Generation of Quality Taxonomies for Auditory Virtual Environments by Means of a Systematic Expert Survey

Andreas Silzle
Institut für Kommunikationsakustik, Ruhr-Universität Bochum, D-44780 Bochum, Email:andreas.silzle@web.de

The evaluation and quantification of a general Quality Taxonomy for interactive Auditory Virtual Environments by means of the systematic expert survey (DELPHI method) is presented.

Introduction and Motivation

A general Quality Taxonomy for interactive Auditory Virtual Environments (AVE) was presented in [1], see Figure 1. The Quality Elements (QE), characterising the implementation of a technical application, influence the Quality Features (QF). These QFs describe a defined part of the perception and quality judgement. A combination of all QFs yields the user's Quality of Experience (QoE). The Quality of System (QoS) describes the combination of the QoE of the user and the cost defined by the implementation of the QEs.

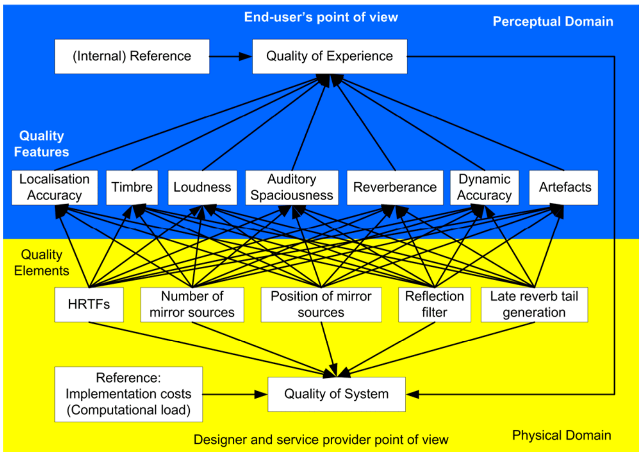

Figure 1: General Quality Taxonomy for an AVE

The QFs are influenced by the QEs in a multivariate way, i.e., every QF is influenced by a number of QEs and every QE influences several QFs. This is the bottom-up, signal-driven part of the perception and quality judgement, see [2]. A second inherent part of the quality judgement is the comparison of these external inputs to the internal references of the listener, the "desired nature of the sound" [3]. This internal reference highly depends on the task and the application. This is the top-down, hypothesis-driven part of the taxonomy.

The aim of the newly developed taxonomy is to describe the components involved in the quality judgement process of AVEs, and to quantify the relations between them for different applications, creating a road map through the multivariate quality judgement process to achieve a more holistic overview.

Quality Taxonomies help the researchers in this field to gain a better overview and insight into the quality assessment process. Additionally, as it is impossible to cover the complete spectrum of the quality assessment process with listening experiments, such Quality Taxonomies also assist the researchers in identifying and designing relevant listening tests to evaluate the quality of their applications.

Quality Taxonomies are very important in the software development process. The designers of such applications require qualitative, or, even better, quantitative models for specific algorithm classes to develop or optimise them. The description and optimisation of only one relation inside this multivariate process – as is often done in the literature – does not sufficiently cover the problem, because of the large number of these parameters and their interdependence.

Generation of Quality Taxonomies for AVEs

A detailed list of the eighteen QE definitions used in this investigation is given in [4]. It is based on [5] and significantly expanded in [6]. The list is developed based on literature knowledge and practical experience with AVE generators. Five of the selected seven QFs are presented in [5]. *Reverberance* and *artefacts* are added to describe better the relevant perceptions in an interactive AVE (for more details see also [4]). These QFs are not orthogonal to each other in the perceptual space and some of them have a multi-dimensional character. However, audio experts are familiar with these QFs and can easily differentiate between them. Furthermore listening tests are not necessary, which was a prerequisite for this investigation.

Three different applications for AVEs are chosen, with which the consulted experts are familiar:

1) Localisation test over Internet
2) Virtual chat-room
3) Edutainment scenario

The relations between the QEs and the QFs are divided into four categories to introduce a quantification of the importance:

1) No relation (weighting factor 0)
2) Less important relation (weighting factor 1)
3) Important relation (weighting factor 2)
4) Very important relation (weighting factor 4)

Evaluation by DELPHI Expert Survey

The new introduction of an expert survey to evaluate and quantify the general taxonomy extends the objectivity by the incorporation of a larger number of experts in the field.

The established DELPHI method (developed by Dalkey and Helmer [7] and described by Häder [8]) works with questionnaires. Its intention is to find a stable status (preferred is a consensus) on the subject of investigation by asking the experts' opinion in several rounds. After every Delphi round the experts get an average of all expert opinions presented. They have then the possibility to apply changes to this average. The experts are kept anonymously to each other, to avoid the influence of strong personalities or names.

Procedure and Subjects

An online-questionnaire with the QEs and QFs, together with a description of the applications was presented to the experts. The experts were asked to complete graphs by adding pointers between the QEs and the QFs and selecting an importance category. There was also the possibility to add new elements or features.

15 of the 28 approached experts from the field of AVE generation, with academic or industrial background (ten of them with a Ph.D. degree), participated in the first DELPHI round of the investigation. In the subsequent second DELPHI round, the number of changes was quite limited and all experts were in agreement afterwards. The detailed description of the averaging (using a cluster analysis) is given in [4].

Results

Because of the limited place in this publication the reader is pointed to [4] for all result plots, statistical analysis and a detailed discussion. As example, Figure 2 presents for the virtual chat room application the sorted results after the second and final DELPHI round. Because the experts agree in most of the cases with the average results of the first DELPHI round, the confidence intervals are zero, visualized in black. For the non-black dots, the experts vary in their judgments about the importance of these relations.

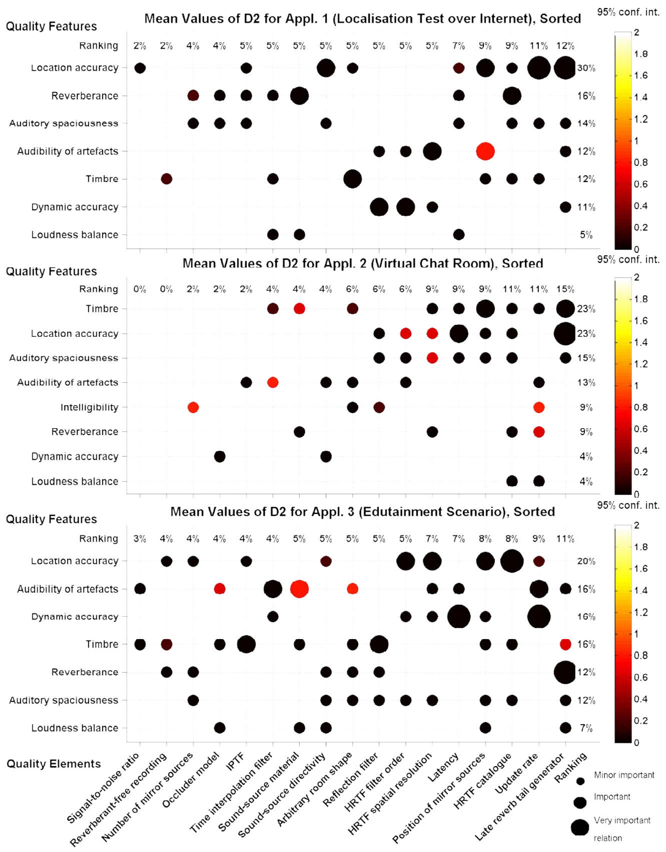

Figure 2: Quantitative interdependences of QEs and QFs after the second DELPHI round (point size: importance; shading: 95% confidence interval)

Figure 2 clearly documents the multivariate character between the QEs and the QFs. The virtual chat-room is the least complex application: it requires the lowest number of relations and has relatively low weightings between the QEs and QFs. The edutainment scenario is the most complex one: it requires the highest number of relations.

The most important QF for all three applications is localisation accuracy. For the virtual chat-room the QF timbre gets the same ranking as the localisation accuracy.

For the QEs, HRTF catalogue is the most important one. This is not surprising because this QE directly influence the QFs localisation accuracy and timbre. Interestingly, the late-reverb-tail generator is together with the number of mirror sources in the second position of the QEs. This is a good forecast of the listening test results, presented in detail in [4].

A Principal Component Analysis (PCA) reveals that five orthogonal dimensions are necessary to describe a minimum of 95% of the variance in the data of the taxonomy investigation.

In listening experiments the same QFs are used. Only three dimensions are necessary to describe the variance of these results. The subjects follow with their judgements much more the over all quality. That denotes, in the results of the taxonomy investigation is more detailed information.

Summary

A general Quality Taxonomy is evaluated and and the QE-to-QF relations are quantified for three different AVE applications. No listening experiments are necessary to generate them with the presented method, only a number of experts, who have experience in this field. Here, 15 experts participated in the DELPHI expert survey. Therefore, a large extent of objectivity of the results has been obtained.

The multivariate relations between the QEs and the QFs are quantified and clearly visible. The important QEs and QFs for the different applications are identified.

The expert survey – as used here – can only reflect the present average knowledge of the experts in their field.

The presented method leads to a much faster overview of the quantitative influence of all known QEs than any other method.

Literatur

[1] Silzle, A. (2005). Taxonomie und Optimierungsmethode für Systeme zur Generierung von auditiven virtuellen Umgebungen (Taxonomy and Optimisation Method for Generation Auditory Virtual Environments). DAGA '05. München

[2] Blauert, J. (2005), Analysis and Synthesis of Auditory Scenes, in Communication Acoustics, J. Blauert, Editor. Springer, Berlin, Heidelberg, New York.

[3] Jekosch, U. (2004). Basic Concepts and Terms of "Quality", Reconsidered in the Context of Product-Sound Quality. Acustica - acta acustica. 90(6), p. 999-1006;

[4] Silzle, A. (2007). Generation of Quality Taxonomies for Auditory Virtual Environments by Means of Systematic Expert Survey. Submitted PhD. Institut für Kommunikationsakustik. Ruhr-Universität Bochum.

[5] Pellegrini, R.S. (2002). A Virtual Reference Listening Room as an Application of Auditory Virtual Environments. PhD. Institut für Kommunikationsakustik. Ruhr-Universität Bochum, Bochum.

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

[7] Dalkey, N.C. and O. Helmer (1963). An Experimental Application of the Delphi Method to the Use of Experts. Management Science. 9, p. 458-467;

[8] Häder, M. (2002). Delphi-Befragungen. Westdeutscher Verlag, Wiesbaden.

The author would like to thank all participants of the expert survey and his colleagues from the Institute of Communication Acoustics.