Grant Agreement No.: 687645
Research and Innovation action
Call Topic: H2020 ICT-19-2015

ORPHEUS
Object-Based Audio Experience

Object-based broadcasting – for European leadership in next generation audio experiences

D5.3: Document on Methodology for Evaluation of User Experience

Version: v1.0

Deliverable type	R (Document, report)
Dissemination level	PU (Public)
Due date	31/05/2017
Submission date	30/06/2017
Lead editor	Werner Bleisteiner (BR)
Authors	Werner Bleisteiner (BR), Martin Ragot (b<>com), Andrew Mason (BBC R&D) Milan Kracht (ECandy), Michael Weitnauer (IRT), Tilman Herberger (Magix), Andreas Silzle (FHG IIS), Benjamin Duval (Trinnov),
Reviewers	Andreas Silzle (FHG IIS); Olivier Warusfel (IRCAM)
Work package, Task	WP 5, T5.4
Keywords	Evaluation of User Experience

Abstract

Deliverable 5.3 outlines the methodologies of objective and subjective evaluation methods applied by partners in T5.4 to examine various quality aspects of end user experience in 'object-based broadcasting'. It is based upon the project's main pillars, the use cases, the pilot architecture and the pilot itself. It sets the basic framework of practical examination and evaluation of user experience by ORPHEUS consortium partners.

Results related to this deliverable will be presented in *D5.6 Report on audio subjective tests and user tests* in month 29.

[End of abstract]

Document revision history

Version	Date	Description of change	List of contributor(s)
v0.1	21/03/2017	First draft and proposal of structuring	wb
V0.12	12/04/2017	Adding Guidelines text p 3 (later tbd)	Olivier, wb
v.02	25/04/2017	Chapter 1, Appendix A	wb (BR)
V0.3	26/04/2017	Chapter 2.2, 3.2, 4.2, Appendix B	Martin Ragot (bcom)
V0.35	27/04/2017	Suggestion for systematic approach	wb (BR)
V0.38	03/05/2017	Input chapter 2.1	Andrew (BBC)
V0.39	03/05/2017	Input chapters 2.3 and 4.2.2.4	Milan Kracht (ECandy)
V0.40	04/05/2017	Changes and addition after WebEx	all
v0.41	08/05/2017	input IRT	Michael Weitnauer (IRT)
v0.5	10/05/2017	input bcom, Magix – unconsolidated	Martin Ragot (bcom), Tilman Herberger (Magix)
V0.6	12/05/2017	input 3.1 by FHG IIS	Andreas (FHG IIS),
V0.7	15/05/2017	restructuring, editing	wb (BR)
V0.8	23/05/2017	Input BR and Trinnov	Benjamin (Trinnov), Werner
V0.89	31/05/2017	Restructuring & additions	wb (BR)
V0.9	02/06/2017	BBC input	Andrew (BBC)
V0.92	06/06/2017	document consolidated	wb
v0.93	08/06/2017	internal revision 1	Andreas Silzle (FHG IIS)
v0.95	15/06/2017	consolidation for internal review	wb (BR)
v1.0	23/06/2017	Final Version for PMC approval	wb (BR)

Disclaimer

This report contains material which is the copyright of certain ORPHEUS Consortium Parties and may not be reproduced or copied without permission.

All ORPHEUS Consortium Parties have agreed to publication of this report, the content of which is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License[1].

Neither the ORPHEUS Consortium Parties nor the European Commission warrant that the information contained in the Deliverable is capable of use, or that use of the information is free from risk, and accept no liability for loss or damage suffered by any person using the information.

Copyright notice

© 2015 - 2018 ORPHEUS Consortium Parties

[1] http://creativecommons.org/licenses/by-nc-nd/3.0/deed.en_US

Executive Summary

A continuously and stringent object-based broadcasting approach offers new and advanced possibilities to create and implement novel ways of encompassing experience and usability features. With audio becoming 'interactive' – within pre-defined producer authorized limitations – new challenges are opening up to make object-based media features accessible, understandable and operable. This applies throughout all stages of the media chain, during production, distribution and reception.

In addition, it's not just the audio itself that matters, but also the additional services and features (e.g. transcripts, additional text-based information or still pictures) becoming integrative components in the assessment of the media experience.

This challenges the development of appropriate user interfaces, in order to make human interaction to control and adjust complex technical metadata and parameters delivered alongside the audio on the different devices appropriate and convenient. Only if this can be achieved, the user will esteem object-based media technology providing an exciting and satisfying experience.

As a consequence, previous separately regarded domains for examining and evaluating 'quality of end user experience' have to be considered inclusively or convergent.

In this deliverable, the ORPHEUS consortium partners give first an outline of well-established and proven methodologies for objective and subjective evaluation in the fields of audio quality, application usability and interactivity appreciation.

Based upon the ORPHEUS project's main pillars, the use cases, the pilot architecture and the pilot itself, we then develop a basic framework for practical examination and evaluation of user experience within an object-based media eco-system as a holistic model.

This model is comprised of three main categories:

- audio experience
- usability experience
- information experience

This is our starting point for further examination in the course of the project. Findings from the pilot phases will be used in order to adjust and refine the model, as well to elicit possible interdependencies of the different categories and their respective characteristics and quality features.

© 2015 - 2018 ORPHEUS Consortium Parties

Table of Contents

© 2015 - 2018 ORPHEUS Consortium Parties

© 2015 - 2018 ORPHEUS Consortium Parties

List of Figures

© 2015 - 2018 ORPHEUS Consortium Parties

List of Tables

© 2015 - 2018 ORPHEUS Consortium Parties

Abbreviations

ADM	Audio Definition Model
BAQ	Basic Audio Quality
CMS	Content Management System
CSUQ	Computer System Usability Questionnaire
DVB	Digital Video Broadcast
EC	European Commission
GUI	Graphical User Interface
HCI	Human-Computer-Interaction
HOA	Higher Order Ambisonics
ITU	International Telecommunication Union
MPEG	Moving Pictures Expert Group
MS-IPM	Multiple Stimulus Ideal Profile Method
MUSHRA	MUltiple Stimuli with Hidden Reference and Anchor method
OBA	Object Based Audio
OLE	Overall Listening Experience
QE	Quality Elements
QF	Quality Features
QoE	Quality of Experience
UCD	User Centric Designs
UHD	Ultra HD Broadcast
UI	User Interface
USP	Unique Selling Proposition
UX	User Experience
VIP	Internal project format of MAGIX SEQUOIA

© 2015 - 2018 ORPHEUS Consortium Parties

1 Quality - End User - Experience in the ORPHEUS architecture

A continuously and stringent object-based broadcasting approach offers new and advanced possibilities to create and implement novel ways of encompassing experience and usability features. With audio becoming 'interactive' – within pre-defined producer authorized limitations – new challenges are opening up to make object-based media features accessible, understandable and operable. This applies throughout all stages of the media chain, during production, distribution and reception.

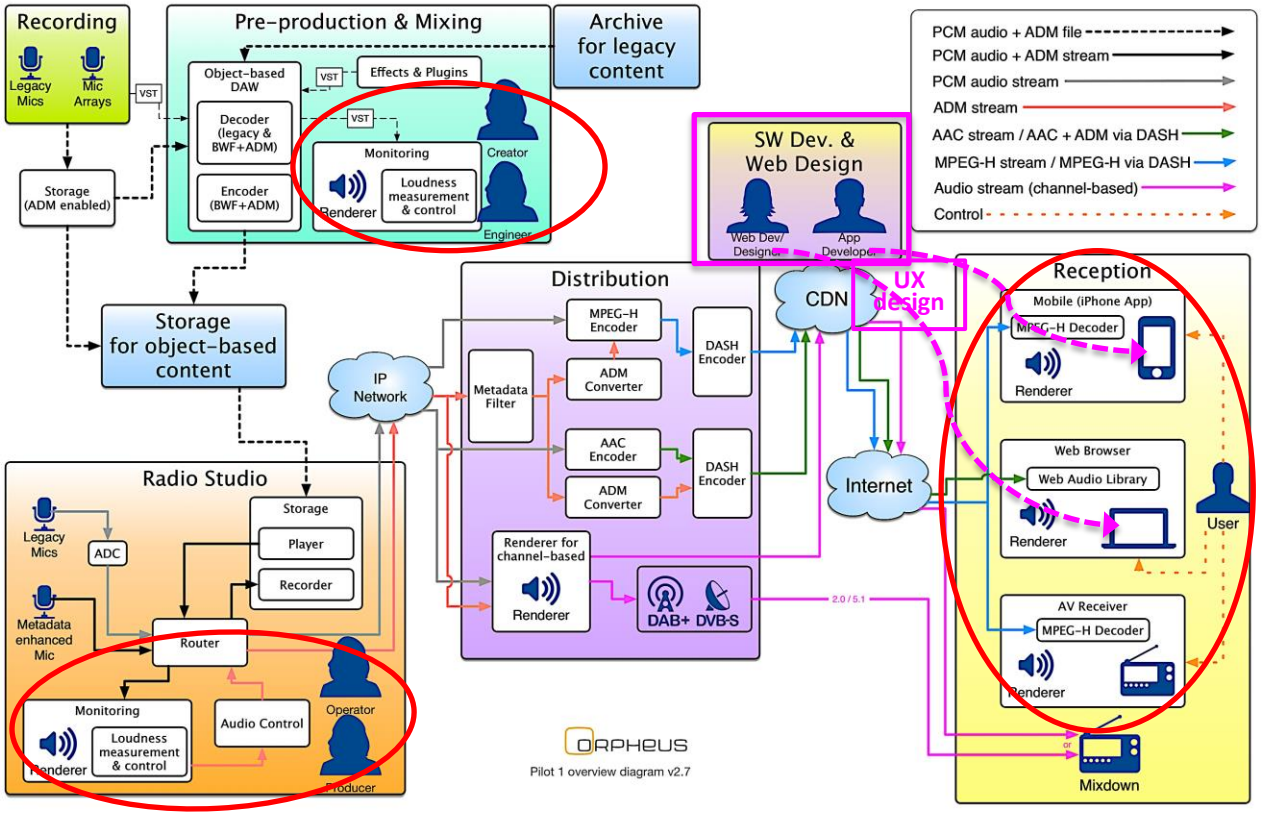


Figure 1: User experience issues in the ORPHEUS Pilot architecture

Regarding the ORPHEUS pilot architecture, the quality assessment of an object-based audio broadcast is mainly to be located in three macroblocks (red circles in *Figure 1*):

- Pre-production & mixing: by the creator/director/Tonmeister and sound engineer of a production (i.e. radio drama, piece of music)
- Radio Studio: by the producer and operator of the live broadcast or play-out of pre-recorded material (rundown schedule)

  Yet, the two production instances will mainly deal with creating and designing the advanced **audio experience**, focusing on the specific 'quality features' of an object-based production like 'spatialisation', 'immersion', 'intelligibility' etc.
- Reception: by the user/listener using one or several devices connected to different distribution channels

The general public 'end user' encounters object-based media within the reception macro-block of our pilot architecture via

- the ORPHEUS App (iOS)
- a Web Browser on a computer (WebAudio API)
- a high-end AV-Receiver as typical consumer electronics home entertainment device, Figure 2.

© 2015 - 2018 ORPHEUS Consortium Parties

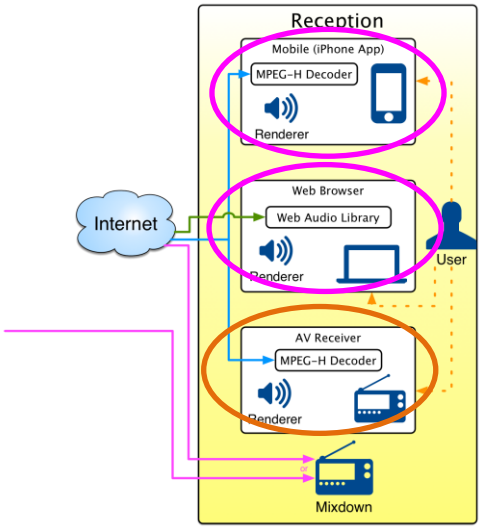


Figure 2: Reception macroblock and devices

Within this block of the ORPHEUS pilot architecture, users have the opportunity to choose between devices with different user interfaces. Nevertheless, all object-based media features provided are contained in the incoming IP stream of the broadcast and it is a matter of technical possibilities to be explored, which one of them are requested and how they can be optimally implemented.

So, the amount of object-based features represented on these devices may vary to some extent, depending on limitations of the used distribution channel, applied audio codec, or the restrictions of the 'man-machine-interface':

Naturally, all three types of devices offer different UIs, display and interaction possibilities. And of course also user demands and expectations for delivering audio and visual representations of media do vary.

As for the mobile app and the web browser, another entity is to be 'virtually' added to the ORPHEUS pilot architecture (Figure 1 magenta square):

- *Software & App Development & Web Design* taking care of implementing the various features of the object-based audio stream into usable GUIs, so to speak 'creating and designing a unique – and broadcaster branded - **user experience**.

This entity may not be regarded as an essential part of the object-based broadcast chain per se, but is nevertheless a prerequisite for the overall process of representation and reproduction.

Within ORPHEUS, the role of design and technical implementation of the app is taken by ECandy, as, so to speak, 'pioneering system integrator'.

Whereas the necessary developments for the browser representation for the ORPHEUS pilots - using WebAudio API, HTML5 etc. – is mainly to be fulfilled by BBC R&D (TASTER: http://www.taster.bbc) and IRT (IRT-lab: https://lab.irt.de).

For the regular operations in broadcasting, the technical requirements of object-based broadcasting representation and reproduction in web technologies have to become integral part of future web content management systems (CMS).

The technical implementation and the UI design for the AV-Receiver (orange) are controlled exclusively by the manufacturer of the device.

© 2015 - 2018 ORPHEUS Consortium Parties

2 Outline of Applied Approaches and Methodologies by Partners

In this chapter, ORPHEUS consortium partners involved in *T5.4 Quality of End User Experience* describe their predispositions, approaches, expertise, and standardisation activities for different categories of quality examination.

2.1 Quality Evaluation at the BBC

2.1.1 Technical Quality

BBC R&D has a long track record of subjective and objective testing to assess the quality of new technologies. The need to do this arises from the desire to ensure that the audience is provided with a high quality experience, within the constraints of cost and complexity. In recent years, two important areas have been the assessment of the degradation introduced by the use of data-reduction codecs (MPEG audio layer II, for example) and of watermarking.

The development of the standard test methods used for this kind of assessment - described in Recommendations ITU-R BS.1116[2] and BS.1534[3] - owes much to the efforts of BBC R&D. One notable characteristic of both these methods is the requirement of known reference stimuli. In the kind of tests mentioned, this is the original audio item, before it was coded and decoded, or watermarked.

The use of object-based audio, with the need for a renderer to generate the signals for loudspeakers or headphones, means that there will not be a known reference stimulus to present to the assessor - it will depend on the renderer. Because of this, BBC R&D is actively involved in the development of a new test method that does not require a known reference. Called the "multiple stimulus - ideal profile method" (MS-IPM[4]) the assessor can indicate what would be their ideal value of some perceptual attribute, rather than having it defined by the properties of a given stimulus.

It is expected that this method will be used for the assessment of renderers (when more than one is available and one must be selected according to a subjective quality criterion). Further, it might be used for the assessment of the results of object-based audio metadata manipulation.

The aim of assessments such as these is ultimately to ensure the quality of experience for the audience. This can be affected by all parts of the programme chain: Standards can be influenced for technologies used in consumer devices, and choices guided for the selection of equipment to be used for monitoring during programme production.

2.1.2 Usability

More than one part of the BBC assesses the “usability” or “audience appreciation” of different products or services. The type of assessment done depends strongly on what is being assessed, and for what purpose. BBC R&D has two dedicated user testing laboratories (one in London, and one in Salford), which are equipped with an observation room behind one-way glass, and equipment for monitoring users’ actions (cameras, microphones).

To expand its capability in user experience research, in 2013 BBC R&D established a partnership with six UK universities[5], creating a virtual centre of excellence. Naturally, there is a strong overlap of its four broad areas of interest stated at the time and the aims of the ORPHEUS project:

[2] ITU-R Recommendation BS.1116-3, Methods for the Subjective Assessment of Small Impairments in Audio Systems. 2014, Intern. Telecom Union, Geneva, Switzerland.

[3] ITU-R Recommendation BS.1534-3, Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems (MUSHRA). 2015, Intern. Telecom Union, Geneva, Switzerland

[4] N. Zacharov, C. Pike, F. Melchior & T. Worch: “Next generation audio system assessment using the multiple stimulus ideal profile method”; 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX); Lisbon, Portugal

[5] The partner universities are: University of Bath, University of Dundee, University College London, Newcastle University, University of Nottingham, and Swansea University

© 2015 - 2018 ORPHEUS Consortium Parties

- The user experience of an IP broadcasting system, developing future services across multiple platforms and creating public service broadcasting for the digital age.
- Designing for new interactions, moving beyond gesture and voice to develop new ways of controlling and displaying digital content for more natural and engaging interfaces.
- Sustainable approaches to user capability, allowing the industry to meet the changing needs of older users, young children and people with disabilities in accessing digital media.
- New production interface technologies, using interface and interaction technologies in new ways to give production teams the most creative and effective tools to craft new forms of content.

It is not known at this stage if, or to what extent, the partnership will be involved in any ORPHEUS evaluations.

BBC R&D's own assessments have often been conducted "in house" rather than with large-scale public trials. With the recent incorporation of the BBC Connected Studio team into R&D, the scope of testing has expanded to include "BBC Taster"[6] which is the audience-facing platform that was built and is run through Connected Studio. BBC Taster is a website that invites our public audiences to try, rate and share the latest digital pilots from across the BBC, showcasing a range of digital tools, techniques and content. It is a home for new ideas, a resource for experimentation and learning around new formats and digital prototypes. The pilots put on BBC Taster are available only for a limited amount of time, which is typically a few months. As with any prototype, learning from the pilot during and at the end of this period, is fundamental.

Assessments done in house have made use of concepts such as the Geneva Emotion Wheel - a theoretical research tool repurposed by BBC R&D for the purposes of interactive testing - to enable comparison across levels of emotional response, across multiple points of stimulation, coupled with biometric research, that studies unconscious responses to stimuli by brainwaves (via EEG monitor), heartrate (via band) and skin / sweat response (via sensor) to learn why people make the decisions they do. This enables the estimating engagement, affinity, vigilance (decision making), and excitement (see Figure 3).

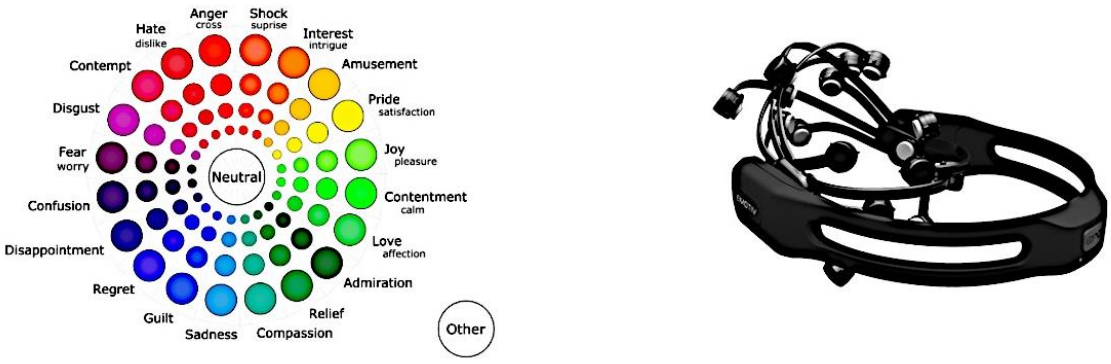


Figure 3: The Geneva Emotion Wheel (left), and an EEG sensor used in audience response assessment

More details of some relevant earlier pilots from BBC Taster can be found in Section 3.3 *Testing user experience in the browser: BBC Taster*.

[6] BBC Taster - http://www.bbc.co.uk/taster

© 2015 - 2018 ORPHEUS Consortium Parties

2.2 Quality Evaluation at b<>com

2.2.1 Objective - Target Groups – Methodologies

In the ORPHEUS project, the objective of the Uses and Acceptability laboratory at b<>com is to assess, analyse and try to improve the user experience through three following main missions:

- A benchmark with the objective of analysing and comparing existing systems in the market, in order to identify relevant functionalities and innovative practices for the ORPHEUS project.

- A heuristic evaluation to support the development of mobile applications by taking into account ergonomic criteria and cognitive knowledge. The aim is to create an interface adapted to human functioning, the simplest, easiest and most usable.

- Finally, user tests, in order to collect real data, directly with real end users, to understand the influence of the innovations created within the ORPHEUS project on the user's perception.

Overall, the main focus for the Uses and Acceptability laboratory at b<>com, in the ORPHEUS Project, is to assess the quality of the information and ease of use of the provided interfaces (mobile application). In this objective, different methodologies are implemented (benchmark, heuristic evaluation, user tests, etc.). Benchmark and heuristic evaluation involve UX design experts. The purpose is to be able to evaluate interfaces quickly and efficiently in the first steps of design. Once these first iterations have been carried out, other methodologies such as user tests make it possible to involve more deeply the end users. So, from the initial design stages, b<>com has involved potential end users to test the mobile application developed by ECandy. Therefore, the main focus is on ease of use and quality of information, because these notions, although not sufficient, are necessary for a quality user experience. The ORPHEUS project is in line with this conception, the User Centric Designs (UCD) approach[7]. The iterative tests, built by multidisciplinary teams, involve users and aim to understand their overall experience, their tasks and their environment, from an innovative perspective.

The benchmark, by analysing the good practices already implemented in the market, allows to take inspiration from the reality, in order to reproduce and create a quality user experience. A number of previously described criteria also made it possible to evaluate the way the information was presented on the interface, in order to draw on good practices and avoid the worst.

The heuristic evaluation, in a second step, made it possible to evaluate the interface already created within the project so that it satisfies ergonomic criteria, necessary for a quality user experience (feedback, quality of information, etc.).

Finally, user tests were able to gather real-time feedback from real users. Beyond the objective indices (time to complete a task, number of successful tasks, etc.), many subjective indices have been extracted. For example, some questionnaires (e.g. CSUQ[8]) were administered to users to understand their perception of the interface and the quality of the perceived information. From this information, it becomes possible to confirm or invalidate the initial design choices and to modify accordingly the already developed interface.

[7] ISO standard 9241-210: 2010

[8] Lewis, J. R. (1995). IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. International Journal of Human-Computer Interaction. (7,1). 57-78.

© 2015 - 2018 ORPHEUS Consortium Parties

2.3 Quality Evaluation at ECandy

2.3.1 Objective - Target Groups – Methodologies

ECandy works in close collaboration with b<>com during the design process of the mobile app. Main focus of quality evaluation here is to get information about usability and information quality for end users. ECandy will pay less attention to user evaluation for audio quality, although it won't be neglected totally. Other parties involved in the ORPHEUS project have better facilities to examine audio quality features for end users, so they will be able to give more attention to this section. As the FHG MPEG-H renderer is used for the conversion from audio objects to down-mixed audio streams, tests of the rendered audio quality can be conducted in the studio (and have been done already extensively before ORPHEUS).

During user tests, executed in collaboration with b<>com, subjective assessment on audio quality can be incorporated through questionnaire surveys.

The design process during the ORPHEUS project has a strong iterative character: the app is developed in different design cycles, which have the following, rough procedure:

1. statement of design goals: which end user needs do we want to accomplish?
2. creating ideas how above goals can be achieved
3. work out solutions on relevant ideas
4. apply solutions to a prototype
5. test prototype
6. analysis of test results

Steps 1 till 4 are executed by ECandy, step 5 by b<>com and step 6 in cooperation by both companies. The analysis of the test results will be input for a next design cycle.

The above procedure offers insight in user behaviour concerning the mobile app. It tells us whether a solution works well for achieving a stated goal, or if it should be improved during a next iteration.

Besides the qualitative evaluation as described above, tools for quantitative research in the app will be applied too. These tools will offer statistic data about user behaviour in a more general perspective. It offers answers to questions like:

- how many users do use the app
- how long do users spend on the app
- how many users handle specific features or functionalities
- how many users start a specific task and which percentage finalises it?

Quantitative results give insight in the acceptance and relevance for different functionalities within the app by the user and thus help to improve it during future design cycles. A detailed description of this quantitative analysis method will be given in section 4.2.

© 2015 - 2018 ORPHEUS Consortium Parties

2.4 Quality Evaluation at Trinnov

2.4.1 Objective - Target Groups – Methodologies

Trinnov products are targeting high-end hi-fi and home-cinema audio. The hi-fi users are very sensitive to audio quality, and the home-cinema users to quality and immersion, so the main objectives of the quality evaluation conducted at Trinnov will focus on these two aspects of the rendering of object-based content.

The main aspects to be evaluated are the newest audio rendering algorithms that are embedded in the demonstrator. First, the object rendering evaluation should be evaluated with contextual testing. This rendering is achieved with the MPEG-H decoder, and the main aspect to be evaluated will be the quality of its integration. But other algorithms are being incorporated in the demonstrator to improve the quality of experience and also need quality evaluation, such as the new bass frequency rendering optimisation.

Unless the ideal user group should be a panel representing the usual users of Trinnov audio processors (audio enthusiasts, but also sound engineers and 3D audio specialists), the panel that we will gather or the tests will only be audio experts, since it is considered as commercially dangerous to ask potential clients to test products that are still being developed. They are usually involved only in final testing for processors that are about to be marketed.

No methodology has yet been chosen, but the Multiple Stimulus Ideal Profile Method (MS-IPM) methodology will be studied. One main advantage of this reference-less methodology is that the test subjects can evaluate separately different pre-chosen attributes and therefore draw a more precise picture of the quality of what is being evaluated.

Apart from audio quality, the usability will also be tested. But since the main focus of Trinnov is about audio quality, these tests will not be formalized as much as those evaluating the audio quality.

© 2015 - 2018 ORPHEUS Consortium Parties

2.5 Quality evaluation at IRT

2.5.1 Objective

The focus of IRT regarding quality evaluation is, very similar as BBC R&D, on the technical quality. The IRT is developing a component, to be used at the distribution stage before, for the automatic reduction and adaptation of audio object scenes to specific limitations. The limitation can be either caused by a certain profile of the distribution codec (e.g. MPEG-H low complexity profile with only 16 simultaneously rendered objects), by end device restrictions due to the computational power or available bandwidth or other reasons. To preserve the creative intent of the sound engineer or the tonmeister and to guarantee the best end user experience is the most important aim for this component. Different algorithms and approaches are being developed by IRT and these need to be evaluated to find the best or most suitable one.

2.5.2 Target groups

Even though the results of the planned evaluations will have influence on the end user experience, the main and probably only listening test subjects will be expert listeners. Expert listeners are in this case on the one hand experienced audio engineers who are used to evaluate different sound examples, produced by e.g. different encoding algorithms. On the other hand, experienced sound engineers and tonmeisters are also planned to assess different pre-processing algorithms regarding the preservation of the creative intent. It is planned to have 15 – 25 experienced listeners for the evaluation with MS-IPM.

2.5.3 Methodologies

We plan to use the "Multiple Stimulus Ideal Profile Method" (MS-IPM) methodology[9] for the evaluation of different algorithms and approaches for our Pre-Processing component. Compared to BS.1116[10] and BS.1534[11], the method described can provide additional data regarding the perceptual performance of the sound systems under study. It provides measures of overall subjective quality, as well as characterising the perceptual nature of these preferences using attributes. Furthermore, the method seeks to establish how well the sound systems under evaluation compare to a perceptual ideal provided by assessors. The MS-IPM is designed to allow evaluation of systems in applications where there is not an explicit reference of target quality or an original unprocessed stimulus. It is primarily an exploratory evaluation method, useful for providing insight into complex multi-dimensional experiences, and for discovering what drives the preferences of listeners.

2.5.4 Content

For that purpose, we will use existing object-based content either produced especially for ORPHEUS and its pilots or from other occasions (e.g. Turning Forest by BBC R&D and the S3A reserach project[12]) but also special content produced by the IRT to test certain algorithmic behaviours. Such special content pieces potentially include certain numbers of objects and combinations of ADM parameters (e.g. special movements of objects or combinations with channel-beds).

2.5.5 Facilities

The evaluation will be conducted at the 3D Audio Lab of IRT (Figure 4). The room is equipped with all loudspeaker configurations listed in ITU-R BS.2051[13] (Advanced Audio Systems) as well as with headphones for binaural playback. The listening test will be driven by a Client-Server based browser solution ("leap"),

[9] N. Zacharov, C. Pike, F. Melchior & T. Worch: "Next generation audio system assessment using the multiple stimulus ideal profile method"; 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX); Lisbon, Portugal

[10] https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1116-3-201502-I!!PDF-E.pdf

[11] https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1534-3-201510-I!!PDF-E.pdf

[12] http://www.s3a-spatialaudio.org/wordpress/

[13] https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.2051-0-201402-I!!PDF-E.pdf

© 2015 - 2018 ORPHEUS Consortium Parties

developed by IRT where the MS-IPM methodology user interface is implemented.

Figure 4: 3D Audio Lab at IRT

2.6 Quality Evaluation at MAGIX

2.6.1 Main Focus and Target Group

The DAW Sequoia is the main contribution of MAGIX to the ORPHEUS project. A typical user of Sequoia is the audio engineer in the pre-production studio. Sequoia provides a professional user interface for these skilled people and will incorporate all new functionalities of the ORPHEUS context under this interface, as seamless as possible. The goal is to let Sequoia trained audio engineers work with the ORPHEUS features with minimal learning effort.

Quality testing has to be done in the areas audio quality and usability quality.

2.6.2 Audio Quality Testing

Testing the audio quality of the ORPHEUS features is not trivial because they require the conversion between the internal project format (VIP), the ORPHEUS standard format (ADM) and export formats (such as MPEG-H).

Especially between VIP and ADM some effects / settings can be transferred directly, some have to be rendered prior to the export and others cannot be transferred at all. The goal is to provide the audio engineer always a clear information, if the selected export target (ADM, MPEG-H) contains all audible and visible data of his project or not. This can be tested in an automatic objective way by creating test projects which contain all classes of effects available in Sequoia. These projects are then exported, reimported and bounced to WAV files. These resulting WAVs can be analysed via difference signal analysis to ensure they contain the very same material as the original project.

The difference signal analysis is common practise at MAGIX and is used for all kinds of internal DSP effects and Plug Ins. So this methodology has to be enhanced to include the new ORPHEUS related features. The expected result is a difference signal with zero content or almost zero because of dithering and rounding effects.

© 2015 - 2018 ORPHEUS Consortium Parties

Also acceptable is information presented to the audio engineer explaining which part of the project cannot be translated into the desired target format, helping him to replace or change these parts. This method of audio quality testing is a standard procedure in the Sequoia release cycle, so it is permanent on-going and has just to be enhanced with new features.

Another important aspect of the quality testing is the audio real time performance. Because of the high demands of typical ORPHEUS projects (e.g. with 32 tracks for 3D microphone array material, special plugins in each track for HOA conversion, multi-track renderer plugins in the master section etc.) the internal audio processing architecture has to be optimized intensively. This results in lower CPU consumption for playback and editing and is monitored during internally testing the program as well as external usability testing. The goal is a lag free playback on typical PC systems with low audio latency.

2.6.3 Usability Testing

For subjective usability testing of the new ORPHEUS features in Sequoia, MAGIX will use its existing group of registered Beta users. These are mainly German broadcast audio engineers who know Sequoia very well and are trained in its GUI workflow.

MAGIX will provide them a short description of the new functionalities and an explanation of how to use them in the typical style of the Sequoia manual and online system. Then MAGIX will get feedback if this worked well for the Beta users or where problems were discovered. In this case, MAGIX will tune the workflow and the GUI in the next iteration and start the external test again. This way, MAGIX managed in the past to adapt even complicated workflows exactly to the needs of the broadcast audio engineers, so we are optimistic that this approach will also come to good results for the ORPHEUS features.

This subjective GUI testing is planned for the last phase of the ORPHEUS project, where all technical features are working and now can be tuned at the GUI level.

2.7 Quality Evaluation at BR

2.7.1 Motivation and Objective

In classical broadcast media, the question of ‘quality’ has diverging aspects: Editorial departments consider mainly the quality of a ‘story’ (i.e. accuracy and relevance in journalism), the grade of interest and impact in fictional and non-fictional story-telling (i.e. radio drama and documentaries), the prominence and significance of musical pieces. Although sometimes the technical quality of broadcast material may objectively or individually be regarded as technically poor, the legitimation of airing such material to the public prevails, due to claimed ‘historic dimensions and character’. This antagonism between ‘content’ and ‘technical’ quality is quite often subject to internal discussions between editorial and technical departments. Of course, ideally both categories, content and technical persuasiveness come together.

In addition, as broadcast receiver devices used to - and will be also in the future to a considerable extent - be supplied by third party consumer electronics companies, the question of ‘usability’ in operating these devices has only been of limited interest for the broadcaster. Only in recent years, with the advent of additional digital services in Digital Broadcasting (DVB, DAB) and the Internet (streaming services, apps) in conjunction with their ‘hybrid’ implementation, the question of the display, accessibility and ease of usage of these services has gained importance for broadcasters. Development of applications, design and interaction interfaces for these digital services are increasingly executed in-house. Therefore, related capabilities and competences for software development and platform deployment are being established. With future next generation object-based audio developments, these questions will gain even more importance, as the experience of media will encompass more dimensions.

2.7.2 Methodologies and facilities

Historically the examination, assessment of and guidelines for ‘technical quality’ for German Broadcasters was mainly delegated to and executed by their common R&D department, the IRT. Therefore, especially for audio experience examination, BR will closely collaborate with IRT for all quality aspects, as they can provide their invaluable facilities and expertise in this field (c.f. chapter 2.5). The methodologies will match

© 2015 - 2018 ORPHEUS Consortium Parties

IRT's recommendations and practices. The target group will be BR sound engineers and specific identified non-technical test users from editorial departments.

In addition, if applicable, in the further course of developments within the pilot phases, we want to apply awareness and acceptance studies for audio experience characteristics, i.e. binaural and multi-channel reproduction, on a broader basis, targeting towards the general end user. For this, we strive to adapt appropriate audio evaluation methodologies with our project partners in the ORPHEUS app and through online surveys in the browser.

As for the 'usability/user' and 'information' experience examination aspects, BR will employ its own, recently established Userlab along with BR's Media Research department. Considering the ORPHEUS app and the Web browser as end user device, we want to evaluate generated usage data as well as apply acceptance and awareness in questionnaires for specific target user groups.

BR's Userlab applies usability and user experience examinations according to the Evidence-Based User Experience Research model by the Nielsen Norman Group[14]. It claims that "elaborate usability tests are a waste of resources. The best results come from testing no more than 5 users and running as many small tests as you can afford"[15], see Figure 5.

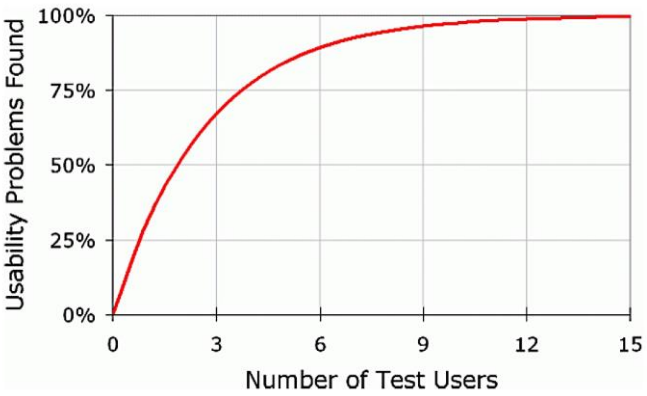


Figure 5: Nielsen/Landauer model of ratio of usability problems and number of test users[10]

2.7.3 Content and Stimuli

As for the examination of 'audio experience', BR provides to ORPHEUS several productions for the different stages of the project's pilot phases. They may all be used by our partners as test material, representing the variety of 'quality features' and characteristics defined in use cases.

Up-to-date the origin of this document there are to be mentioned:

- A radio football report – 5': immersive sound, fgd/bgd adjustment, transcript – single language (Ger)
- A report "Experience object-based audio" 15': immersive sound, fgd/bgd adjustment, transcripts, multiple-languages (Ger/Eng)
- A radio documentary on the "History and Art of Foley Making" – 15': immersive sound, fgd/bgd adjustment, transcripts, multiple-languages (Ger/Eng/Fr)
- A music piece "Gigue" by Mozart – 3': immersive sound

[14] Nielsen, Jakob, and Landauer, Thomas K.: "A mathematical model of the finding of usability problems," *Proceedings of ACM INTERCHI'93 Conference* (Amsterdam, The Netherlands, 24-29 April 1993), pp. 206-213. c.f.

[15] Nielsen, Jakob: Why You Only Need to Test with 5 Users https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

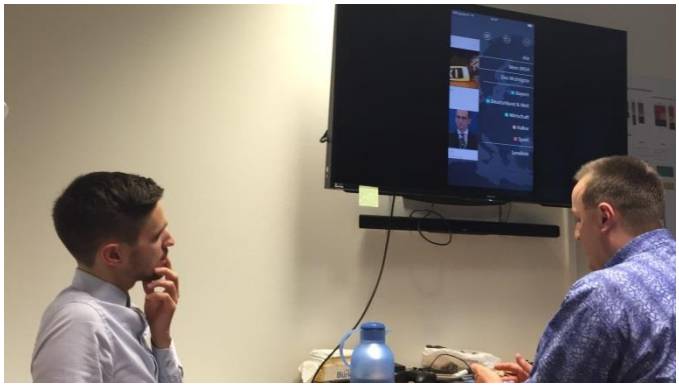
© 2015 - 2018 ORPHEUS Consortium Parties

2.7.4 Participants and Target Groups

Audio experience: Listening experience tests in different spatial reproduction formats are considered to be conducted in various stages of ORPHEUS' proposed end-to-end object-based broadcasting chain.

- during pre-production: with sound engineers and creators
- during playout: with operating engineers and producers
- in reception: with listeners and users in various target group

Usability and information experience: Evaluation will focus mainly on the BR version of the ORPHEUS app and a possible implementation in web browser on a test site. Here we will develop the appropriate methodology with BR's Userlab, see Figure 6 & 7



Figures 6 & 7: test scenarios in BR Userlab

© 2015 - 2018 ORPHEUS Consortium Parties

3 Approaches and Examples of Applicable Key Methodologies

3.1 Audio Experience Methodologies

In the following section, several established and new listening test evaluation methods are presented and compared, to explain which method is needed where and to mention the areas with which they cannot be applied, and/or identify questions they cannot answer.

ITU-R BS.1534[16] "Methods for the subjective assessment of intermediate quality level of audio systems" is a Multi-Stimulus test with Hidden Reference and Anchor (MUSHRA) with a given reference. It is probably the most used method for quality evaluation of audio codecs. Here, the reference is an unprocessed, original audio item that is clearly defined. The test subject makes a relative judgement between the reference and the different processed test items on a continuous 100-point scale with five verbal anchors: bad, poor, fair, good and excellent. The test method is appropriate for evaluating medium and large impairments.

In comparison ITU-R BS.1116[17] "Methods for the subjective assessment of small impairments in audio systems" is specifically designed to evaluate small differences compared to the reference signal. It is a double-blind triple-stimulus test with hidden reference that compares every processed item individually against the reference. It is again, a relative rating against the reference on a five grade scale.

MPEG-H 3D Audio is one of the main distribution audio codecs used in ORPHEUS. As one of the main, final tasks of the standardization effort in ISO/ICE, a verification test was conducted and the report published in January 2017[18]. Four tests were conducted to assess performance of the Low Complexity Profile of MPEG-H 3D Audio. The tests covered a range of bit rates and a range of "immersive audio" use cases (i.e. from 22.2 down to 2.0 channel presentations). Seven test sites participated in the tests with a total of 288 listeners. This resulted in a data set of 15576 individual scores.

For test 1 "Ultra HD Broadcast" uses the bitrate 768 kbit/s, 12 items in the formats 22.2, 7.1+4H+3obj (7 loudspeaker on ear level, 1 LFE, 4 loudspeakers above ears plus 3 objects) and HOA+obj, see Figure 8. This was an ITU-R BS.1116 test with loudspeaker reproduction, evaluated at the sweet spot. The test showed that a bit rate of 768 kbit/s, MPEG-H 3D Audio easily achieves "ITU-R High-Quality Emission" quality, as needed in broadcast applications.

[16] ITU-R Recommendation BS.1534-3, Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems (MUSHRA). 2015, Intern. Telecom Union, Geneva, Switzerland.

[17] ITU-R Recommendation BS.1116-3, Methods for the Subjective Assessment of Small Impairments in Audio Systems. 2014, Intern. Telecom Union, Geneva, Switzerland. p. 33

[18] ISO/IEC N16584, MPEG-H 3D Audio Verification Test Report, JTC1/SC29/WG11, Editor. 2017. http://mpeg.chiariglione.org/sites/default/files/files/standards/parts/docs/w16584_%283D_Audio_Verification_Test_Report%29.docx.

© 2015 - 2018 ORPHEUS Consortium Parties

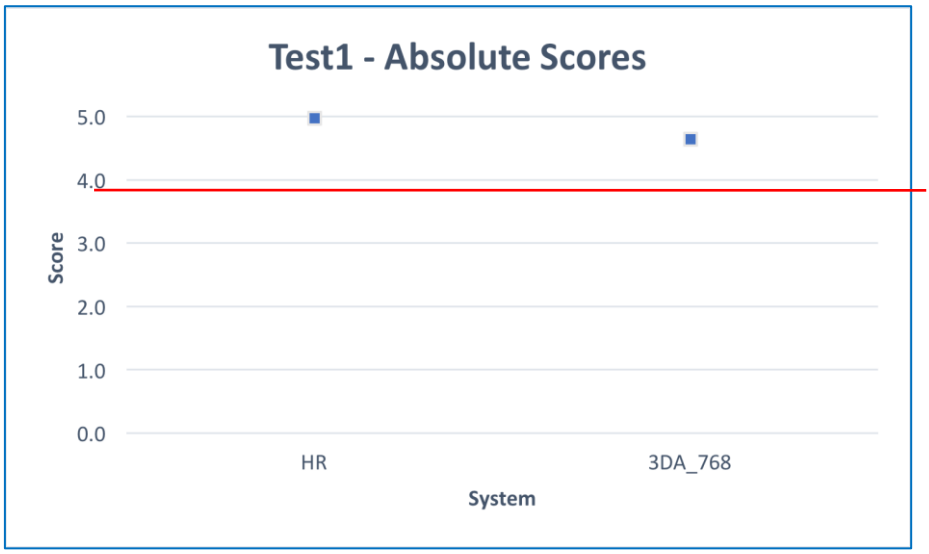


Figure 8: Test 1 "Ultra HD Broadcast", HR hidden reference, 3DA_768 3D audio material in 768 kbit/s, the red line shows the ITU-R requirement for "high-quality emission," i.e. mean value of 4.0, The size of the confidence interval is so small that it is within the size of the marker used for the mean[18].

Test 2 "HD Broadcast" or "A/V Streaming" used three different bitrates: 512, 384 and 256 kbit/s, see Figure 9. Two different anchor signals low pass filtered at 7.0 and 3.5 kHz were used. Twelve items with 7.1+4H channels, 5.1+2H+3obj and HOA format were used. MPEG-H 3D Audio achieved a quality of "Excellent" on the MUSHRA subjective quality scale.

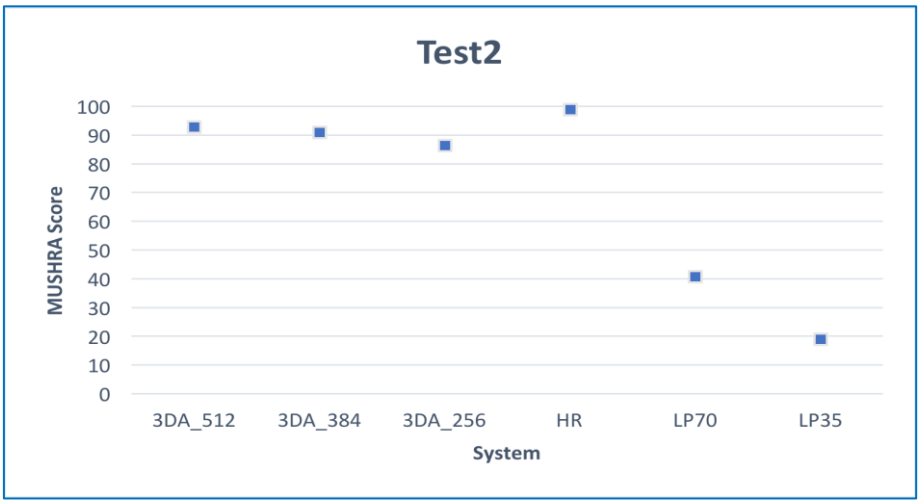


Figure 9: Test 2 "HD Broadcast" or "A/V Streaming". Three different bitrates: 512, 384 and 256 kbit/s. The size of the confidence interval is so small that it is within the size of the marker used for the mean[18].

In test 3 "High Efficiency Broadcast" the following formats are used: 5.1+2H, 5.1, 2.0, HOA, the bitrates for the different formats are listed in Table 1, the results are in Figure 10. The test showed that for all bit rates, MPEG-H 3D Audio achieved a quality of "Excellent" on the MUSHRA subjective quality scale

	Formats	**5.1+2H**	**5.1**	**2.0**	**HOA**
1	Hidden reference				
2	3D Audio	256 kb/s	180 kb/s	80 kb/s	256 kb/s
3	3D Audio	192 kb/s	144 kb/s	64 kb/s	192 kb/s
4	3D Audio	144 kb/s	128 kb/s	48 kb/s	144 kb/s
5	Anchor 1				
6	Anchor 2				

Table 1: Bitrate and format combinations for test 3.

© 2015 - 2018 ORPHEUS Consortium Parties

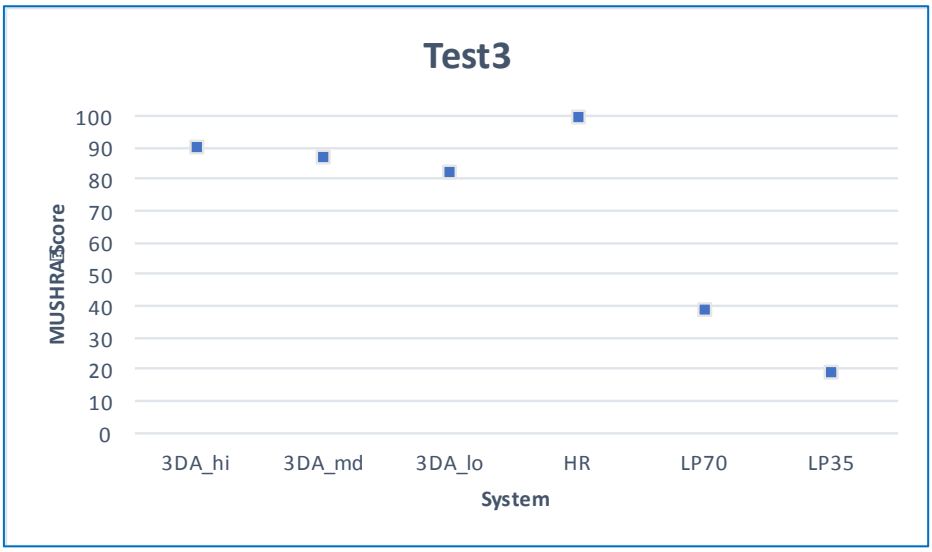


Figure 10: Test 3 "High Efficiency Broadcast". For formats and bitrates see Table 1. The size of the confidence interval is so small that it is within the size of the marker used for the mean[18].

Test 4 measured performance for the “Mobile” use case, in which audio material was coded at 384 kb/s, and presented via headphones. The MPEG-H 3D Audio FD binauralisation engine was used to render a virtual, immersive audio sound stage for the headphone presentation. The 384 kb/s bit streams from test 2 were used. The test showed that at 384 kb/s, MPEG-H 3D Audio with binauralisation achieved a quality of “Excellent” on the MUSHRA subjective quality scale, see Figure 11.

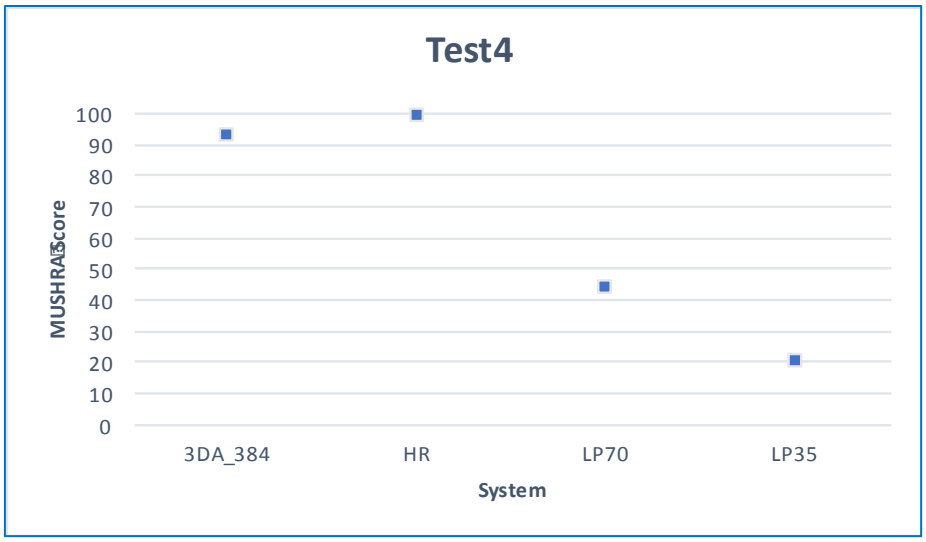


Figure 11: Test 4 “Mobile” [18]

Not every audio quality evaluation test has a clear reference, e.g. binaural reproduction over headphone. Here the normal in-head localised stereo down-mix is one extreme and the reproduction of the content over a good loudspeaker setup in a good listening room is the other extreme. The binaural reproduction lies in between these two extremes. The discussion about how to make such an evaluation of a binaural reproduction and how to evaluate the listening test results with cluster analysis can be found in[17]. In this evaluation, the standard down-mix signal is just a comparison signal and the evaluated renderers can be better or worse when compared to it. Additionally, the benefit of clustering the results is illustrated here. Over all listeners and all items, there was no difference between the two binaural reproduction methods compared to the down-mix signal. But there were clearly two clusters of listeners: the so-called “binaural lovers” and “down-mix lovers”, see Figure 12.

© 2015 - 2018 ORPHEUS Consortium Parties

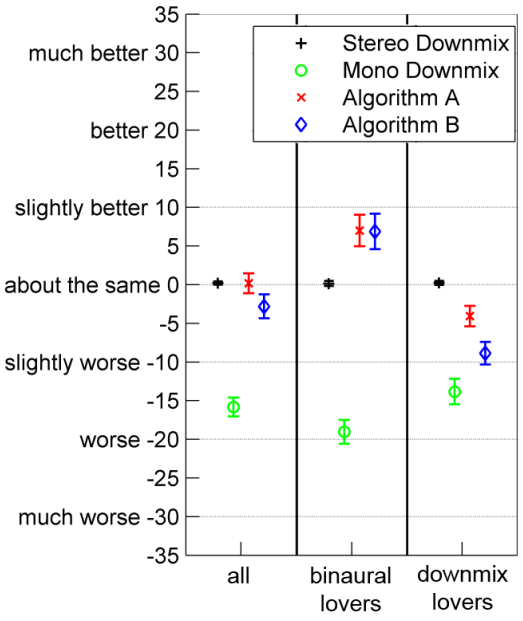


Figure 12: Using a bi-polar scale and cluster analysis for comparison of binaural renderers to a standard downmix of 5.1 items [19]

The usage of a reference and its influence on the evaluation of different 3D loudspeaker reproduction systems can be significant at the upper end of the scale, see Figure 13 (taken from[20]). With a given reference, there is a relative rating compared to the reference, judged at 100 points. Without this reference, there is an absolute rating, where the verbal anchors on the scale, the used anchor signal (here stereo downmix D2) and the internal references of the listeners define the ratings of the subjects.

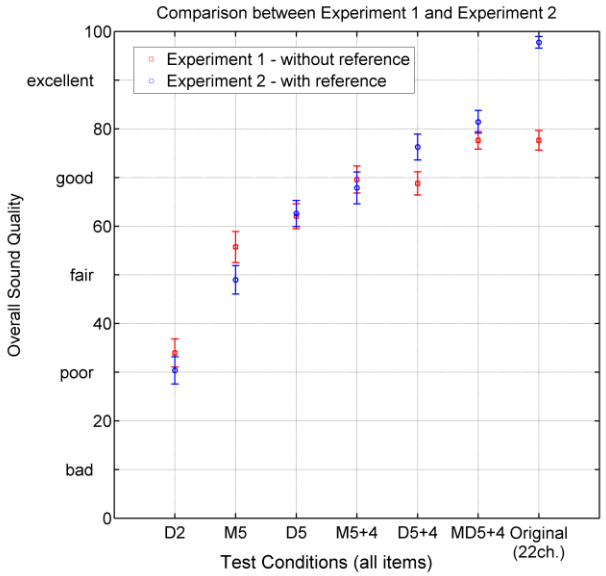


Figure 13: Comparison between different loudspeaker reproduction setups, with and without reference; x+y: x loudspeaker on ear height, y elevated loudspeakers, D downmix, M leaving out channels [20]

[19] Silzle, A., et al. Binaural Processing Algorithms: Importance of Clustering Analysis for Preference Tests. 126th AES Convention. 2009. Munich, Germany, preprint #7728

© 2015 - 2018 ORPHEUS Consortium Parties

There is a new reference-less method proposed at ITU-R: Multiple Stimulus Ideal Profile Method (MS-IPM)[21]. This method is useful for cases like the above mentioned which possess no clear reference. In this method, additional attribute ratings give further insight, see Figure 14. It has also the advantage that the test subjects can define ideal points for these attributes, which may not lie at the end of the scale. Such an ideal profile provides a new insight towards the preference of test subjects.

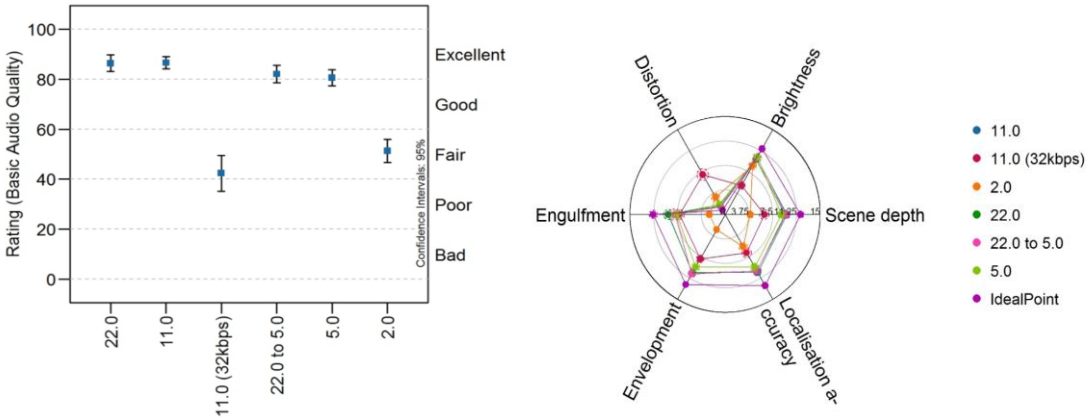


Figure 14: Overall system basic audio quality ratings for different loudspeaker reproduction setups; b) Spider plot of attribute rating with 95% confidence circles, from[21]

With the above-mentioned methods, the Basic Audio Quality (BAQ) of the technical implementation of an algorithm, unit, or setup is evaluated by the test subjects. There is the scientific discussion about how much important this "technical quality" is for the end user, or how much the content, the accessibility, the price and other issues are important. The newly introduced Overall Listening Experience (OLE) measures the overall satisfaction of user experiences or Quality of Experience (QoE) when being presented with a service or technology. In[22] "the participants were instructed to take everything into account what is personally important to them". Moreover, they were asked to rate the stimuli according to how much they enjoy (or like) listening to them. Following this interpretation of the term OLE, OLE can be seen as a term used to describe someone's experiences considering all aspects in the context of listening to something. Moreover, a rating reflects to which degree such experiences are perceived positively. The more an audio system fulfils all expectations and needs of a listener, the more he or she appreciates this listening experience. To this end, ratings of the OLE, or the "Quality of OLE", are to some degree equivalent to QoE in the context of listening." A comparison between OLE and the standard Basic Audio Quality (BAQ), as used in ITU-R BS.1534 for different reproduction systems is given in Figure 15.

[20] Silzle, A., et al. Investigation on the Quality of 3D Sound Reproduction. International Conference on Spatial Audio (ICSA). 2011. Detmold, Germany.

https://www.researchgate.net/publication/259974979_Investigation_on_the_quality_of_3D_sound_reproduction.

[21] Zacharov, N., et al. Next Generation Audio System Assessement Using the Multiple Stimulus Ideal Profile Method. 8th International Conference on Quality of Multimedia Experience (QoMEX). 2016. Lisbon, Portugal.

[22] Schoeffler, M., A. Silzle, and J. Herre, Evaluation of Spatial/3D Audio: Basic Audio Quality Versus Quality of Experience. IEEE Journal of Selected Topics in Signal Processing, 2017. 11(1): pp. 75-88, DOI: 10.1109/JSTSP.2016.2639325.

© 2015 - 2018 ORPHEUS Consortium Parties

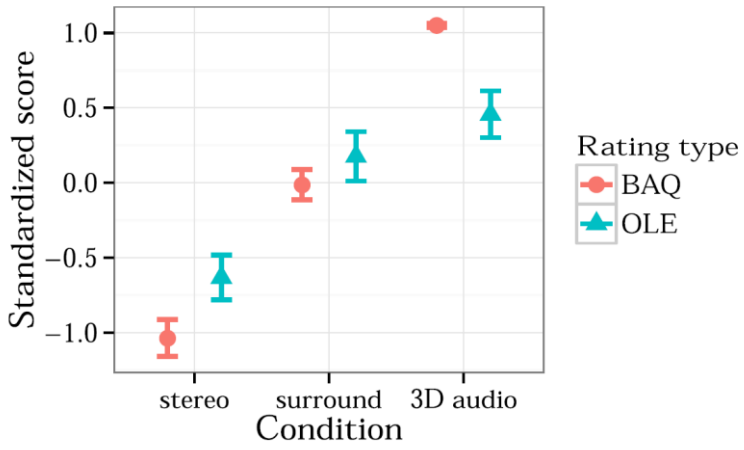


Figure 15: Comparison between Basic Audio Quality (BAQ) and Overall Listening Experience (OLE) for different reproduction systems[22].

Table 2 lists the above mentioned listening test methods with their main features.

Method or Standard	External Reference	Scale	Measuring	Evaluation method
ITU-R BS.1534 (MUSHRA)	yes	0..100	BAQ	Multi-stimulus
ITU-R BS.1116	yes	1..5	BAQ	A-B comparison
MS-IPM	no	0..100	BAQ	Multi-stimulus
OLE	no	5 grade Likert scale	OLE	Absolute rating

Table 2: Comparison of the above mentioned listening test methods

In[23] a general overview about the process of evaluating 3D audio is given: what kind of Quality Elements (QE) in the physical domain can influence which kind of Quality Features (QF) or attributes in the perceptual domain, see Figure 16. There is a multivariate relationship between the Quality Elements and Quality Features, i.e. every element can influence every feature, and every feature can be influenced by every element. In a listening experiment, it is only possible to evaluate the dependency of a small number of QEs to very few QFs or only the Quality of Experience (QoE), which is the BAQ or OLE, dependent on the test method. Therefore, a method using an expert survey was developed to get an overview about this multivariate relationship and identify the most important aspects for a given application, to then select them for listening test evaluation[24]. In the Quality of System block, the decision between the amount of effort, e.g. available computational power on the target platform, and the wished QoE the Basic Audio Quality is done.

As examples for these multivariate relations between QEs and QFs Table 3 quantifies and sorts them by importance for the application of a ***home reproduction with loudspeakers***, Table 4 for a ***mobile reproduction with headphones***. Six experts participated in the expert survey.

[23] Silzle, A. 3D Audio Quality Evaluation: Theory and Practice. 2nd International Conference on Spatial Audio (ICSA). 2014. Erlangen, Germany.

[24] Silzle, A., Generation of Quality Taxonomies for Auditory Virtual Environments by Means of Systematic Expert Surveys. Institut für Kommunikationsakustik, Ruhr-Universität Bochum, Doctoral Dissertation. 2007, Shaker Verlag, Aachen. ISBN 978-3-8322-7040-7

© 2015 - 2018 ORPHEUS Consortium Parties

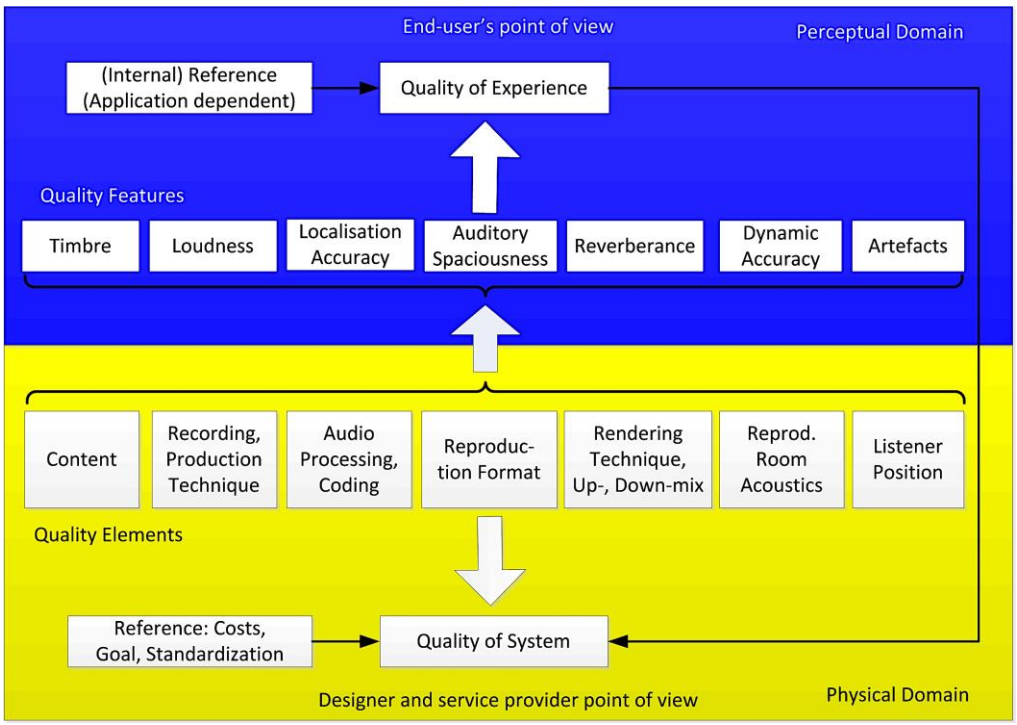


Figure 16: Quality Taxonomy for 3D audio systems, with the Quality Elements in the physical domain and the Quality Features in the perceptual domain. The Quality of Experience and the Quality of System are comparisons between their input parameters and the given references[23]

Quality Features	1,8	1,9	1,9	2,0	2,1	2,1	2,5	average
Auditory spaciousness	2,3	2,8	2,3	3,0	2,2	3,0	3,2	2,7
Localisation accuracy	1,7	3,3	3,7	4,0	1,4	1,6	2,8	2,6
Timbre	2,3	2,2	2,0	1,3	2,2	2,6	2,8	2,2
Reverberance	1,7	1,4	1,2	1,2	1,4	3,0	3,2	1,9
Dynamic accuracy	1,7	1,0	1,4	1,4	1,6	1,8	2,0	1,6
Artefacts	1,0	0,4	0,8	0,8	3,6	0,8	1,0	1,2
Quality Elements	Content	Position of loud-speakers	Listener position	Number of loud-speakers	Audio processing, coding	Reproduction room acoustics	Recording technique, mic setup	

Table 3: Quality Element to Quality Features relation matrix for the application home reproduction with loudspeakers, with sorted average values over rows and column. Categories of QE-to-QF relations: No relation (0), Less important relation (1), Important relation (2), Very important relation (4)[23]

Quality Features	1,5	1,7	2,0	2,2	2,4	2,6	average
Localisation accuracy	1,4	3,2	1,8	3,2	3,0	3,5	2,7
Timbre	1,8	0,4	2,3	4,0	2,8	2,8	2,3
Auditory spaciousness	1,6	2,0	1,8	2,3	3,0	3,3	2,3
Artefacts	1,2	2,4	3,5	1,8	0,8	2,7	2,0
Reverberance	2,6	0,8	1,5	0,8	2,5	2,0	1,7
Dynamic accuracy	0,5	1,4	1,5	1,3	2,3	1,7	1,4
Quality Elements	Content	Head tracking	Audio processing, coding	HRTFs	Recording technique, mic setup	Binaural Rendering	

Table 4: Quality Element to Quality Features relation matrix for the application mobile reproduction with headphones, with sorted average values over rows and column. Categories of QE-to-QF relations: No relation (0), Less important relation (1), Important relation (2), Very important relation (4)[23]

© 2015 - 2018 ORPHEUS Consortium Parties

3.2 Application usability experience methodologies

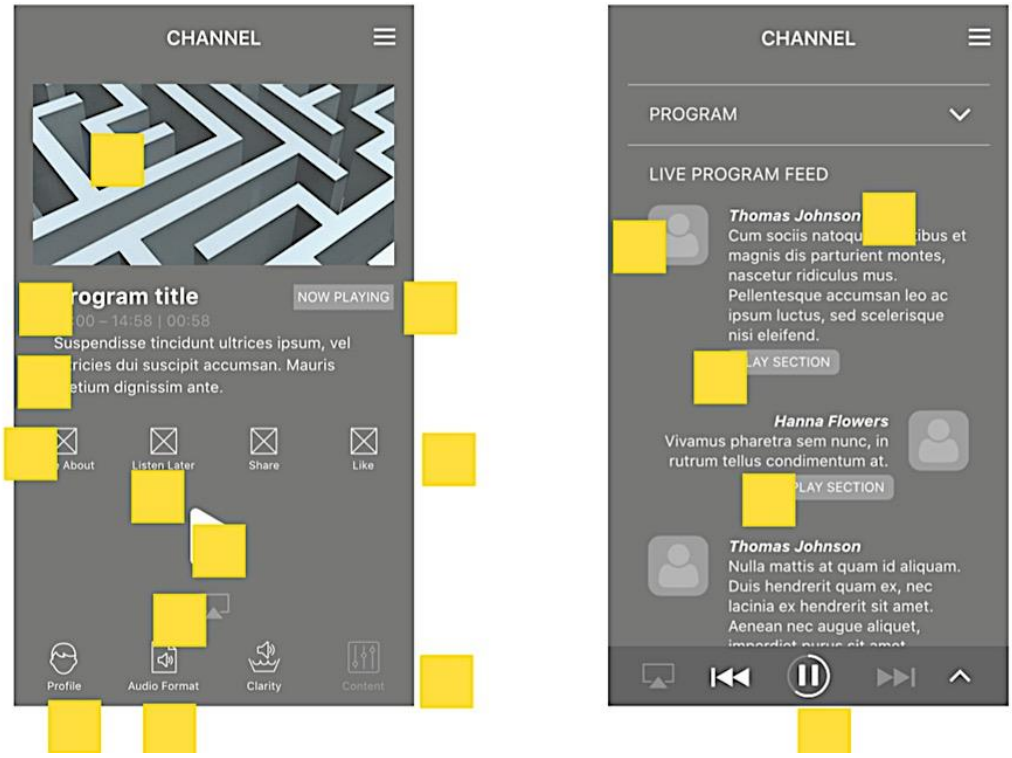


Figure 17: Heuristic evaluation in designing the ORPHEUS app

To evaluate the quality of the information provided by the mobile interface and its usability, various methods and evaluations have been implemented: a heuristic evaluation including experts from UX design and user tests involving real users.

Benchmark or "competitor analysis" aims to create a comparative study of competitors' products which exist on the market today. Ergonomic benchmarking can be used as an appropriate way to set up best practises in the evaluation of the technological aspects of the selected products. The basic principle at work here is to study similar products and applications so as to understand the best and worst elements which they display at the present time.

This benchmark is a comparative study of Object-based broadcasting, and more generally of the 3D audio field, consisted of a usability audit.

Heuristic evaluation is a usability inspection method. It helps to identify usability problems in the user interface design. Based on Jacob Nielsen method[25] a heuristic evaluation is based on a number of standards, criteria, and good practices for evaluating an interface. b<>com, in close collaboration with ECandy, was able to implement this type of method within the ORPHEUS project. More specifically, an iterative heuristic design and evaluation process have been introduced. Concretely, ECandy has developed many mockups, which have been evaluated in collaboration with b<>com. Figure 17 shows an example of mockups and elements that were discussed within the project.

The benchmark methodology performed by b<>com is composed of the following steps:

- Needs analysis
- Individual meetings with the partners
- Identification of the relevant products and features
- Census of known products

[25] Nielsen, J. (1994). Heuristic evaluation. In Nielsen, J., and Mack, R. L. (Eds.), Usability Inspection Methods, John Wiley & Sons, New York, 25-64.

© 2015 - 2018 ORPHEUS Consortium Parties

- Listing of main features
- In depth product search
- Internet research with different keywords:
    - 3D audio
    - Object-Based-Audio (OBA)
    - Spatial sound
    - Surround sound, etc.
- Selection and screenshots of relevant existing interfaces
- Global ergonomic inspection of interfaces
- Detailed analysis of relevant interfaces according to the main features
    - Program control
    - Sound Control / Audio quality
    - 3D control sound
    - Data / Metadata content
    - Interactivity (non-linear playback)
- Deliverable writing

In order to analyse more generally how the information was presented to users, based on 18 heuristics and recommendations (Bastien and Scapin[26], ten heuristics of Colombo and Pasch[27]), this benchmark identified 53 existing relevant interfaces.

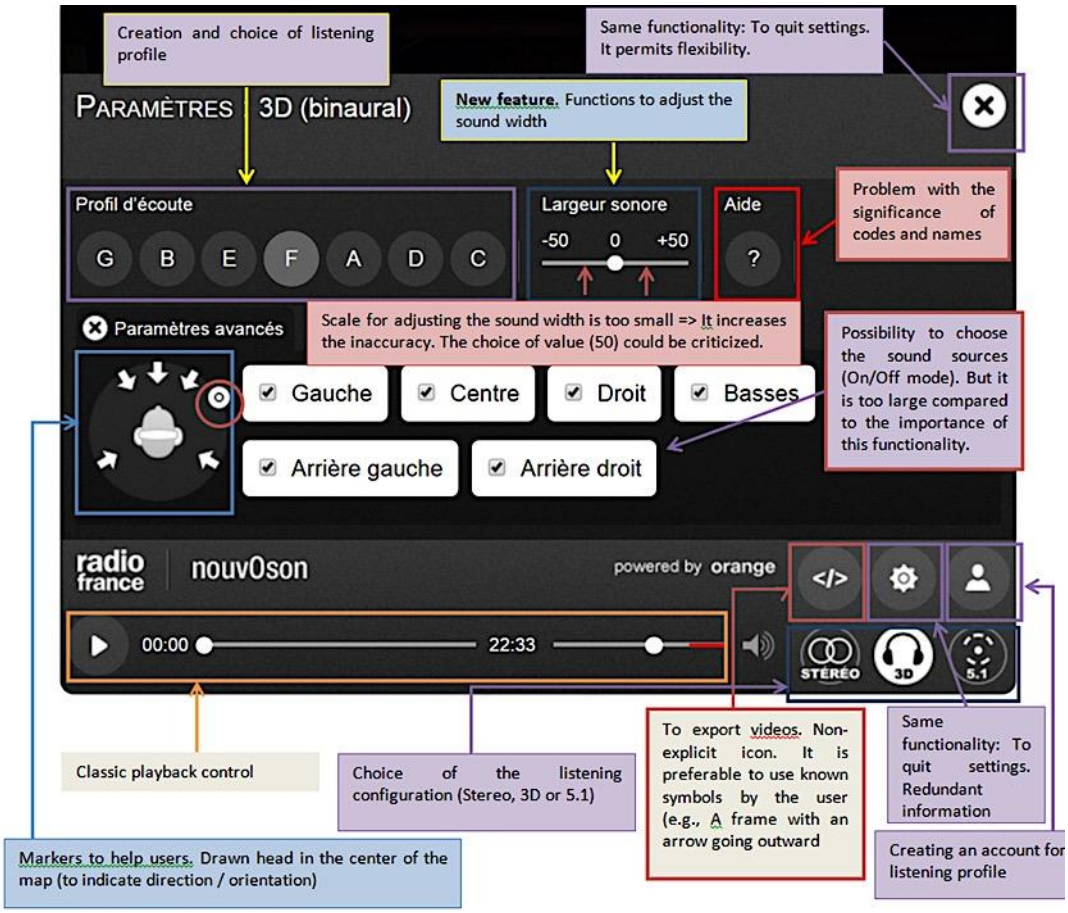


Figure 18: Example of an aspect of an analyzed interface

[26] Bastien, J.M.C., Scapin, D. (1993) Ergonomic Criteria for the Evaluation of Human-Computer interfaces. Institut National de recherche en informatique et en automatique, France

[27] Colombo, L. and Pasch, M. (2012). 10 Heuristics for an Optimal User Experience. Proceedings of the CHI'2012 Altchi. ACM Press.

© 2015 - 2018 ORPHEUS Consortium Parties

Each important interface has been analysed. Moreover, different objectives have constituted this study. First it was important to highlight new features that could be potentially interesting for our project. Then, we had to investigate how information is presented on the interfaces, how it is possible to interact with these products, what is the usability, etc. Around twenty interfaces have been more specifically analysed. These analyses have been made for 3D audio interface and the other interesting interfaces. For this study screenshots of different interfaces have been analysed. For each screenshot, many annotations have been inserted. For more readability, a colour code has been used, according to the following 10 main categories of remarks:

- Guidance
- Workload
- Explicit control
- Adaptability
- Error management
- Consistency
- Significance of codes
- Compatibility
- Bad practices / Errors
- Remarks

The main goal was not to compare sound rendering between applications, nor to make subjective tests in this first step. It was necessary to identify new features which could be interesting for our project. It was also useful to understand the way in which information was presented to users and how they can interact with the main features, etc. This benchmark faced many difficulties since the interfaces came from very different devices and from very different fields with very few similar interfaces to our own. This difficulty was increased by the fact that many interfaces were not available to be tested directly. Testing had to be done, therefore, via screenshots. Unfortunately, these heterogeneous interfaces and access difficulties (obligation to purchase, inaccessibility in some countries, apps in development, etc.) made it impossible to perform user testing which is very useful for robust comparisons between applications. The solution was chosen to make heuristic evaluations, focusing on the main objective: to observe and make a record of the latest existing features in this area, so as to better understand how and why they were implemented.

For similar products to those proposed by the ORPHEUS project, it is nonetheless interesting to focus on the types of interaction / interfaces which sometimes may not seem innovative. Many conventional solutions are already available to users. Many standards exist on how to control playback, change radio and playlist, etc. We should probably be building products based on these standards, conventions and codes to speed up the handling of any future interface and to minimize the learning curve of future users. Innovations can achieve their full potential through this dual position: offer a new audio experience while relying on the conventions that users know. Colombo and Pasch (2012) use the term of “innovative appropriation”. For them, users should be able to customize and manipulate the system according to their characteristics and preferences. They should feel at ease with the system as if it had been designed specifically for them, a system which mixes the familiar and the innovative.

3.2.1 Heuristic evaluation

For heuristic evaluation, different tools were used: iOS Human Interface Guidelines and Bastien and Scapin criteria.

6 main heuristics compose the iOS Human Interface Guidelines:

- Aesthetic integrity
- Consistency
- Direct Manipulation
- Feedback
- Metaphors
- User control

© 2015 - 2018 ORPHEUS Consortium Parties

Consistency, for example, can be seen as implementing familiar standards and paradigms by using system-provided interface elements, well-known icons, standard text styles and uniform terminology. The app must incorporate features and behaviours in ways people expect. It must also for the same action use the same information and the same words.

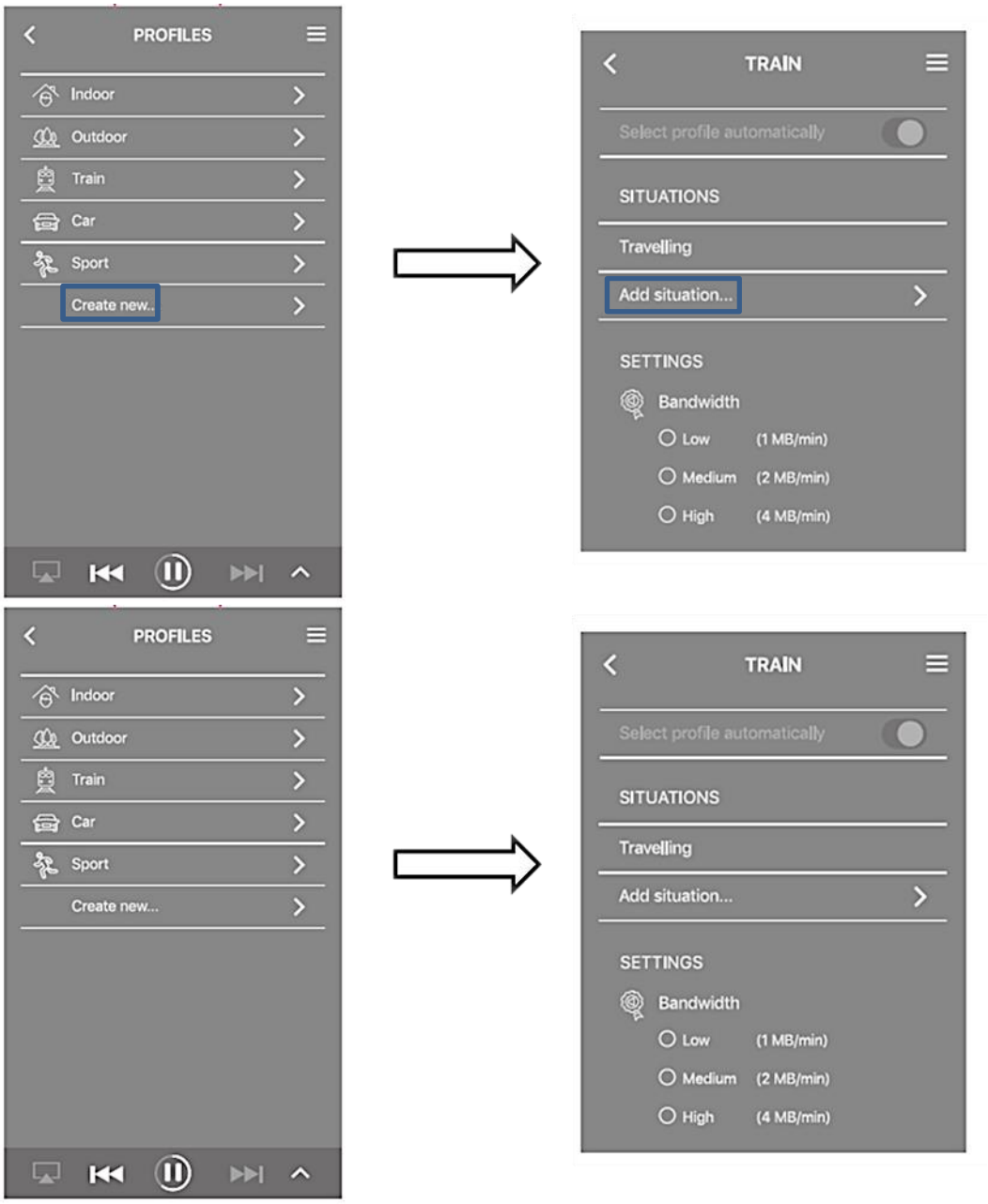


Figure 19: Example of remarks related to heuristic evaluation (here, the objective is to homogenize the terms used, referring to the same user actions)

Moreover, we based our analysis on the research developed by Bastien and Scapin via their list of 18 elementary criteria, and including their 8 main criteria.

3.2.2 User Tests

User tests refer to User Centric Designs (UCD approach). UCD approach is based on 6 principles:

- Base design on explicit understanding of users, tasks and environments
- Involve users in design and development
- Orient the design through user-centric evaluation
- Iterate until the desired result is achieved

© 2015 - 2018 ORPHEUS Consortium Parties

- Cover the user experience in its entirety
- Establish a multidisciplinary team.

The objectives for this first test are multiple:
- Get user feedback;
- Acquire information about usability;
- Test A and B scenarios;
- Etc.

In this first user test, the objective is to compare two versions A and B, provided by ECandy. In terms of methods, there are two possibilities for comparisons in user tests: between-individual comparisons or within-individuals comparisons. In our study, the between-individual comparison was chosen, each tester using only one scenario.

Two experimental groups were created. Each participant was assigned to an experimental group. The first group represents the scenario A. The second group represents the scenario B.
In scenario A, the situation is defined within the profile. A situation is a combination of conditions which determine if a profile is selected automatically. So when a user creates or edits a profile, he first defines the conditions of a situation and then adjusts the audio settings. In this scenario, profile and situation can be defined in one screen. Situation has to be defined separately for each profile.
In scenario B, situations and profiles are considered as different sections. When creating or editing a profile, the user needs to choose from a list of already defined situations. So when someone wants to create a new profile, at least one situation needs to be set already. Situations are defined separately. One or more situations can be connected to a profile. According to scenario B, someone can apply a profile to different situations, which can be a relevant scenario. In scenario A the user has to set a situation for each profile independently.

Quantitative analysis of user interfaces
Quantitative analysis can be executed by special tools which are incorporated in the mobile app. These tools register all kinds of user interaction within the app, being converted into data which can be analysed afterwards.
Typical user interaction registration consists of:
1. heatmaps
2. user session recordings
3. conversion funnels

A heatmap is a graphical representation of user's clicking, tapping and scrolling behaviour on an interface. It gives insight in the quantitative use of certain interaction elements within a screen. These statistic data can tell whether elements are recognizable and well positioned.
User session recordings show the complete user journey through the app, including time spent on every screen. This offers an in-depth view on user behaviour within the app.
Conversion funnels are meant to test usability of delimited, segmented tasks within the app. It gives statistic insight in how many users complete a task or where they abort it. These statistics will help to identify points for improvement.
Data from user interaction registration offer a vast collection of useful statistic information on the mobile app's usability. Due to the extent of data, focussing on certain functions or tasks within the app is necessary. For the ORPHEUS mobile app, at least the following aspects should be investigated by quantitative evaluation:
- channel screen interaction (adjusting the play head, reading live transcript, etc.)
- profile creation/editing
- use of options to adjust audio rendering (audio format, audio clarification, translation, etc.)
- use of certain settings at specific locations or activities

Results of quantitative evaluation can be used in an iterative process where weak aspects of an application are improved during future design cycles.

© 2015 - 2018 ORPHEUS Consortium Parties

3.2.3 Participants and Target Groups

Overall, the main focus for the Uses and Acceptability laboratory at b<>com, in the ORPHEUS Project, is to assess the quality of the information and ease of use of the provided interfaces (mobile application). In this objective, different methodologies are implemented in close collaboration with ECandy (benchmark, heuristic evaluation, user tests, etc.). Benchmark and heuristic evaluation involve UX design experts. The purpose is to be able to evaluate interfaces quickly and efficiently in the first steps of design. Once these first iterations have been carried out, other methodologies such as user tests make it possible to involve more deeply the end users. So, from the initial design stages, b<>com has involved potential end users to test the mobile application developed by ECandy.

3.2.4 Planning

Several prototypes of applications are planned within the ORPHEUS project. It is expected that each major development within the prototypes will be subject to a specific evaluation. This evaluation will take place via user tests, in order to involve the potential and future end user/consumer in the interface evaluation.
Overall, in the ORPHEUS project, at least 2 user tests are still scheduled:

- Test 2: User Test with more features (audio, etc.);
- Test 3: Final user test.

© 2015 - 2018 ORPHEUS Consortium Parties

3.3 Testing user experience in the browser: BBC TASTER

3.3.1 Case study: BBC Proms in Binaural

Figure 20: *BBC TASTER page for BBC Proms in Binaural tests*

This pilot was described as "Exclusive immersive 3D audio mixes giving you the full binaural experience through your regular headphones! Short of being there this has to be the closest you can get to the Proms."

Analytics from the pilot show basic data including how many people tried it, when, on what kind of device, from which geographical area, as well as their rating, as shown in Figure 21.

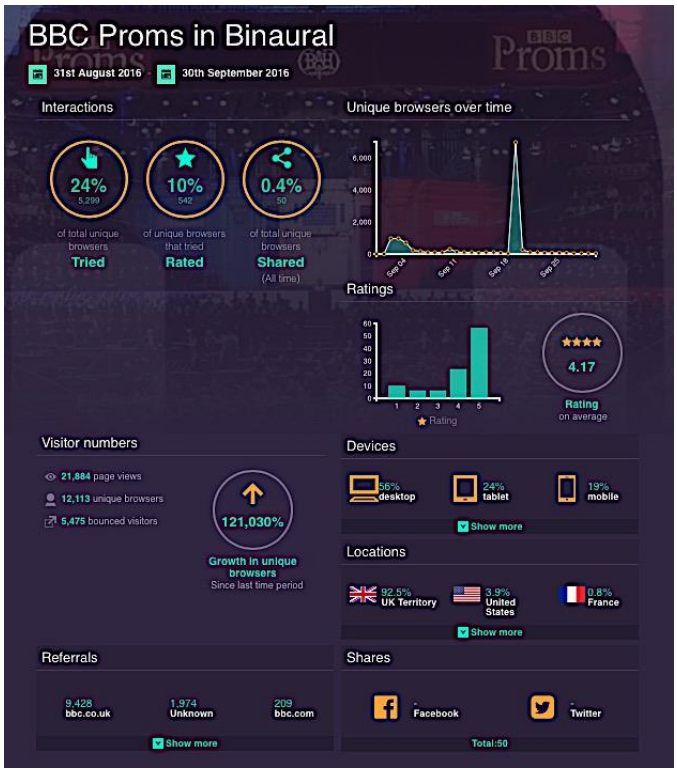


Figure 21: Basic analytics data from BBC Proms in Binaural pilot on BBC Taster

A questionnaire included in the pilot was used to gather opinions and some technical information, as shown in Figure 22.

© 2015 - 2018 ORPHEUS Consortium Parties

Figure 22: Responses to questionnaire in BBC Proms in Binaural pilot on BBC Taster

3.3.2 Case study: Essential Classics Interactive

Figure 23: BBC TASTER page for BBC Essential Classics Interactive

This was a classical music radio show[28] where at certain moments during the programme, the presenter invited listeners to choose alternative content to which to listen, if they desired, by pressing a 'swap' button. This gave audience members a choice within their listening experience, yet keeps a conventional linear broadcast central to it. The ability to swap elements of the show enables them to follow a thread of content, but keeping the desired 'lean back' nature of radio.

[28] http://www.bbc.co.uk/taster/projects/essential-classics-interactive/inside-story

© 2015 - 2018 ORPHEUS Consortium Parties

It is interesting to note that in this, and in other Taster pilots that have been studied, how critical it is to think about technology and the editorial concept as symbiotic things. It was seen that concepts that didn't resonate with audiences reduced the impact of the technical advances. Similarly, where technology limited what audiences could do, people just couldn't get into the story. The “Essential Classics” programme gave a great example. Where audiences connected with the content, they tended to enjoy the pilot more overall, while those who didn't enjoy the content at all, would still seek out the same technology format, just with different content, as shown in Figure 24.

Figure 24: Basic analytics from Essential Classics Interactive pilot on BBC Taster

There is more detailed information on location and referrer, when required, as shown in Figure 25

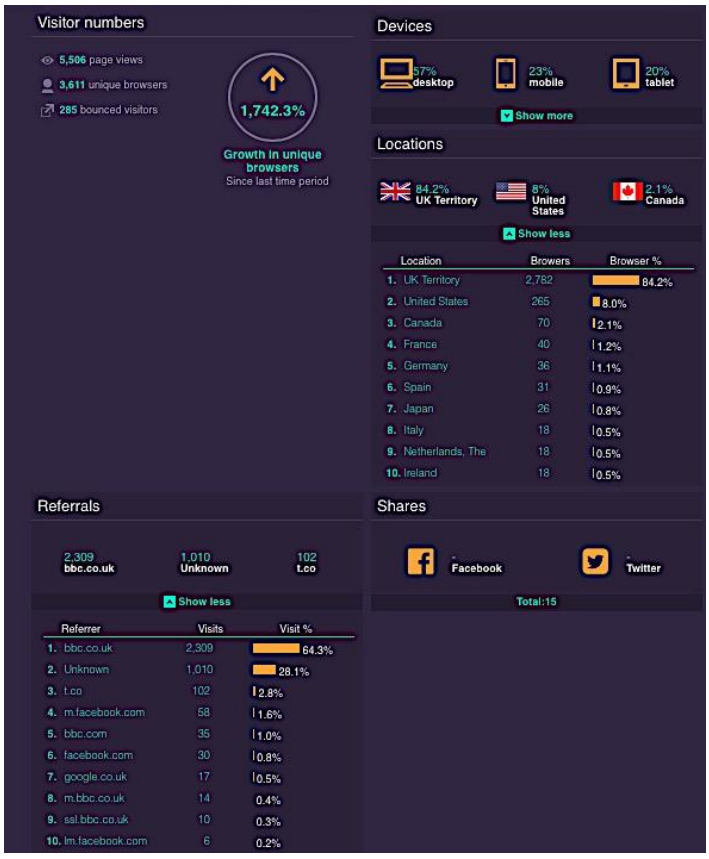


Figure 25: Location and referrer analytics from Essential Classics Interactive pilot on BBC Taster

A basic questionnaire was included in the on-line public pilot, and the responses from people who choose to fill it in are shown in Figure 26.

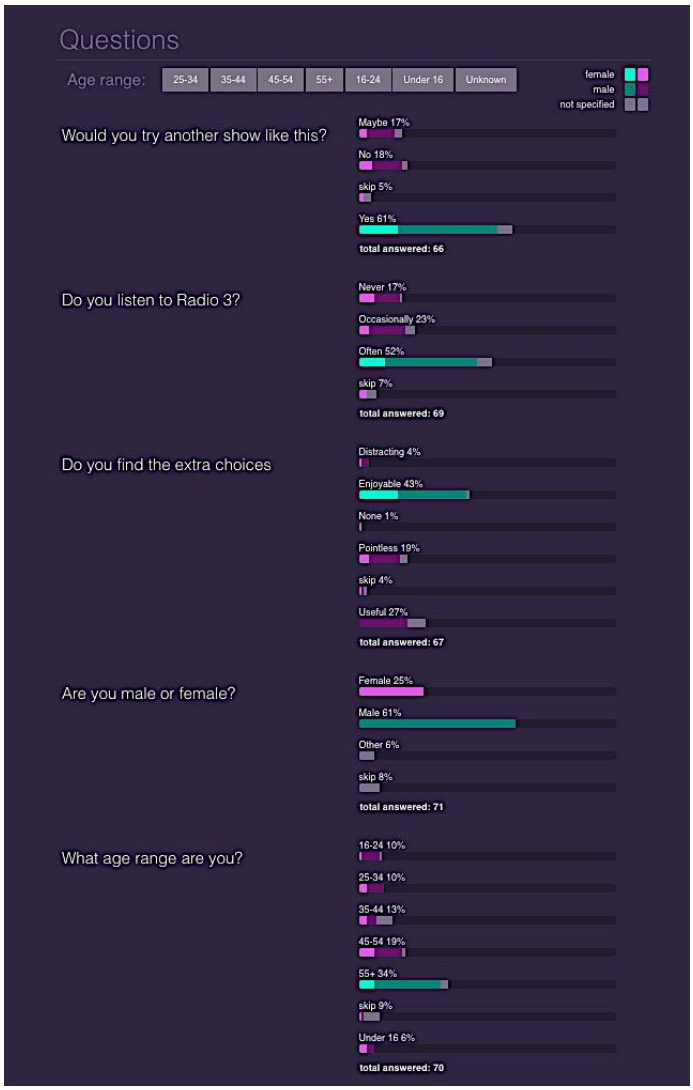


Figure 26: Questionnaire results from Essential Classics Interactive pilot on BBC Taster

Much more detailed information was collected by asking a representative UK sample a longer series of questions, including attitude towards new technology, use of gadgets, social media, demographics of the household, opinion on other companies such as Buzzfeed, YouTube, Netflix, and so on.

3.3.3 Case study: Unearthed

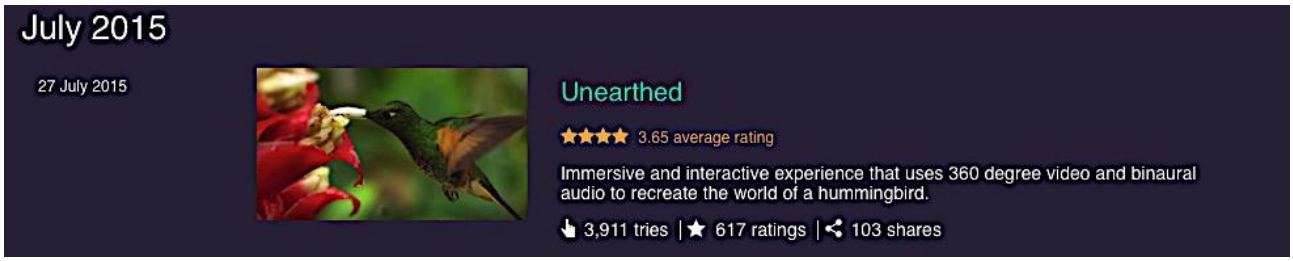


Figure 27: BBC TASTER page for Unearthed

This pilot had the description, “Step into the rainforest as you get up close and personal with the habits of the hummingbird. This interactive pilot proves this bird’s day-to-day is anything but humdrum.”

© 2015 - 2018 ORPHEUS Consortium Parties

"Unearthed"[29] was an interactive and immersive story following a day in the life of a hummingbird, featuring binaural sound and 360˚ video. The project, run through Connected Studio, was a collaboration between the BBC Natural History Unit, BBC R&D's audio team, and an agency called Realise.

To produce the sound for this 360˚ scene, a novel tool was developed that combined BBC R&D's real-time binaural rendering software with 360˚ video rendering for the Oculus DK2 headset. The sound source positions were displayed as overlays on the video, which were presented synchronised to the audio from the DAW. This allowed the engineer to be immersed in the scene whilst mixing it and to explore it interactively as they work.

The basic analytics show a rating of 3.65 out of 5. The questionnaire is aimed at finding out what people thought of the sound and the interactivity, giving some insight into the overall rating, and what they learned (because it had an educational aspect), as summarised in Figure 28.

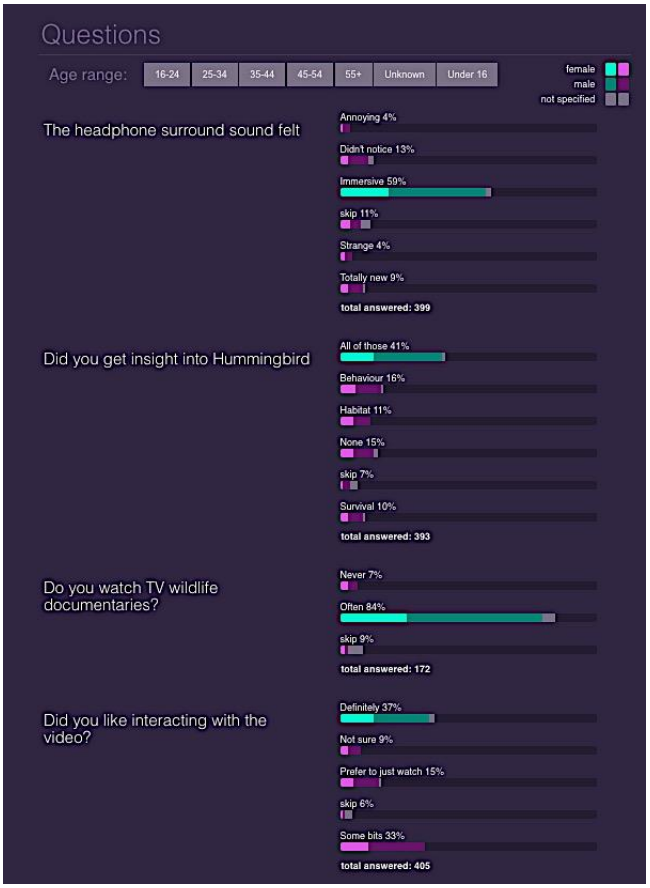


Figure 28: Results of questionnaire of Unearthed pilot on BBC Taster

[29] http://www.bbc.co.uk/rd/blog/2015-07-unearthed-interactive-360-sound-and-video-in-a-web-browser

© 2015 - 2018 ORPHEUS Consortium Parties

4 Defining a Working-model for 'QoE' Assessment in an Object-based Media Eco-system

The new domain of object-based audio production, proliferation, distribution and reception of media bears not only the opportunity but even the necessity to reconsider and reshape existing approaches to examine and evaluate the quality of media experience. Especially the new object-based possibilities for making audio accessible, immersive, interactive and personalized sets-out the various planes to be considered, expanding dramatically the field in examinations of experience in audio quality. Therefore the development of evaluative methodologies, novel and innovative metrics, techniques and tools for schemes of evaluation of user experience has to be addressed – focussing on three simple questions:

- Which quality?
- What experience?
- Who is the end user?

4.1 Basic Quality Features from the Use Case Definitions

In the course of *WP 2.3 Use case definitions* user requests for object-based audio broadcasting were collected. For this purpose, ORPHEUS partners BBC R&D and BR organized several in-house workshops with creative and technical staff, asking them what kind of features and requirements are to be considered as necessary or desirable for next generation audio production and distribution.

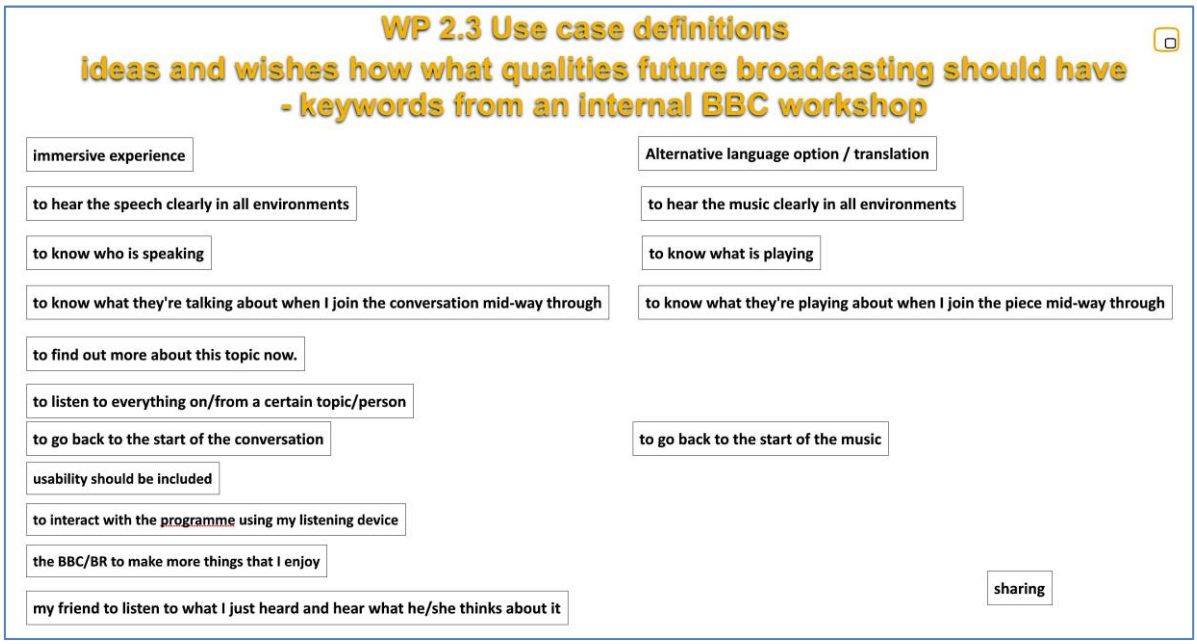


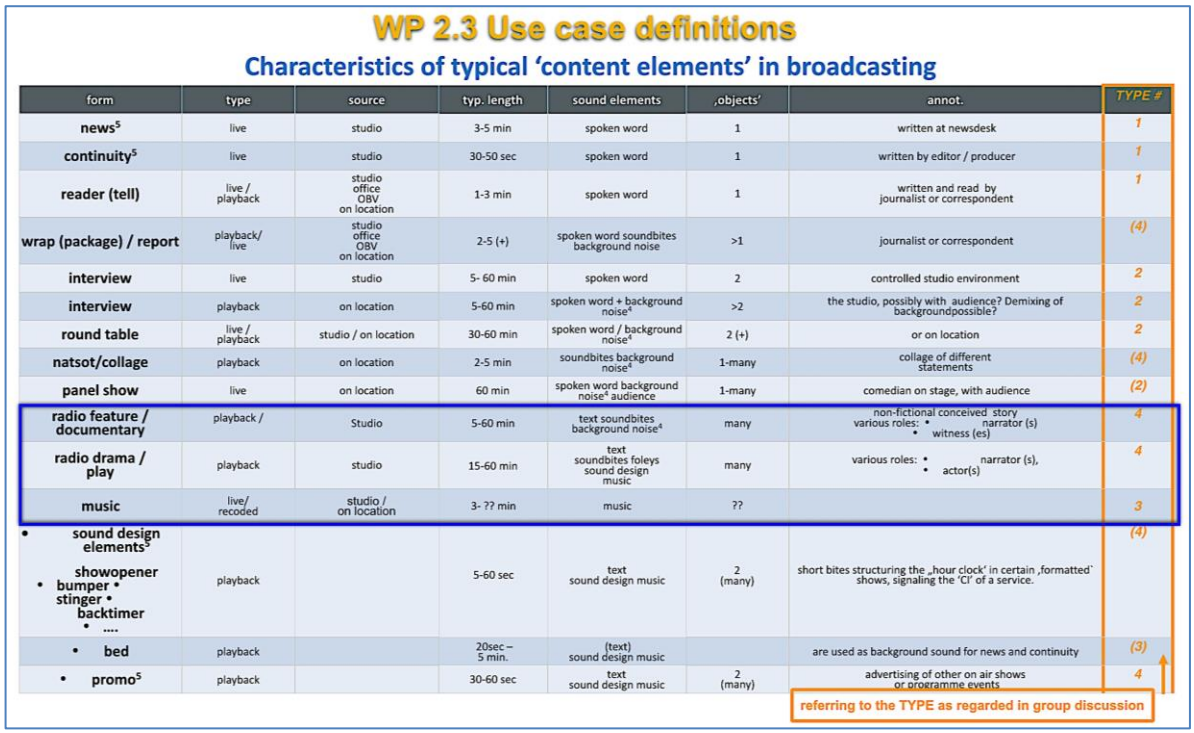
Figure 29: User requests from a use case definitions workshop

The implications of these basic requests were then first enriched in brainstorming sessions, finding applications and more detailed descriptions for practical scenarios.

In a next step, their relation and relevance to existing typical broadcast genres and formats was discussed (types 1-4 defined by the group internally for quick assignment). At this stage, also the implications for

© 2015 - 2018 ORPHEUS Consortium Parties

production and distribution workflows under the predictable change of broadcast infrastructure were considered[30].



WP 2.3 Use case definitions

Characteristics of typical 'content elements' in broadcasting

form	type	source	typ. length	sound elements	‚objects'	annot.	TYPE #
news[5]	live	studio	3-5 min	spoken word	1	written at newsdesk	1
continuity[5]	live	studio	30-50 sec	spoken word	1	written by editor / producer	1
reader (tell)	live / playback	studio office OBV on location	1-3 min	spoken word	1	written and read by journalist or correspondent	1
wrap (package) / report	playback/ live	studio office OBV on location	2-5 (+)	spoken word soundbites background noise	>1	journalist or correspondent	(4)
interview	live	studio	5- 60 min	spoken word	2	controlled studio environment	2
interview	playback	on location	5-60 min	spoken word + background noise[4]	>2	the studio, possibly with audience? Demixing of backgroundpossible?	2
round table	live / playback	studio / on location	30-60 min	spoken word / background noise[4]	2 (+)	or on location	2
natsot/collage	playback	on location	2-5 min	soundbites background noise[4]	1-many	collage of different statements	(4)
panel show	live	on location	60 min	spoken word background noise[4] audience	1-many	comedian on stage, with audience	(2)
radio feature / documentary	playback /	Studio	5-60 min	text soundbites background noise[4]	many	non-fictional conceived story various roles: • narrator (s) • witness (es)	4
radio drama / play	playback	studio	15-60 min	text soundbites foleys sound design music	many	various roles: • narrator (s), • actor(s)	4
music	live/ recoded	studio / on location	3- ?? min	music	??		3
• sound design elements[5] • showopener • bumper • stinger • backtimer •	playback		5-60 sec	text sound design music	2 (many)	short bites structuring the „hour clock' in certain ‚formatted' shows, signaling the 'CI' of a service.	(4)
• bed	playback		20sec – 5 min.	(text) sound design music		are used as background sound for news and continuity	(3)
• promo[5]	playback		30-60 sec	text sound design music	2 (many)	advertising of other on air shows or programme events	4

referring to the TYPE as regarded in group discussion

Figure 30: Typical broadcast content elements –types -

Through this approach, the colloquially uttered basic requests could more clearly be related and assigned to the key-features or USPs of object-based broadcasting: accessibility, interactivity, immersion and personalization.

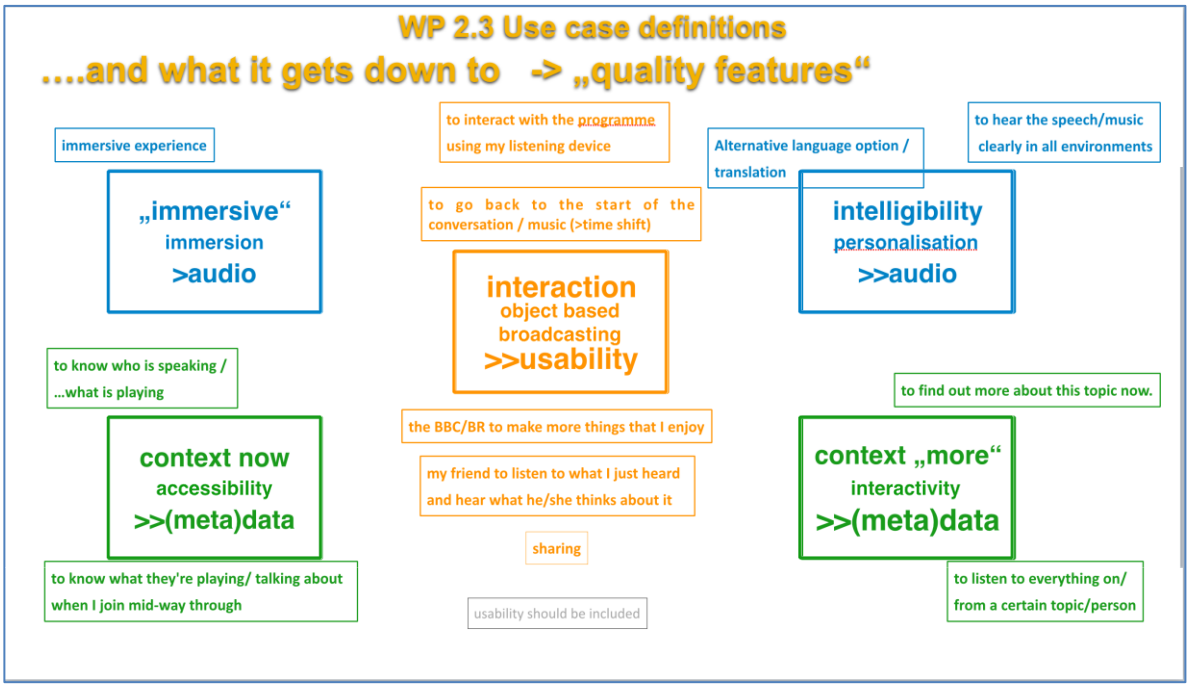


Figure 31: User requests assigned to quality features

[30] as described in ORPHEUS *D3.1 Requirements, designs and workflows of an object- based production environment*

4.2 Generic categories of experience to observe

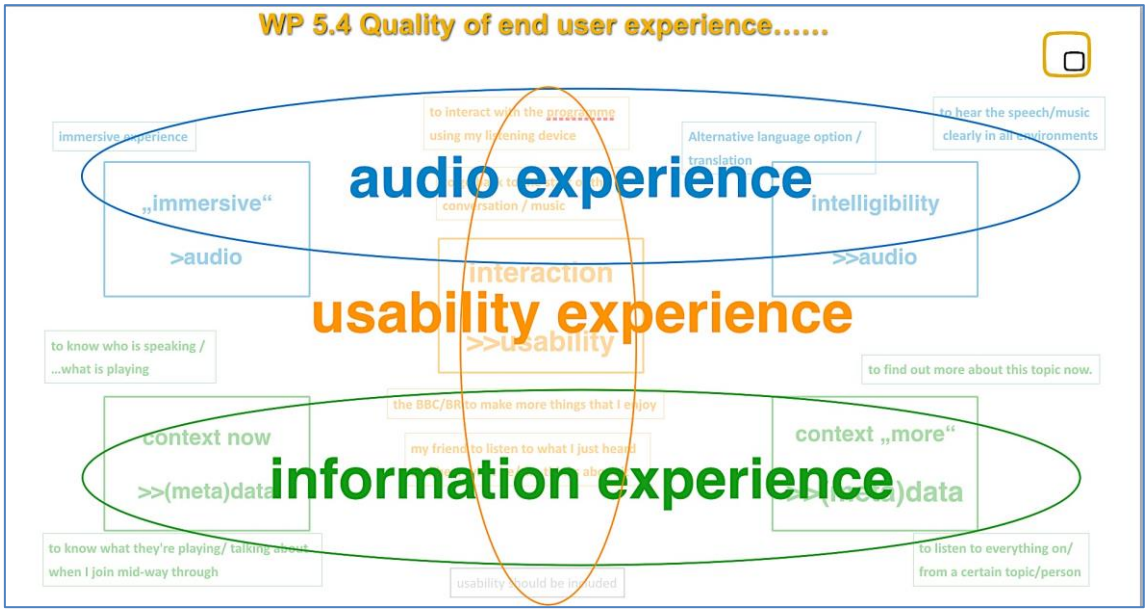


Figure 32: Quality features, characteristics and categories

As consecutive step, the developed schematics was finally transformed into an abstract model, defining three main spheres of qualities to be observed within ORPHEUS' WP 5.4:

- audio experience: referring to the key features 'immersion' and 'intelligibility'
- usability experience: referring to the key feature of human 'interaction' with devices and surfaces
- information experience: referring to the key feature of 'contextual metadata of content'

Proven by an additional brief survey[31], to double-check the relevance of this generic model, all in WP5.4 participating partners of the ORPHEUS have dedicated interests and fields of expertise. This ensures that all fields are covered adequately.

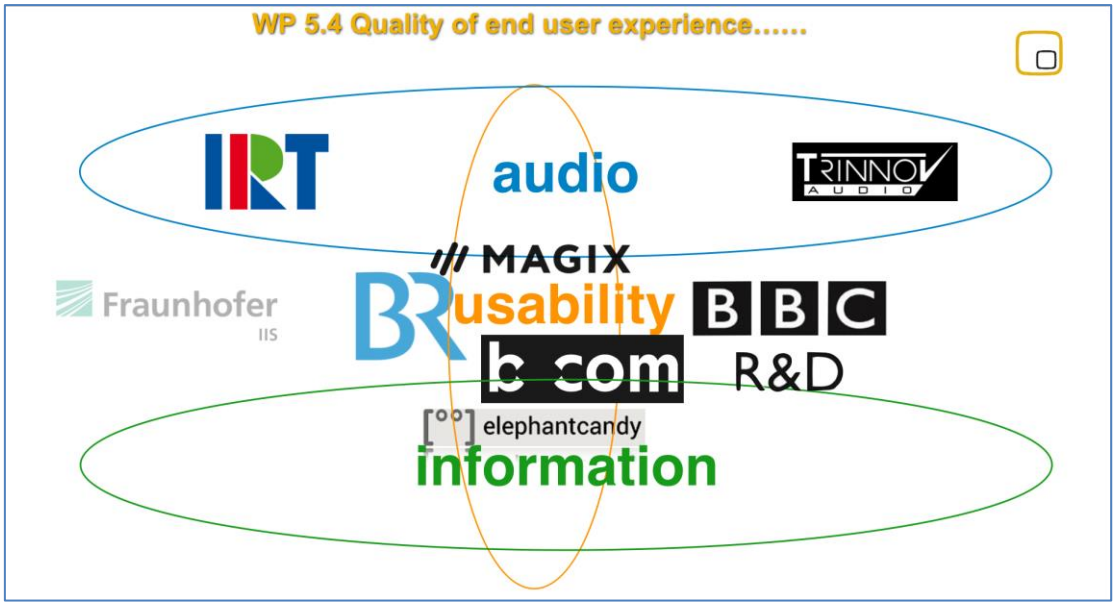


Figure 33: Categories and partners interests and expertise

[31] c.f. Appendix A 1

4.3 General Types of End Users

As the consortium of ORPHEUS is made up of different companies and organizations, each of them has different customers representing their typical 'end user'. Asked in the initial survey who that is, three main types could be identified:

- any user ('end consumer')
- consumer with special interest in audio ('audiophile' but non-professional)
- professional audio user ('sound engineer')

Whereas 'sound engineer' and 'audiophile non-pro' are - to a certain extent also overlapping - subsets of 'any user'.

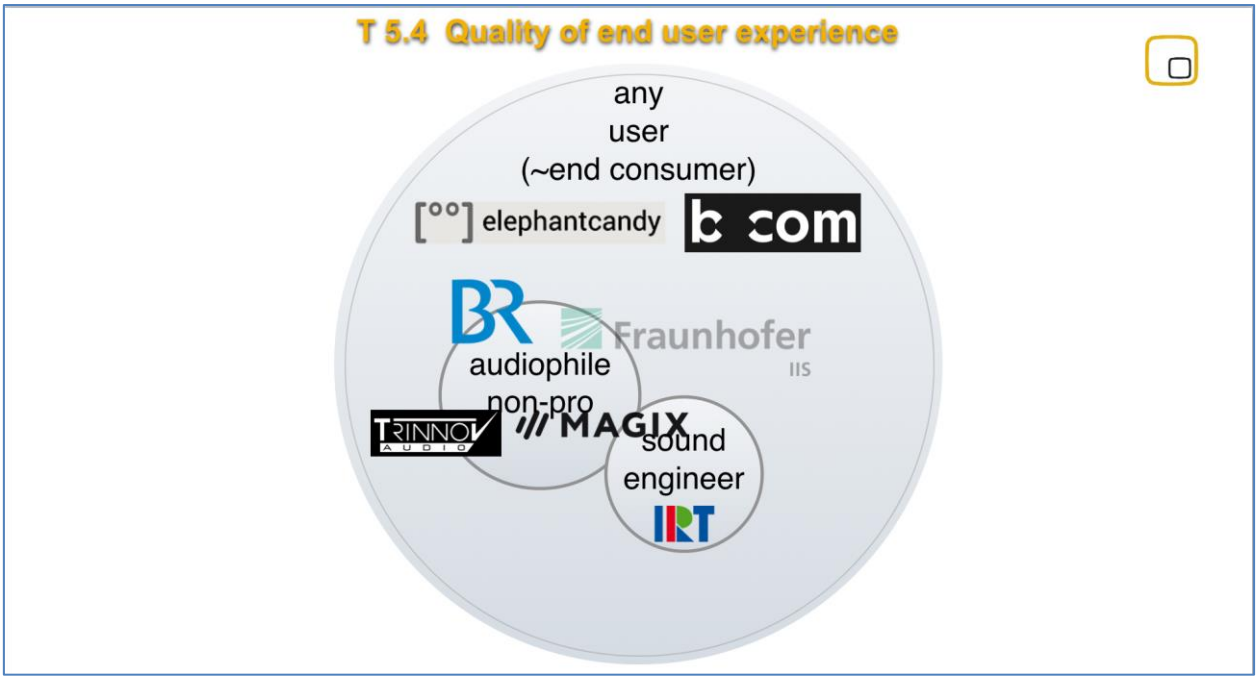


Figure 34: Classification of end users at partners

Set into a frame with identified categories of qualities and experiences, this picture indicates roughly their relationship of sets and subsets and their respective main interests or consideration.

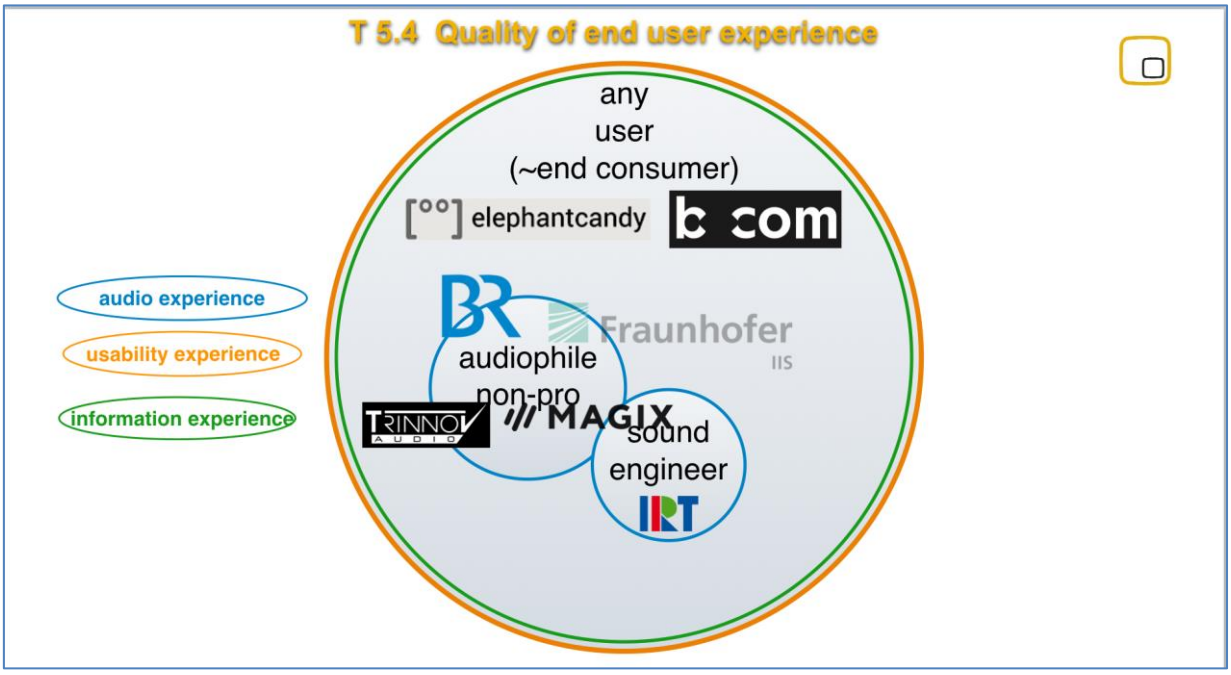


Figure 35: Target audience, interests and fields of focus of partners

© 2015 - 2018 ORPHEUS Consortium Parties

4.4 Focusing the Model within the ORPHEUS Pilot Architecture

Having identified the here main categories of examination

- audio experience
- usability experience
- information experience

as well as the three main profiles of 'end users'

- end consumer
- 'audiophile' listener
- sound engineer,
- media professional

we can now hotspot clearly the different points of applicability within ORPHEUS' pilot architecture

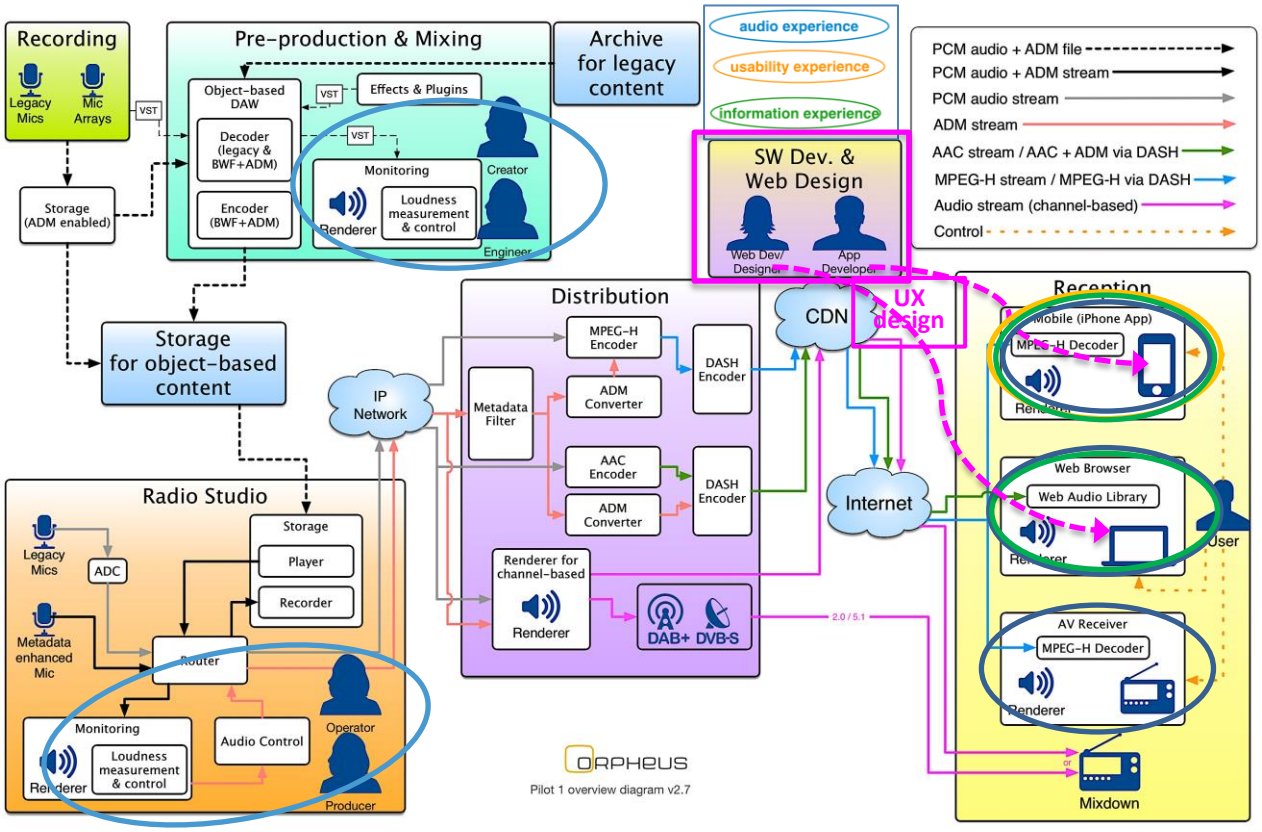


Figure 36: User experience issues in the ORPHEUS pilot architecture

The assessment of

- audio experience is to be focused on
  - the pre-production and radio studio with sound engineers and media-professionals
  - with audiophile and general end users on the ORPHEUS Receiver, Web Browser representation and ORPHEUS Mobile APP
- usability with any user is to be applied mainly on
  - the ORPHEUS Mobile App and
- information experience with any user is to be applied in
  - the ORPHEUS App
  - Web Browser representation.

The necessary tasks are distributed to the respective partners.

© 2015 - 2018 ORPHEUS Consortium Parties

5 Conclusions

The object-based media approach is an essential component for emerging cross-media demands, being integrative, scalable and genuine IP-based. Within the media-chain, starting with the capturing process, the editing and mixing, play-out, distribution until the reception process, a plethora of adjacent metadata offers ample possibilities to optimise the reproduction process according to the reception device, situation environment and personal preferences.

The model developed here, comprised of three main categories

- audio experience
- usability experience
- information experience

is a first attempt, to regard the assessment of 'quality of end user experience' in a holistic way. The partners in the ORPHEUS consortium can provide both, a wide range of expertise in various fields of quality assessment and the necessary facilities to conduct research in this new field.

During the progress of the pilots, the ORPHEUS partners will conduct, orientate and adapt their activities along this model. In a later stage we intend to examine interdependencies and possible relations of the various categories, characteristics and single quality features for further refinement and development of the model.

© 2015 - 2018 ORPHEUS Consortium Parties

Appendix A

A.1 Internal survey on ‘QE’ approaches and expertise by ORPHEUS partners

Werner Bleisteiner 11.07.2016

T5.4 QoEUE questionnaire

You may tick multiple boxes – unless otherwise stated text boxes are limited to 256 characters.

company:

name:

1. Within the ORPHEUS project, in which domain do you want to examine “quality of end user experience”?
   (rate “top”=1; “medium” = 2; “low”=3, no= 0)
   - audio quality
   - usability quality
   - information quality
2. What method(s) do you want to apply when doing research on “quality of end user experience”?
   - quantifiable data
   - questioning people

   please specify/remarks:
3. How do you want to examine characteristics of “quality of end user experience”?
   - online (i.e retrieved data)
   - off-line (i.e. in a lab)

   please specify/remarks:
4. Are there special resources you will use or can provide for the research?
   - lab
   - data analysis tools

   please specify/remarks:
5. Do you apply or focus upon any specific Media User Typologies?
   - no
   - yes

   please specify which model:
6. Who are your target groups of your recruited probands?
   - audio professional
   - audio enthusiast
   - any consumer
   - computer affinitive (nerd)
   - young audiences
   - other

   please specify/remarks:

T5.4 QoEUE questionnaire-v03 page 1 of 2

Werner Bleisteiner 11.07.2016

7. What is the quantity of probands you are likely to examine?
   - up to 10
   - 11-50
   - 51-100
   - more than 100 please estimate
8. what is/are the end user device(s) or situations you will use for the testing scenarios?
   - smart device (mobile, tablet)
   - pc
   - home entertainment system
   - dedicated lab set-up/prototype
   - other

   please specify/remarks:
9. Please state some of the models or methods you intend to use or describe your test scenario or any further comments, suggestions . . .
   (not restricted text box)

Thank you!
wb

T5.4 QoEUE questionnaire-v03 page 2 of 2

Figure 37: Overview internal questionnaire

09.01.2017

WP 5.4 Quality of End User Experience BASIC QUESTIONS

question	BBC	b<>com	BR	ECANDY	IRT	MAGIX	(FH IIS - no task member)	question
1. Within the ORPHEUS project, in which domain do you want to examine "quality of end user experience"? (rate "top"=1; "medium" = 2; "low"=3, no= 0)		audio quality 3 usability quality 1 information quality 2	audio quality 1 usability quality 1 information quality 1	audio quality 1 usability quality 3 information quality 1	audio quality 1 usability quality 0 information quality 0	audio quality 1 usability quality 2 information quality 2	audio quality 2 usability quality 3 information quality 2	1. Within the ORPHEUS project, in which domain do you want to examine "quality of end user experience"? (rate "top"=1; "medium" = 2; "low"=3, no= 0)
2. What method(s) do you want to apply when doing research on "quality of end user experience"?		quantifiable data questionning people	questionning people	quantifiable data	quantifiable data questionning people	questionning people	quantifiable data questionning people	2. What method(s) do you want to apply when doing research on "quality of end user experience"?
		We would like to apply both. During the tests, it's possible to collect simultaneously objective (logs, user actions, clicks…) and subjective data (questionnaires,interviews…) before use (e.g acceptability) and after use (e.g acceptance)	we would like to apply both, but have to develop methods in both ways. But we're depending on technical solutions within the system and supporting efforts of ORPHEUS partners		Questioning might be rather asking for attributes of the listened stimuli or so. Furthermore, potential methodologies for the subjective assesment of audio reproduction may be the "Multiple Stimulus Ideal Profile"		quantifiable simple data like like/dislike, more advanced questions by quesionaire	
3. How do you want to examine characteristics of "quality of end user experience"?		online (retrieved data) off-line (in a lab)	online (retrieved data)	online (retrieved data)	off-line (in a lab)	off-line (in a lab)		3. How do you want to examine characteristics of "quality of end user experience"?
		Usability tests of the interface could be made in our lab. Mock-up analysis could be made remotely.	we don't run a proper lab on our own. So we're likely to depend solely upon data to be analysed					
4. Are there special resources you will use or can provide for the research?		lab data analysis tools	none specific	data analysis tools	lab data analysis tools	lab		4. Are there special resources you will use or can provide for the research?
			BR has a media research department. Depending on different methods we agree to use, we want to make them participate and support the research with special focus on		A server-based listening test software to conduct listening tests of all kinds in browsers and (probably) all over the world; our 3D audio lab			
5. Do you apply or focus upon any specific Media User Typologies?		no	yes	no	no	no		5. Do you apply or focus upon any specific Media User Typologies?
			In media research: Medien-Nutzer-Typologie, Sinus-Milieus, etc.					
6. Who are your target groups of your recruited probands?		any consumer	any consumer youn audiences other	any consumer	audio professional	audio professional audio enthusiast	audio professional any consumer young audiences	6. Who are your target groups of your recruited probands?
			as for intelligibility of speech and contextual information : target group of 'best agers' and beyond (65+) and hearing impaired audiences are to addressed					
7. What is the quantity of probands you are likely to examine?		<10	>100	11-50	11-50	up to 10		7. What is the quantity of probands you are likely to examine?
			possible, depending on the spread of the pilots					
8. what is/are the end user device(s) or situations you will use for the testing scenarios?		smart devices pc	smart devices pc home entertainment system other	samrt device	smart device (mobile, tablet) pc dedicated lab set-up/prototype	smart device (mobile, tablet) dedicated lab set-up/prototype		8. what is/are the end user device(s) or situations you will use for the testing scenarios?
			as it's aimed to provide pilot phase 1 also in legacy broadcast, evaluation of this process is to be considered					
9. Please state some of the models or methods you intend to use or describe your test scenario or any further comments, suggestions . . .		To sum up, in the UA Lab, at b<>com, we propose 3 actions. The first would be a benchmark, to compare existing relevant products (if you have any ideas on existing relevant products, we are interested). The second would be a mock-up analysis. The last would be usability tests, with real users, in lab. This final step is yet to be defined. A priori, users should follow specific scenario tests. Then, we will observe usability problems and subjective perception (with logs, questionnaires, interviews, focus group, etc.)		we will monitor the use of the app by inventorizing clicks, time between clicks, attention spans etc	Potential tests/experiments might include: - the comparison of object-based / NGA codecs such as Dolby Atmos and MPEG-H with respect to the overall audio quality in dependency of the conding efficiency - certain attributes of the rendering of object-based codecs (spatiousness, envelopment, externalisation, …) - reference-less evaluation of object-based reproduction on different end devices - renderer behavior for certain critical audio scenes - downmix issues - mapping of objects to loudspeakers	Sequoia Beta Group Sequoia Forum internal tester		9. Please state some of the models or methods you intend to use or describe your test scenario or any further comments, suggestions . . .

Figure 38: Overview internal survey results

[end of document]