloerb, D.W. (1995). *"A quantitative measure of telepresence",* Presence: Teleoperators and Virtual Environments, 4, 64-80.

[9] Hendrix, C., Barfield, W. (1996). *"The sense of presence within auditory virtual environments",* Presence No. 3, pp. 290-301.

[10] Slater M., Usoh M., Steed A. (1994). *"Depth of presence in virtual environments"*, Vol. 3, pp. 133-144.

[11] Ramsdell, D.A. (1978). *"The psychology of the hard-of-hearing and the deafened adult"*, In H. Davis & S. R. Silverman (Eds.), *"Hearing and deafness"* (4^{th} ed., pp. 499-510), Holt, Rinehart, & Winston, New York.

[12] Hollier M., Voelcker R. (1997). *„Objective performance assessment: Video quality as an influence on audio perception"*, 103^{rd} Convention of the Audio Engineering Society, New York, September 1997, Preprint No. 4590.

[13] Beerends J., Caluwe F. (1999). *"The Influence of Video Quality on Perceived Audio Quality and Vice Versa"*, JAES No. 5 pp. 355-364 Vol. 47.

[14] Begault D. R. (1998). *"Auditory and non-auditory factors that potentially influence virtual acoustic imagery"*, 16^{th} international conference on spatial sound reproduction of the AES, Rovaniemi, Finland.

[15] ITU-R Recommendation BS-1387 PEAQ (1999). *" Method for objective measurements of perceived audio quality* ", International Telecommunications Union, Geneva.

[16] Pellegrini R. S. (2001). *„A virtual reference listening room as an application of auditory virtual environments"*, Ph. D. Thesis, institute of Communication acoustics, Ruhr-University of Bochum, Germany.

[17] Sheridan T. (1996). *"Further Musings on the Psychophysics of Presence"*, PresenceVol. 2, No. 5.

[18] Djelani, T., Pörschmann, C., Sahrhage J., Blauert, J. (2000). *"An interactive virtual-environment generator for psychoacoustic research II"*: Collection of head-related impulse responses and evaluation of auditory localization, ACUSTICA / Acta Acustica (in press).

[19] Boff K.R. (Chapter by Welch B., Warren D.) (1986). *"Handbook of perception and human performance: Intersensory Interactions"*, Chapter 25, pp. 1-32, Wiley, New York.

[20] ITU-T (1992): Handbook on Telephonemetry. Genève, Switzerland.