"The original publication is available at http://www.aes.org/journal/"
Altinsoy, M.E. (2012). „The Quality of Auditory-Tactile Virtual Environments," Journal of the Audio Engineering Society, Vol. 60, No. 1/2, pp. 38-46, Jan.-Feb. 2012.

The Quality of Auditory-Tactile Virtual Environments

M. Ercan Altinsoy, *AES Member*
ercan.altinsoy@tu-dresden.de

Dresden University of Technology, Chair of Communication Acoustics, Dresden, Germany

In our daily lives, we usually perceive an event via more than one sensory modality (e.g., vision, hearing, touch). Therefore, multimodal integration and interactions play an important role when we use objects and for event recognition in our environment. A virtual environment (VE) is a computer simulation of a realistic-looking and interactive world [1]. VEs should take into account the multisensory nature of humans and communicate with the user not only through vision but also through other modalities. In addition to vision, hearing and touch are the most commonly used communication channels. Recently, a variety of products with additional tactile input and output capabilities have been developed (e.g., Apple iPhone and other touch-screen devices, Nintendo Wii, etc.). Some of these devices provide new possibilities for interacting with a computer, including the auditory modality. Binaural synthesis and rendering are becoming key technologies for multimedia products. Virtual environments are no longer limited to academic research; they have commercial applications, particularly in medicine, game, and entertainment industries. Thus, the quality of VEs is becoming increasingly important. User interaction with a VE is a key issue in the perception of its quality. Several studies have discussed the quality of displays, input and output devices (for different modalities) as well as software and hardware issues; however, multimodal user interaction should also be examined. This paper focuses on the parameters that influence the quality of audio-tactile VEs.

0 INTRODUCTION

Functionality and safety are the main quality issues of virtual environments. Without a doubt, they should be designed to satisfy the user. According to the German standard DIN 55350, quality is defined as the "physical nature of an entity with regards to its ability to fulfill predetermined and fixed requirements" [2]. Jekosch extended this definition: *"Quality is a descriptor of the adequacy of the perceived characteristics of an entity with regard to required features…The required features are formed by the totality of the features of individual expectancies and/or social requirements and/or proper demands"* [3]. Taking into account these definitions, the quality assessment of VEs may be based on user expectation of the VE and the functional requirements of the specific application or task. The layers of quality, which have previously been described [4, 5], are also valid for assessing VEs. Quality judgments are based on physical (i.e., elements) and perceptual (i.e., features) layers [6]. Whilst quality elements are the building blocks for engineering the quality of an entity, quality features are components resulting from an analysis of the perceived nature of the entity [2].

1 QUALITY ASPECTS

Fig. 1 shows the principle architecture of an auditory-tactile VE [7]. The classic components include the input devices, virtual reality (VR) engine, auditory and tactile drivers, and output devices. The physical properties of the tracking systems include the lag (overall latency), update rate, interference (sensitivity to environmental factors, e.g., lighting conditions, magnetic noise, etc.) as well as the accuracy, resolution, and working space. The VR engine creates the virtual world and contains auditory and tactile renderers. Auditory and tactile renderers use the same world representation. From a quality point of view, this architecture can be divided into hardware components (e.g., input/output devices), software (e.g., VR engine), and databases [1].

Fig. 2 shows the detailed architecture of the auditory portion [8]. The position of the user's head is required as input information for the auditory virtual environments (AVE). Sound source and sound fields are two architecture modules. Sound-source signals can be obtained through recording or synthesis. A detailed overview of sound-source synthesis has previously been documented [9, 10]. The behavior of the sound field in which the user and sounds stay should be either physics- or perception-based. A complex model might result in a more authentic reproduction but result in longer computer processing times (i.e., reproduction delay). Perception-based models are derived from psychoacoustical investigations (for an overview, see [11]). Another important part of auditory rendering is the reproduction-based renderer (e.g., Wave-field synthesis, Ambisonics, etc). Depending on the reproduction technique, the user's head-related transfer functions may be necessary.

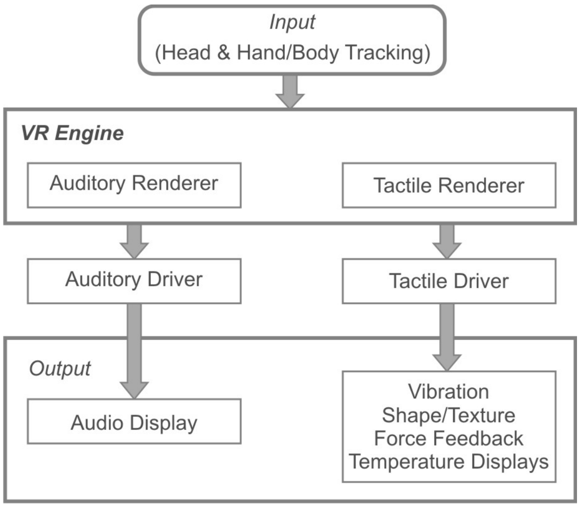

Fig. 1. The principle architecture of an auditory-tactile VE.

The sound-source signals, reflection and directivity filters, filter parameters of the HRTFs, and algorithms of the reproduction renderer are the input parameters of the signal-processing module.

The quality elements of auditory VEs include its frequency resolution, bandwidth (see [11]), spatial and temporal resolution as well as its dynamic behavior. Alternatively, its HRTFs, number of mirror sources, position of mirror sources, reflection filter, and late reverb tail generation are also quality elements of auditory VEs [12]. Quality features include its loudness, auditory spaciousness, timbre, localization accuracy, reverberation, dynamic accuracy, and artifacts [12]. Numerous elicitation experiments have been conducted to define the quality attributes of multichannel audio (e.g., reproduction and rendering) that are a part of the auditory VE [13, 14, 15, and 16]. These attributes also include localization, source width, envelopment, source distance and depth, space perception, and naturalness [17]. Previous research developed a model that contained these attributes to predict multichannel audio quality [18].

Fig. 3 shows the modules of the tactile VE and its data-flow paths. The required input variables are the position and orientation of the user's hand and fingers, their applied force, temperature and, in some applications, their body's position and orientation. Similar to auditory VEs, tactile interactions can be modeled physically or perceptually. Physical models are primarily based on Newtonian law [18]. Depending on the interaction, various components of physical modeling (e.g., collision detection, surface deformation, grasping, texture, gravity, and friction) determine the dynamic behavior of virtual objects and tactile feedback [18]. The measured forces, accelerations, positions, and temperature are sent to the interface controller that drives the tactile interfaces. Tactile interfaces can be categorized based on the type of feedback (e.g., force feedback, vibratory feedback [finger or whole-body], texture feedback, and temperature feedback).

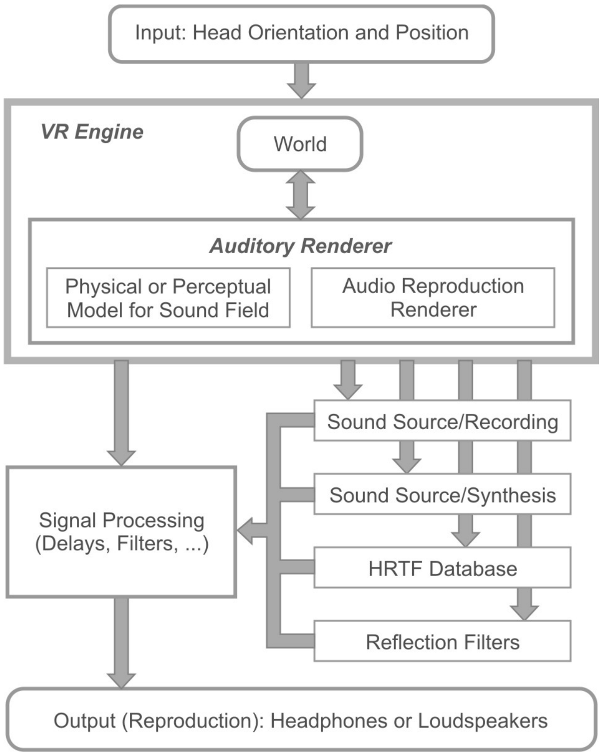

Fig. 2. The architecture of the auditory portion of the VE.

The criteria for a good force-feedback interface have previously been defined (see [19]):

• The interface should be extremely light and strong and have negligible inertia, friction and vibration.

• The interface should simulate a hard stop (e.g., a wall).

• The interface should simulate Coulomb friction without sponginess or jitter.

• The interface should simulate a mechanical centering detent with crisp transitions and no lag.

Similar criteria can be defined for a tactile interface that simulates surface textures:

• The interface should be able to mimic sandpaper from a grit range between 40 (coarse) and 1000 (super fine).

• The interface should be able to mimic smooth surfaces (e.g., glass or laminated wood).

• The interface should be able to mimic grooved surfaces in all possible groove width dimensions from daily life.

The quality elements that affect the quality of hand vibration, whole-body vibration, and texture reproduction include the interface's system linearity, flat frequency response, crosstalk between axes, and harmonic-vibration attenuation component. Although strong crosstalk between axes degrades quality, in some cases, weak crosstalk increases feedback plausibility [20].

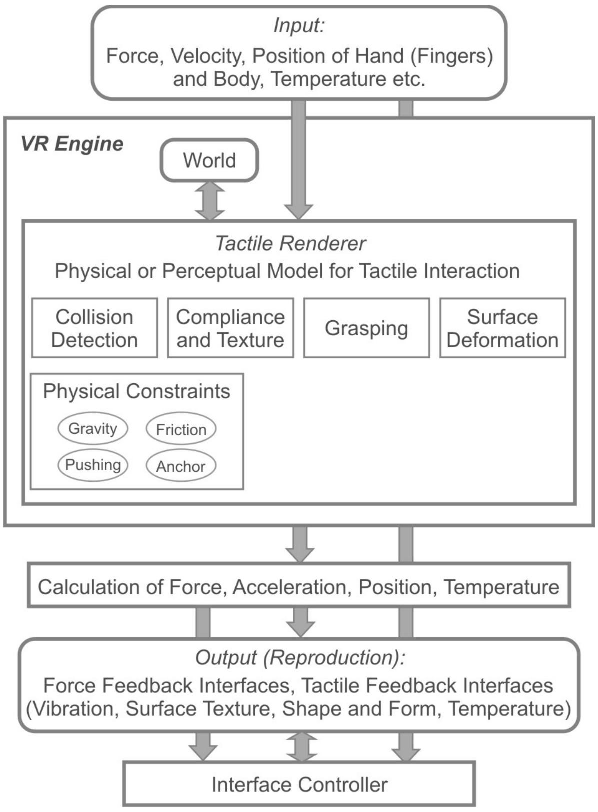

Fig. 3. The architecture of the tactile portion of the VE.

To summarize, important physical properties of almost all tactile interfaces, which have strong influence on the feedback quality, are the bandwidth of the device, the linear frequency response, latency, resolution, maximum feedback amplitude, power-to-weight ratio, and power-to-volume ratio. An optimum interface should

- match (or possibly exceed) human sensory and control capabilities;
- be linear and have a flat frequency response;
- have a low latency;
- have a just-noticeable-difference adequate resolution;
- be negligibly light and sufficiently rigid; and
- be able to simulate sufficient output.

Most of the studies that have measured task performance evaluated a tactile interface or system (for examples, see [21, 22, and 23]). Task completion time and error rate are two criteria used to evaluate quality. Additional research methods include recognizing an object or a texture, directly comparing simulated tactile feedback with the real-world tactile feedback of commonplace objects (for examples, see [24 and 25]).

2 QUALITY LAYERS OF AUDITORY-TACTILE VIRTUAL ENVIRONMENTS

The previous section presented the architecture of auditory-tactile VEs and their quality elements and features. This section discusses the quality layers of the entire auditory-tactile VE.

User activities cause real time changes in the VE. In other words, the users communicate with the VE during their active exploration. Users' expectations, experiences, motivations, memories, emotions, and attitudes, as well as their familiarity with the environment, are user dependent factors in respect to quality evaluation of the auditory-tactile VE (Fig. 7). Quality evaluations of telecommunication services are formed likewise [26]. The information exchange between the user and the VE should be optimally designed; this point is true for both original designs and generic reproductions.

Depending on the VE application, the references of the user vary for quality assessment. Most auditory-tactile VEs seek to reproduce the physical behaviors within real-world environments (e.g., driving simulators or virtual tennis games). Vibration produces sound in the real world. Therefore, sound and vibration are coupled to each other. The experiences, which we did in our childhood and play a role in the rest of our life, are based on this physical relationship. When people interact with an object in a VE, they expect feedback. Feedback is a type of communication with the VE. We expect to integrate the multisensory stimuli, which are generated by a multimodal event in the VE. Temporal asynchrony between the auditory and tactile events or spatial origin disparity can result in the segregation of the auditory and tactile events into two isolated percepts for each modality, instead of a unified multimodal percept. The perceptual consequence of the segregation and the disappointment of our expectation is the quality degradation. A practical example from the audio-visual domain can be found in the broadcasting applications, where the presence of detectable audio-video temporal asynchronies results in a reduction of quality [27]. Each of us experienced this problem while watching TV at least once in life. A virtual tennis game is a good example of an auditory-tactile VE. When the user hits a "tennis ball" with his or her "racket", he or she wants to hear and feel feedback from the virtual contact of the ball and racket. Some physical conditions, which can cause a perceptual segregation of auditory and tactile events, are the delay between the hitting event and the sound, the changes of the sound parameters (e.g., loudness, timbre, etc.) based on hitting parameters (e.g., velocity, force, etc.) and the location of the contact and sound. The perceptual segregation threshold values provide important hints for the effective design of VEs without causing quality degradation.

Synchronization/Perceived Synchrony

Over time, people learn that one physical event often generates multiple sensory stimuli (auditory, visual, tactile, etc.). The temporal correlations of these stimuli help the brain integrate them with each other and differentiate them from stimuli that are unrelated. Synchronizing the different modalities of multimedia applications is difficult. Technical issues such as data transfer time, computer-processing time, and the delays that occur during the feedback generation process constrain synchronization. As the asynchrony between different modalities increases, users' sense of presence and realism will decrease.

In multimodal VEs, each unimodal information can be delayed with respect to user action. For example, in auditory-tactile VEs, both auditory and tactile feedback can be delayed with respect to the action, when both information are delayed by the same amount of time, auditory and tactile events still are synchronous. A multimodal VE system latency is the time that elapses between unimodal feedbacks (e.g., auditory-visual, auditory-tactile or visual-tactile). If a user hits an object with his or her hand, the central controller should receive information (e.g., applied velocity, event location, listener head's location, etc.) related to the hitting event and transmit this information to the auditory, tactile, and visual renderers. Each renderer makes the required calculations and then the tactile renderer transmits the force-feedback information to the tactile actuator, the auditory renderer sends sound data to the loudspeakers or the headphones, and the visual renderer transmits the data to a head-mounted display or a projection screen.

Fig. 4 shows the latencies that are important to the design of VE generators. The time t_{act} is the moment at which the action occurs; t_{vis} is the arrival time of visual information; t_{aud} is the arrival time of auditory information; and t_{tac} is the arrival time of tactile information. L1, L2, and L3 are the latencies for the each unimodal subsystems: visual, auditory, and tactile, respectively. L4, L5, and L6 are the latencies between modalities: visual-auditory, auditory-tactile, and visual-tactile, respectively. An approximate latency for an auditory-tactile VE can be estimated to be between 20 ms and 40 ms.

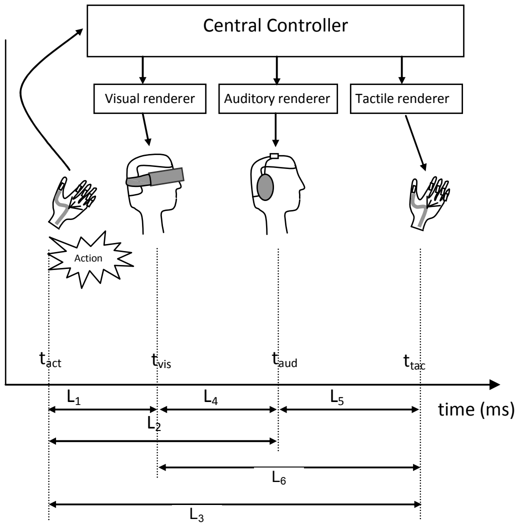

Fig. 4. VE system latencies.

Several studies have discussed the perceived simultaneity of multimodal stimuli. Previous research defined a multimodal synchronization threshold as the maximum tolerable temporal separation between the onsets of two types of sensory stimuli, such that the accompanying sensory objects are perceived as synchronous [28]. Different psychophysical measurements have attempted to measure this threshold, (for a detailed review, see [9 and 30]). The obtained results vary depending on the kinds of stimuli and the psychometric methods employed.

Altinsoy reported perceptual threshold values for auditory-tactile asynchronies [28]. The tactile stimulus was a sine wave presented at the tip of a participant's index finger via a mini-electrodynamic shaker (passive tactile stimulation) located inside a wooden box that contained a circular hole on which the participants placed their index finger. The shaker delivered the stimuli to the skin via a 4-mm vibrating probe. The auditory stimulus was a burst of white noise. We randomly presented audio-tactile stimuli with an audio delay between -150 ms and 150 ms with varying step sizes to participants. Negative delay values indicate that the auditory stimulus was presented first; positive delay values indicate that the tactile stimulus was presented first. We presented each condition 12 times. Participants reported whether the audio signal and the tactile signal were synchronous or asynchronous. The proportions of the synchronous responses are shown in Fig. 5. We obtained a psychophysical model by fitting ogive results using a Gaussian fit with an exponential background. The perceptual threshold values were 50 ms for the audio lag and the 25 ms for audio lead. The results indicated that the synchronization between auditory and tactile modalities must be at least within 25 ms. Thus, the auditory-tactile delay is even more critical than the auditory-visual delay. People actively touch objects in most commonplace situations (e.g., playing a piano or typing with a keyboard), causing auditory and tactile feedback. Adelstein et al. [29] and Levitin et al. [30] investigated the perceptual asynchrony threshold values for an active tactile interaction situation (e.g., playing a drum). The former group found a threshold value of 42 ms [29], whereas the latter group found that the thresholds varied between 18 ms and 31 ms depending on the stimulus duration [30]. Differences between the psychophysical measurements of these studies possibly caused their results to diverge. However, all these studies have reported that many participants have low threshold values (approximately 10 ms [28, 29, and 30]). In particular, musicians have smaller thresholds than the general population, possibly because of their training [9]. Therefore we suggest the synchronization requirement of 10 ms for auditory-tactile virtual environments.

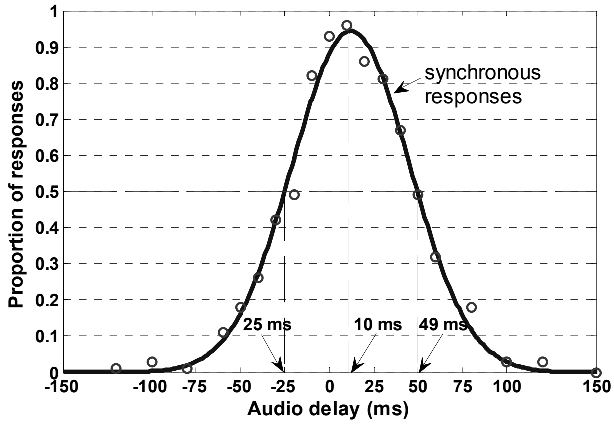

Fig. 5. The proportions of the synchronous responses. The maximum point on the synchronous curve indicates the PSS. Fifty percent of responses indicate the asynchrony detection thresholds were in the direction of delayed audio and delayed tactile stimuli.

The results of our investigation show that the point of subjective simultaneity (PSS) does not necessarily coincide with the point of objective simultaneity (0 ms [28]). We found that the PSS occurred at an audio delay of approximately 8 ms. The most interesting finding was that participants detected audio advances more accurately than audio delays. This fact might be related to physical rules (e.g., the speed of sound). The distance between human hands and ears is approximately 1 m. Therefore, sound takes approximately 3 ms longer to reach the brain compared to tactile stimuli. In addition, the physiological transduction times along the auditory and somatosensory neural pathways are different. Previous research showed that reaction times were 13 ms shorter for auditory stimuli compared to tactile stimuli [28]. Therefore, the human perceptual system may have evolved to tolerate longer audio delays than tactile delays. Pilot experiments showed that a short audio delay (between 1-8 ms) leads to a higher perceived quality than does synchronous reproduction. Similar tendencies have also been found for auditory-visual perception [29].

Location/Localization

Spatial origin is an important cue by which humans determine whether auditory and tactile signals originate from the same event or object. Naturally, if auditory, tactile, and visual information are generated by one multimodal event, the locations of the auditory and tactile events should coincide.

We investigated the minimum angle difference between auditory and tactile events necessary for a listener to conclude that the auditory and tactile event locations do not coincide [32]. Participants scraped a surface with their finger tip (a common multisensory event). In this experiment, nine loudspeakers presented the acoustic stimulus. We placed the loudspeakers 75 cm in front of the participant and activated them in a random order. We used electro-tactile finger stimulation to present tactile information at the tip of their index finger. Touch-induced sounds came from different loudspeakers along with a simultaneous tactile feedback when participants made the scraping movement (always at the same place). We asked participants whether they perceived that the sound was caused by their index finger. The percentages of "yes" responses are shown in Fig. 6. A probit regression fit to a cumulative Gaussian distribution modeled the psychometric function. The minimum angle necessary for participants to notice that the locations of the auditory and tactile events did not coincide was 5.3°. This result shows that humans are sensitive to differences in spatial source.

Frequency/Pitch

The frequency of sound and vibration are coupled to each other by physical laws. Human response to vibrations (i.e., tactile feedback) and sounds depends on their frequency. Therefore, we expect to observe conformity between the frequency of auditory and tactile stimuli in auditory-tactile VE applications.

Human sensitivity to discrepancies in frequency of auditory and tactile stimuli has previously been investigated [33]. The tactile stimuli were sinusoidal vibrations varying in frequency (4, 10, 50, 63, 80, and 100 Hz). Auditory stimuli were pure tones at fifteen different frequencies (31.5 Hz, 40 Hz, 50 Hz, 63 Hz, 80 Hz, 100 Hz, 125 Hz, 160 Hz, 200 Hz, 250 Hz, 315 Hz, 500 Hz, 630 Hz, 1000 Hz, and 2000 Hz). A tactile feedback mouse presented vibrations to participants. This mouse contained a motor that relayed vibrations to its guiding hand. In this experiment, subjects should imagine that the vibration and auditory information were produced by any device (e.g. a razor or a hair dryer) which they want to imagine. The authors simultaneously presented tactile and auditory stimulus pairs four times in a random order. Participants reported whether or not the same device/event caused the auditory and tactile information. As expected, participants tended to prefer stimuli pairs that had similar auditory and tactile frequencies to find the most suitable multimodal combination for integration. The frequency deviation threshold of the tactile stimuli was about 60 % of the auditory stimuli. In most cases, participants also judged second or third harmonics of the vibration frequency as suitable for the auditory frequency. The results of this study show that human sensitivity to frequency discrepancies between auditory stimuli and whole-body vibration are similar (approximately 60 % [34]).

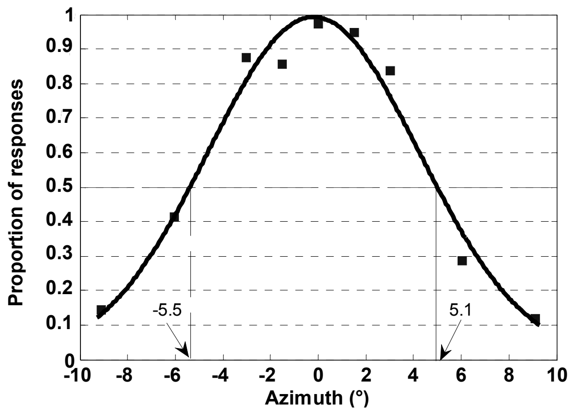

Fig. 6. The proportions of "yes" responses. Participants reported whether they perceived a sound as being caused by their index finger.

Intensity, Loudness, and Strength

Sound generation requires acoustic energy that, for the most part, structural movement supplies. This movement can be the result of tactile interactions with objects. Therefore, the sound pressure and force-feedback levels (by hitting) are coupled. These levels provide important cues for the brain to integrate various sensory data, such as simultaneity. For example, the reflected force-feedback information of an object and the associated sound when it is struck are coupled together. During perceptual development, infants learn that when they strike an object with more force, the associated sound becomes louder. When people strike an object, they expect to hear a loud sound; quiet sounds are less plausible. Thus, people will have difficulty integrating strong force-feedback information with quiet sounds.

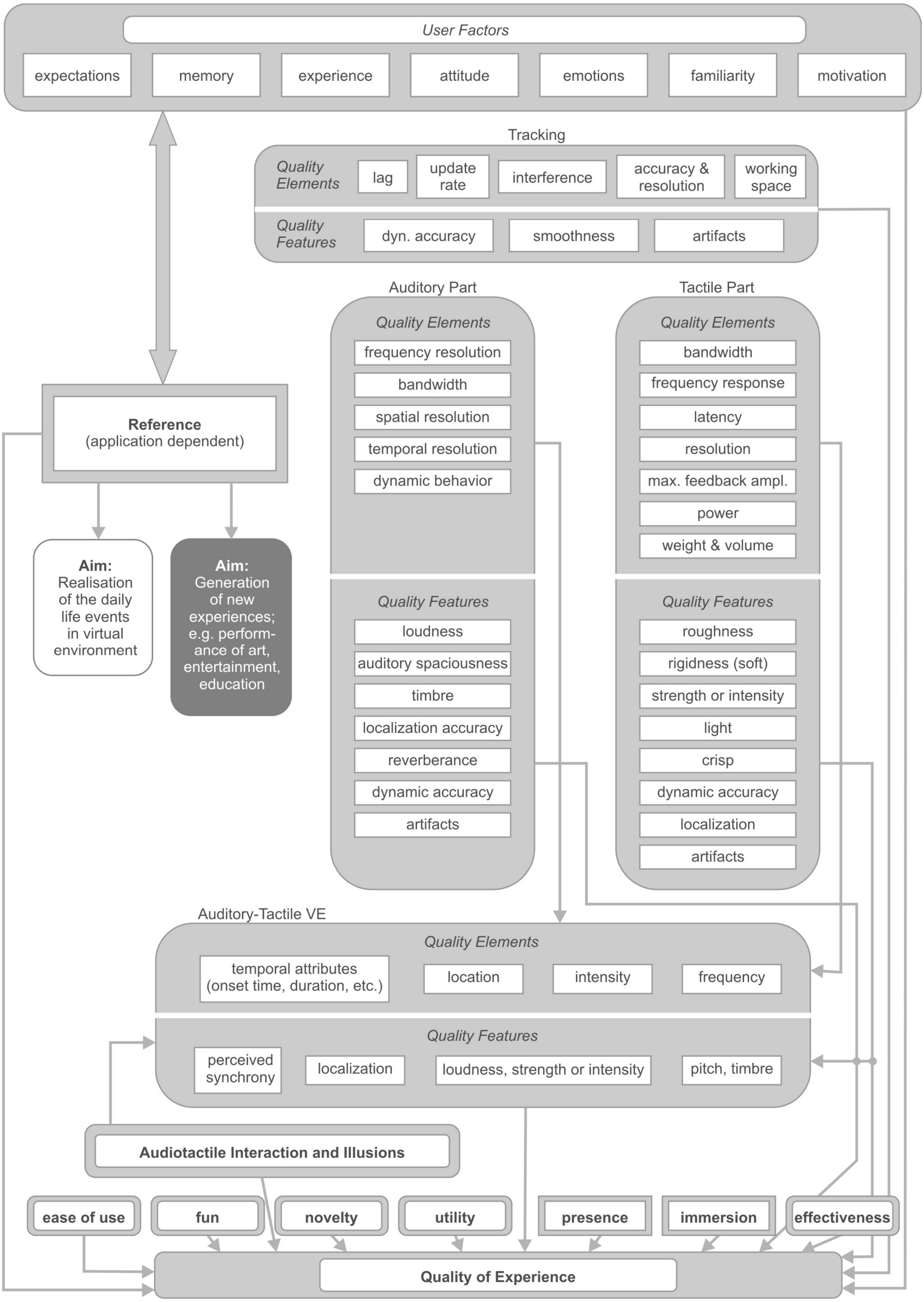

Fig. 7. The judgment of the quality of auditory-tactile VEs.

An investigation measured the perceptual threshold values between auditory and tactile events (e.g., playing of a virtual drum [35]). The sound pressure level of the auditory stimuli and the force-feedback level of the tactile stimuli varied within the experiment. The threshold was 17.6 dB as the level increased and 11.2 dB as the level decreased. People encounter different physical conditions and interact with different physical objects throughout their lives. These differences might lead to the ability to integrate different intensities across two sensory modalities, thus explaining the above results.

3 IMPROVING THE QUALITY OF AUDITORY-TACTILE VIRTUAL ENVIRONMENTS THROUGH THEIR INTERACTION

In addition to feedback, object and event recognition are essential for the user to interact with a VE. VEs are able to independently generate information in different modalities. Information in one modality can replace or alter information that is perceived within another modality. Audio-tactile illusions can be used in the conception of VEs. They can even improve the quality of audio-tactile VEs. The interface technologies currently available produce insufficient and low quality sensory outputs compared to the capabilities of human perception. When considering haptic interfaces, generating virtual walls as rigid as real walls is a problem. Simulating rigid surfaces is not yet possible. The appropriate use of auditory information may overcome this type of haptic interface limitation [9]. A psychophysical experiment investigated the effect of loudness on the perception of tactile force-feedback ("strength") by playing a virtual drum [35]. This investigation showed that auditory information could change the perception of a tactile stimulus. In fact, sound induced a tactile illusion; namely, when an auditory stimulus of different sound-pressure levels accompanies a constant haptic force-feedback stimulus, the auditory stimulus modulates haptic perception, and the magnitude of strength increases with increasing loudness despite a lack of force-feedback change.

Similar effects were observed in multisensory texture explorations. Roughness is an important physical and perceptual dimension of texture. Perceiving surface texture by touching it (e.g., scraping with fingertip) is a multimodal task in which auditory, tactile, and visual information is available. Previous research varied modulation frequency and the loudness of a sound to study their perceptual consequences [36]. The perceived tactile roughness was substantially altered towards the roughness that the auditory stimulus alone produced. Decreasing the modulation frequency increased the perception of tactile roughness, even though the tactile information was smoother than the auditory information. Increasing sound pressure level (by approximately 4 or 6 dB) also increased the perception of tactile roughness.

People often experience the vibrations generated by music. The air-borne sound causes seat vibrations or excites the skin surface directly. For some instruments (e.g. an organ) structure-borne sound is transmitted directly from the instrument to the listener. The synchronous presentation of vertical whole-body vibrations during a DVD reproduction of a concert can improve the perceived quality of the concert experience [37]. The whole-body vibration signal can be generated using the low-pass filtered audio signal. However, the frequency content of the vibration signal has an important influence on the quality judgments. The ideal low pass cutoff frequency is dependent on the particular music sequence.

Considering the results of the investigations, it is clear that benefits of the auditory-tactile interaction are very promising for engineers who develop VEs.

4 SUMMARY AND CONCLUSIONS

Fig. 7 presents the model of quality auditory-tactile VEs that contain the above-mentioned criteria. The judgment of quality is based on the user traits. These traits are grouped into factors. The quality elements and features of the auditory and tactile portions of the VE are presented in two separate modules. These elements and features are the preliminary stages of the auditory-tactile VE module.

Whether and to what degree these module factors affect judgments of quality depends on the reference of the user, in other words the aim of the auditory-tactile VE. If the VE reproduces the physics of a real environment, most of its elements will affect the judgment of its quality. In particular, the synchronicity and the spatial origin of auditory and tactile stimuli are important quality features. Sections 2 and 3 provide some guidelines based on experimental data. Auditory-tactile interactions and illusions are promising ways to improve the quality of a VE. Therefore, we present them as a module. If the aim of the VE is to generate new experiences (e.g., performance art, entertainment, and education), most of the factors in these modules do not have an effect because many users are open to experiences that do not conform to physical laws. Therefore, factors such as fun, novelty, and ease of use are important.

This study introduced a quality model for auditory-tactile VEs. This model describes relevant quality elements and features which influence the overall quality judgment with different relative weightings. Future efforts are necessary to determine the weightings of the individual elements and individual features on the overall quality judgment.

5 REFERENCES

[1] G. Burdea and P. Coiffet, *Virtual Reality Technology* (John Wiley & Sons, New York, 1994).

[2] DIN 55350: *Begriffe der Qualitätssicherung und Statistik. Part 12, Merkmalsbezogene Begriffe* (Beuth-Vertrieb, Berlin, 1989).

[3] U. Jekosch, "Meaning in the Context of Sound-Quality Assessment," *ACUSTICA united with ACTA ACUSTICA* 85, 681-684 (1999).

[4] J. Blauert and U. Jekosch, "Sound-Quality Evaluation − a Multi-Layered Problem" *ACUSTICA united with ACTA ACUSTICA* 83, 747−753 (1996).

[5] J. Blauert and U. Jekosch, "A Layer Model of Sound Quality," *in Proc. of Third Int. Perceptual Quality of Systems Workshop*, Bautzen, Germany (2010).

[6] U. Jekosch, "Basic Concepts and Terms of "Quality", Reconsidered in the Context of Product-Sound Quality," *ACUSTICA united with ACTA ACUSTICA* 90, 999-1006 (2004).

[7] J. Blauert, H. Lehnert, J. Sahrhage, and H. Strauss, "An Interactive Virtual-Environment Generator for Psychoacoustic Research," *ACUSTICA united with ACTA ACUSTICA*, 86, 94-102 (2000).

[8] A. Silzle, P. Novo, and H. Strauss, "IKA-SIM: A System to Generate Auditory Virtual Environments," *116th AES Convention*, Berlin, Germany, preprint number #6016 (2004).

[9] E. Altinsoy, *Auditory-Tactile Interaction in Virtual Environments* (Shaker Verlag, Aachen, 2006).

[10] P. Cook, *Real Sound Synthesis for Interactive Applications* (A K Peters Ltd. Natick, MA, 2002).

[11] R. Pellegrini, *A Virtual Reference Listening Room as an Application of Auditory Virtual Environments* (dissertation.de, ISBN 3-89825-403-8, 2002).

[12] A. Silzle, *Generation of Quality Taxonomies for Auditory Virtual Environments by Means of Systematic Expert Survey* (Shaker Verlag, Aachen, 2008).

[13] J. Berg and F. Rumsey, "Identification of Quality Attributes of Spatial Audio by Repertory Grid Technique," *Journal of Audio Eng. Soc.* Volume 54 Issue 5 pp. 365-379 (2006).

[14] N. Zacharov and K. Koivuniemi, "Unravelling the Perception of Spatial Sound Reproduction: Language Development, Verbal Protocol Analysis and Listener Training," *Presented at 111th AES Convention* at New York, September 2001 Preprint number 5424 (2001).

[15] S. Choisel and F. Wickelmaier, "Extraction of Auditory Features and Elicitation of Attributes for the Assessment of Multichannel Reproduced Sound," *118th AES international convention*, Barcelona, Spain, Preprint Number: 6369 (2005).

[16] S. Bech and N. Zacharov, *Perceptual Audio Evaluation - Theory, Method and Application* (Wiley GB−Chichester, 2006).

[17] S. George, "Objective Models for Predicting Selected Multichannel Audio Quality Attributes," Ph.D. Thesis, University of Surrey, (2009).

[18] G. C. Burdea, *Force and Touch Feedback for Virtual Reality* (John Wiley and Sons, Inc, New York, 1996).

[19] H. Jex, "Four Critical Tests for Control-Feel Simulators," in *Proc. of 23rd annual Conference on Manual Control*, (Cambridge, MA, 1988).

[20] M. Rosenthal, "Reflex™ Technology: Actuators Capable of Direct Control of Haptics from Audio Signals" in *Proceedings of Fourth International Workshop on Haptic and Audio Interaction Design*, (Dresden, Germany, 2009).

[21] D. Kontarinis and R. Howe "Tactile Display of Vibratory Information in Teleoperation and Virtual Environments," *Presence*, Vol. 4., No. 4, pp. 387-402 (1994).

[22] C. Power, "On the Accuracy of Tactile Displays," in *ICCHP 2006*, K. Miesenberger, et al., Eds. (Springer, Berlin, Germany, 2006), LNCS 4061, pp. 1155–1162.

[23] M. E. Altinsoy and S. Merchel, "Audiotactile Feedback Design for Touch Screens," in *Haptic and Audio Interaction Design*, M. E. Altinsoy, U. Jekosch, and S. Brewster, Eds. (Springer, Berlin, Germany, 2009), LNCS 5763, pp. 136-144.

[24] A. E. Kirkpatrick and S. A. Douglas "A Shape Recognition Benchmark for Evaluating Usability of a Haptic Environment" in *Haptic Human-Computer Interaction*, (Springer, Berlin, Germany, 2001), LNCS 2058, pp. 151-156.

[25] M. Stamm, M. E. Altinsoy, and S. Merchel, "Identification Accuracy and Efficiency of Haptic Virtual Objects Using Force-Feedback" *in Proc. of Third International Perceptual Quality of Systems Workshop*, (Bautzen, Germany, 2010).

[26] S. Möller, "Quality of Transmitted Speech for Humans and Machines," in *Communication Acoustics*, J. Blauert, Eds. (Springer, Berlin, Germany, 2005).

[27] S. Rihs, "The Influence of Audio on Perceived Picture Quality and Subjective Audio-Video Delay Tolerance," *in Proceedings of the MOSAIC Workshop Advanced Methods for the Evaluation of Television Picture Quality*, (Institute for Perception Research, Eindhoven, The Netherlands, 1995), pp. 133-137.

[28] M. E. Altinsoy, "Perceptual Aspects of Auditory-Tactile Asynchrony," *in Proceedings of the Tenth International Congress on Sound and Vibration*, (IIAV, Stockholm, Schweden, 2003), pp. 3831-3838.

[29] B. D. Adelstein, D. Begault, M. Anderson, and E. Wenzel, "Sensitivity to Haptic-Audio Asynchrony" *in Proc. Int. Conf. on Multimodal Interfaces*, (New York, NY, 2003) pp. 73-76.

[30] D. J. Levitin, K. MacLean, M. Mathews, L. Chu, and E. Jensen, "The Perception of Cross-modal Simultaneity," *in Proc. of Computing Anticipatory Systems.* AIP Conf. Proc. 517, (2000) pp. 323-329.

[31] A. Kohlrausch and S. van der Paar, "Audio-visual Interaction in the Context of Multi-Media Applications," in *Communication Acoustics*, J. Blauert, Eds. (Springer, Berlin, Germany, 2005).

[32] M. E. Altinsoy, "The Effect of Spatial Disparity on the Integration of Auditory and Tactile Information" in *Haptic and Audio Interaction Design*, R. Nordahl, F. Fontana, S. Serafin, and S. Brewster, Eds. (Springer, Berlin, Germany, 2010), LNCS 6306, pp. 20-25.

[33] M. E. Altinsoy, "The Influence of Frequency on the Integration of Auditory and Tactile Information," *in Proc. of the 18th International Congress on Acoustics*, (Kyoto, Japan, 2004).

[34] M. E. Altinsoy and S. Merchel, "Cross-Modal Frequency Matching: Sound and Whole-Body Vibration" in *Haptic and Audio Interaction Design*, R. Nordahl, F.

Fontana, S. Serafin, and S. Brewster, Eds. (Springer, Berlin, Germany, 2010), LNCS 6306, pp. 37-45.

[35] M.E. Altinsoy, J. Blauert, and R.H.Y. So, "Effect of Loudness on the Haptic Force-Feedback Perception in Virtual Environments," *J. Acoust. Soc. Am.* 114, 2330-2331 (2003).

[36] M.E. Altinsoy, "The Effect of Auditory Cues on the Audiotactile Roughness Perception: Modulation Frequency and Sound Pressure Level" in *Haptic and Audio Interaction Design*, A. Pirhonen and S. Brewster, Eds. (Springer, Berlin, Germany, 2008), LNCS 5270, pp. 120-129.

[37] S. Merchel and M.E. Altinsoy, "Vibratory and Acoustical Factors in Multimodal Reproduction of Concert DVDs", in *Haptic and Audio Interaction Design*, M. E. Altinsoy, U. Jekosch, and S. Brewster, Eds. (Springer, Berlin, Germany, 2009), LNCS 5763, pp. 119-127.

THE AUTHOR

M. Ercan Altinsoy

M. Ercan Altinsoy studied mechanical engineering at the Technical University of Istanbul and became a research and teaching assistant at the chair of mechanical vibrations and acoustics. In 2000, he was accepted to the Institute of Communication Acoustics, Ruhr-University Bochum, Germany as a Ph.D. student and participated in the International Graduate School for Neuroscience at the Ruhr-University Bochum. In 2005, he completed his Ph.D. thesis: "Auditory-Tactile Interaction in Virtual Environments". From 2004 to 2006, Ercan Altinsoy worked at HEAD acoustics as NVH Consulting Engineer. In 2006, he started lecturing at the Dresden University of Technology, Chair of Communication Acoustics. He is leading the „Audiotactile Interaction“ group at the same university. His research interests include psychoacoustics, vibroacoustics, product sound and vibration design, vehicle acoustics, auditory and haptic interfaces for virtual environments, tactile psychophysics and psychoacoustics.

Dr. Altinsoy is a member of the Audio Engineering Society, the German Society of Acoustics (DEGA) and the VDE Association for Electrical, Electronic & Information Technologies.