

Audio Conferencing Enhancements

Leena Vesterinen

University of Tampere
Department of Computer Sciences
Interactive Technology
Master's Thesis
Supervisor: Kari-Jouko Rähkä
June 2006

University of Tampere
Department of Computer Sciences
Interactive Technology
Leena Vesterinen: Audio Conferencing Enhancements
Master's Thesis, 54 pages, 4 pages of appendix
June 2006

Audio conferencing allows multiple people in distant locations to interact in a single voice call. Whilst it can be very useful service it also has several key disadvantages. This thesis study investigated the options for improving the user experience of the mobile teleconferencing applications. In particular, the use of 3D, spatial audio and visual-interactive functionality was investigated as the means of improving the intelligibility and audio perception during the audio conference calls. User experience studies: web based user survey, subjective tests, demonstrations and focus groups were conducted in order to collect feedback from the audio conference users and identify the issues with current audio conference systems.

Keywords: 3D, spatial audio, audio conferencing, interaction, user interfaces, visualisation, HRTF, Head Related Transfer Functions, user experience

Contents

1. Introduction.....	1
2. Audio Conferencing.....	4
3. Enhancements on Audio Conferencing.....	7
3.1 Audio Terminology.....	7
3.2 Issues with Audio Conferencing.....	9
3.3 Spatial 3D Audio.....	10
3.3.1 Head Related Transfer Functions (HRTF).....	12
3.3.2 Stereo Panning	13
3.4 Visual – Interactive Functionality.....	14
4. Subjective Audio Testing with the MUSHRA Method	18
4.1 Research question	19
4.2 Participants.....	20
4.3 Acoustic Requirements	20
4.4 Test Design and Implementation	21
4.5 Data analysis	21
4.6 Part 1: Preliminary Hearing Test	22
4.6.1 Test Task design	22
4.6.2 Test Procedure	23
4.6.3 Results.....	24
4.7 Part 2: Subjective Test on intelligibility	26
4.7.1 Test Task Design.....	26
4.7.2 Test Procedure	30
4.7.3 Results.....	31
4.8 Part 3: Subjective Test on Perception	32
4.8.1 Test Task Design.....	33
4.8.2 Test Procedure	34
4.8.3 Results.....	34
5. WEB Based User Survey	36
5.1 Survey Design.....	36
5.2 Respondents	37
5.3 Survey procedure	37
5.4 Results.....	38
5.4.1 User Profile and Usage	38
5.4.2 User Experience	39
5.4.3 Improvement suggestions	40
6. Product Demonstration and Focus Groups	42

6.1 ACE Demonstrator.....	42
6.2 Participants.....	43
6.3 Procedure	44
6.4 Feedback Results	44
7. Discussion	47
8. Summary	51
References.....	52
Appendix	

1. Introduction

Modern network technologies and groupware applications enable groups to “collocate virtually” when the group members are not physically in the same place. Some of the advantages of the distant collaboration tools, such as audio conferencing, are to save time and money spent on travelling. [Olson & Teasley, 1996]

Current audio conferencing systems are still successfully used for “virtual collocation” purposes even if video conferencing, Voice over Internet Protocol (VoIP) and chatting are making headway in the distant conferencing culture. Audio conferencing allows several participants to join a single voice call, a conference call, through a landline telephone or a mobile phone. Especially, mobile based conferencing is becoming increasingly popular among business users due to its flexibility, portability and freedom.

Whilst audio conferencing can be a very useful service, it also has several key disadvantages. Effective communication between people requires that the communicative exchange take place with respect to some level of common ground. In other words, common ground is based on the knowledge that participants share. Traditionally audio conference calls have suffered from the difficulty of establishing a common ground, because conference participants find it difficult to follow the conference calls or identifying the other participants on a call is often impossible due to issues of intelligibility or audio perception [Olson & Olson, 2000]. The issues with speech intelligibility, such as acoustical problems, can make the conference call hard to hear and follow due to ambient noise, echoes and amplitude differences in each ear [Brungart et al., 2002]. Additionally, inconsistent call quality between participants can cause distractions. Whereas, in a face to face meeting, a person can determine who is talking using visual cues, directional clues picked up by the

ears or combination of both. In an audio conference call, these cues are lost and hence it can be difficult to determine who is talking, or when a new speaker begins. Therefore, identifying and remembering who is participating in the conference and who or what (company, organisation, team) they are representing is often difficult [Billinghurst et al., 1998].

Based on the human natural ability to hear spatially and to process multiple audio streams of information simultaneously [Arons, 1992], previous research [Goose et al., 2005; Baldis, 2001; Marentakis & Brewster, 2005] has shown that spatial, 3D audio can enhance speech intelligibility and audio perception of the audio conference calls. Spatial, 3D audio could be used to imitate the human binaural hearing system by emitting sounds from a real stereophonic sound source (such as headphones / speakers). These processed sounds would appear to be coming from phantom positions around human head. In other words, sounds would be positioned around the user, creating a virtual, 3D sound stage. Therefore, spatial, 3D audio could help increasing voice separation between conference participants, reducing listener fatigue and creating a more natural environment for the users.

Another way of eliminating issues with audio conferencing could be to introduce visual-interactive functionality. By visualising the audio conference participants on a graphical user interface of a mobile device [Goose et al., 2005; Mortlock et al., 1997; Hindus et al., 1996] could further help in identifying the speaking participants. Further visualisations in combination with interactive functionality within the conferencing application could reduce the issues of intelligibility and perception.

The research reported in this paper was conducted as a part of the Audio Conferencing Enhancements (ACE) project at Vodafone research and development department in United Kingdom. The goal of the research was to investigate the user experience of the current audio conferencing systems and to find the ways to enhance them. In particular, the ACE study concentrated on solving the problems of intelligibility and perception in order to differentiate between participants during an audio conference call. Therefore, a spatial, 3D

audio and visual-interactive functionality were investigated as means of enhancing the audio conferencing systems.

Two techniques, Head Related Transfer Functions (HRTF) and stereo panning, for reproducing spatial audio were applied in the ACE study. These techniques will be further discussed later in this document.

Two major questions were presented for this research study:

1: Can the spatial, 3D audio improve the speech intelligibility and audio perception of the audio conference systems?

2: What are the user requirements for the visual-interactive functionality on a mobile based audio conferencing application?

Performance differences for 3D, monophonic and stereophonic audio conferences were tested through subjective testing sessions. Demonstrations and focus groups were conducted in order to gain understanding of the user requirements for the visual-interactive functionality of the mobile based audio conferencing. The visual-interactive functionality would of course be dependent on a device capable of displaying the required information enabling users to view the information during a conference call. In the situations where concentration and coordination of the hands and eye is important, this could prove to be dangerous.

2. Audio Conferencing

The time spent on business travels results in decreased productivity and a great amount of money is spent on travelling between remote sites. Therefore, many companies are increasingly evaluating and deploying technologies in order to save time and money while doing business. [Goose et al., 2005]

Currently, videoconferencing is making fast progress however the audio conferencing continues to remain in an important role in a distant collaboration culture. Virtually, every place in the world has at least analogue telephone service which enables audio conferencing to be universally available. Throughout the years, global businesses and organisations have benefited from this communication medium extensively, linking separately located colleagues, business partners and clients. A common example of a multi-party conferencing facility is a fixed phone line audio conference set up in a conference room. This conference call set up allows several participants to be present at the same time in the same conference room and conference calls are established through conference phone which consist of speakers and built-in microphones. Users may adjust the output volume of the call or one can mute itself, however traditional fixed line audio conference phones have had very limited functionality.

In addition to above, audio teleconferencing is based on the good conversational skills. In communication, we apply conversation as a medium for decision making and through conversation we generate, develop, validate and share knowledge. Conversation has been said to have two major characteristics:

- 1.) Talking is an intensely interactive intellectual process and is seen as an outstanding method for eliciting, unpacking, articulating, applying, and re-contextualising knowledge.
- 2.) Conversation is fundamental social process [Erickson & Kellogg, 2000].

Good conversational skills therefore are etiquettes which apply in audio conferencing. The most familiar conference call etiquette is turn taking, requiring that speakers pause in their speech in order to let others talk [Aoki et al., 2003]. If this etiquette is not followed in an audio conference, a simultaneous conversation may result in unpleasant communication experience. Role taking is also an important part of the audio conference culture. As an extension to above, in order to create more natural, flexible and open audio conferencing system, Greenhalgh and Benford [1995] proposed that effective social spatial skills should be considered in the people interaction. Therefore, Greenhalgh and Benford researched into creating awareness between the conference participants by introducing audio, visualisation and interaction in the audio conference functionality.

Recent studies [Baldis, 2001; Billinghamurst et al., 2002; Goose et al., 2005; Williams, 1997] have shown that people are performing worse in the audio / video conferencing conditions than in a face to face collaboration. However, face to face interaction is not confirmed to be any better than speech-only interaction for cognitive problem solving. Visual cues instead can be beneficial for tasks requiring negotiation. In the face to face meeting, a person can determine who is talking using either visual cues or cues picked up by their ears (or a combination of both). In an audio conference, these cues are lost and hence it can be difficult to determine who is talking, or when a new speaker begins. Therefore, identifying and remembering who is actually participating in the conference and which company, organization or department they represent is often difficult. Thunderwire study [Hindus et al., 1996] supports the findings of Baldis and Billinghamurst by investigating the effectiveness of audio in communication space. Study showed that the audio alone may be sufficient for decent interaction between people and that participants communicated naturally and socially in the audio communication space. However, some major problems were pointed out in the research, e.g. users were not able to tell who was present in a conferencing space. Also the lack of visual cues made audio only communication difficult.

Despite the increased demand of the audio conferencing, the audio conferencing user experience is still inadequate. Typically in the audio conference call, participant voices are sent through one audio output channel, resulting in confusing and unnatural conference experience. Research on memory, speech intelligibility and participant identification shows that spatial, 3D audio may improve conference performance [Baldis, 2001]. Several other researches [Yamazaki & Herder, 2000; Burgess, 1992] indicate that spatial audio can improve separation of multiple voices from each other.

Spatial audio can be reproduced through headphones to give the listener the feeling that the sounds are located in a space around the listener: front, rear, right, left or even above. In addition to spatial audio study by [Goose et al., 2005] shows that interactive graphical representation of an audio conference may aid with the issues of speech intelligibility and perception.

Later, movements in human-computer interaction and increased usage of small screen devices: mobile phones and personal digital assistants (PDA) resulted in escalating portability of computing and communication facilities. Current technology and continuous research within portable devices assure that mobile conferencing facilities continue to develop [Goose et al., 2005; Billinghamurst et al., 2002].

3. Enhancements on Audio Conferencing

This chapter will concentrate on the audio conferencing enhancements, introducing 3D, spatial audio and visual-interactive functionality as means of improving the audio conference systems.

3.1 Audio Terminology

Monaural, also known as *monophonic* (mono) audio is a reproduction of sound through a single audio channel. Typically there is only one loudspeaker, and if multiple speakers are used, the audio is perceived evenly from the left and right, causing an unnatural interaction environment. During a traditional desktop conference, the multiple sound sources originate from a set of *monophonic* speakers. In other words, mono sound is outputted through one audio channel arriving to listener's both ears at the same time. [Baldis, 2001]

Traditionally, *stereophonic* or *binaural* reproduction of sound uses two audio channels: left and right. These channels appear to distribute the sound sources recreating a more natural listening experience. For example, spatial, 3D¹, surround sound in cinemas is based on the multi channel stereophonic audio output technology. The human's natural way of hearing sounds, in our audio listening environment, is based on the *binaural* experience. The binaural hearing means the human ability to perceive locations of sounds based on the interaural² differences in time and intensity. The neurons located in various parts of the auditory pathway are very sensitive to disparity in the sound arrival time and sound intensity

¹ 3D audio has the three dimensions: length, width and depth.

between the two ears. The sound arrival time and the intensity in addition with some other interaural hearing factors create our spatial hearing experience in the real world. [Shaw, 1996].

When talking about binaural hearing we can also associate it with a term spatial hearing. Binaural, *spatial hearing* is thought to be one of the most complex biological abilities. During every day conversations, people have an ability to ‘tune out’ disturbing noises from their environment [Vause & Grantham, 1998]. An ability to selectively focus on one single talker among a cacophony³ of conversations is known as the “cocktail party effect”. Therefore, the listener is able to focus on one interesting audio signal at a time in an environment where many simultaneous audio streams are present. In addition, the listener has an option to switch attention between the audio signals in a listening environment. [Arons, 1992; Stifelman, 1994]

In the spatial, 3D listening environment the actual sounds are emitted from a real stereophonic sound source, such as headphones or speakers. The sounds emitted are perceived as coming from phantom positions around the human head. For example, 3D audio technology is used in film and game industry, as well as in the safety critical systems such as aircraft cockpits [Johnson & Dell, 2003]. In order to represent the audio in 3D, advanced audio processing technologies are required.

Typically, fixed line phone operates using frequency response ranges from 300 Hz to 3400 Hz. Voice transmission requires a bandwidth usage of about 3000 to 4000 cycles, 3 to 4 kHz per second. GSM communications network instead operates in the 900 and 1800 MHz frequency bands. Therefore, subjective tests carried out in this study using a GSM audio quality was remarkably better than the voice transmission quality using the monophonic audio.

² Interaural means the mechanism we have in our ears to perceive sounds. Human interaural properties are based on the neurons.

³ Harsh, mixed joining of sounds.

When talking about audio signals, the term frequency is very common. Frequency is measured in Hz and it means the number of cycles or complete vibrations experienced per second. The frequency of a sound wave determines its pitch. A hearing frequency range of a young person is about 20 to 20,000 hertz. But closer to middle age, hearing range decreases to 15 Hz to 18 kHz. Therefore, most of the audio frequencies used in the subjective audio testing samples were within the human hearing frequency range. Audio filtering was also applied using high- and low-pass filters. A high pass filter was used to pass frequencies above a certain value and to attenuate frequencies below that value. A low-pass filter therefore was to pass frequencies below a certain value and to attenuate frequencies above that value.

3.2 Issues with Audio Conferencing

According to findings of Brungart et al. [2002], the complexity of the multi channel listening problem is based on the various issues with audio intelligibility and perception. Most frequently, problems are caused by ambient noise, the high number of competing talkers, similarities in voice characteristics, similar voice frequency levels, location of the talker and a listener's prior knowledge and experience about the listening task.

Ambient noise is a common issue in audio communication, which can be caused by the listening environment or the noise in the communication network. When the number of competing talkers in the listening environment increases, identifying different participants on a call becomes more difficult. This is caused by the possibility of overlapping and interfering conversations. Voices of different talkers may vary in many ways, including: speech frequency, accent or voice intonation. Talkers representing different sex and age groups are easier to recognize from each other, but when the voices are similar and from same sex, identification can become complex. Therefore, an important improvement in the speech intelligibility of the multi channel listening systems could be achieved by increasing the volume levels among the talkers. Binaural audio could solve the issues of intelligibility and perception in noisy environments leading to easier separation of the participants from each other. [Brungart et al., 2002]

In addition to the findings of Brungart et al., other research studies [Mortlock et al., 1997; Billinghamurst et al., 2002; Baldis, 2001; Goose et al., 2005] show that current audio conferencing systems with one audio output channel limit the naturalness of the communication between participants. In order to increase the naturalness of the current audio conference communication, virtual conferencing is introduced. Through virtual, spatial audio, interaction between larger groups of people could become pleasant.

3.3 Spatial 3D Audio

What is spatial 3D audio?

Spatial audio is considered to be a good way of enhancing the audio conferencing facilities. The actual sounds appear to be coming from phantom positions in the audio space creating a 3D feel for the listeners on a conference call. [Evans et al., 1997]

Monaural presentation of sound is attained when outputting sound through single earphones such as mobile phone speakers or hands-free kit with a single earpiece. However, the human auditory system localizes the sounds based on the binaural cues, the time differences when audio signals arrive at right and left ears. Therefore by using a monaural sound output an impression of spatial audio is corrupted and sound localization is not accurate. Current audio technology, synthetic sound production, processes sounds in a way that when reaching human ears, listener feels as if the sound would be located externally around the listener, 'around the head'. [Marentakis & Brewster, 2005; Goose et al., 2005]

Binaural, spatial hearing is evidenced to provide important environmental cues to the human. Therefore spatial audio is used to improve speech intelligibility on audio conferences. Some research [Burgess, 1992; Baldis, 2001] shows that spatial audio may improve the intelligibility of the conference calls by:

- Allowing user to distinguish between other conference participants easier

- Providing more natural listening environment and reducing listening fatigue when listening through headphones by introducing ‘around the head’ feel to the audio
- Enhanced audio only information can be used in hands busy / eyes busy environment
- Potential solution for large number of conference participants due to extended sound field

An experiment [Goose et al., 2005] was carried out to gain a deeper understanding of human ability to hear mixture of audio cues through ears. The primary cues facilitating a human spatial audio perception were described as follows:

Volume - The longer the distance is between the listener and the object, quieter the sound is.

Interaural intensity difference (IID) - The sound reaching the listener from the right will sound louder in the right ear than in the left ear.

Interaural Time Difference (ITD) - The sound originating from a source to the listener's right side will reach the right ear approximately one millisecond before the left ear.

Reverberation - The reflections of sound within a closed space is known as reverberation. The sound effects are fully dependent on the shape and size of the room where the sounds are produced.

In order to output a spatial 3D audio, stereophonic headphones or dual stereo earpieces are required. This requirement for the audio conferencing enhancements has been criticized due to isolation of the users from their real world audio environment while on a conference call. [Marentakis & Brewster, 2005]

Spatial audio listening experience could be produced by various different techniques. However, in the Audio Conferencing Enhancements project we have looked into two different potential solutions:

1. Head Related Transfer Functions
2. Stereo panning technique

These audio reproduction techniques will be discussed in the next section.

3.3.1 Head Related Transfer Functions (HRTF)

Sounds generated in space reach listener's ears as sound waves. When we hear a sound from our left side, the sound reaches our left ear before the right ear. Our ear acts as a tone controller and is dependent on the incident sound. Unconsciously, the human uses the time difference, intensity difference and tonal information at each ear to locate the sounds in our environment. [Gardner, 1999]

The idea of the Head Related Transfer Function is to measure the transformation of sound from a point in space to the ear canal [Gardner, 1999]. HRTF are based on mathematical transformations on the spectrum of the sound that simulate the time difference, intensity difference and tonal information in each ear. They also involve outer ear: pinna geometry, inner ear: ear canal geometry and diffraction reflection in order to perceive spatial audio experience. To gain an 'around the head' feel more than a thousand different functions have to be generated. The HRTF are based on a coordinate system of a human head, defining the centre of a head as a point halfway between the ears. [Johnson & Dell, 2003; Kan et al., 2004]

Sound localization cues for each ear are reproduced after the sound signals are processed by the digital filter and listened to through headphones. At this stage, a listener should perceive a sound at a location specified by the HRTF. However, the localization performance becomes inaccurate when directional cues are synthesized by using the HRTF measures taken from different sized or shaped head [Gardner, 1999]. Therefore, this means that the HRTF must be individualized for each listener in order to gain full, accurate 3D listening experience. In practice, despite of differences between human head sizes and shapes, non-individualized HRTF are often used in applications.

Usage of non-individualized HRTF can cause sound localization errors, where listeners are unable to tell whether the sound is coming from front or behind them. This phenomenon is

known as the front/back confusion. In other words, this would mean that some listeners may not be able to perceive the rear sounds as coming from behind, especially if the sounds were presented in combination with front and side sounds. In practice, a sound would be panned from the front, around to the side and to the rear, but the listener would hear the sound coming from the front to the side and back to the front. [Burgess, 1992; Gardner, 1999]

The elevation errors are also another common issue with spatial audio processing. In practice, when a sound is moved directly to the right and directly upwards, this may create a feeling as if the sound would be moving from the right to the front. Especially, this is commonly experienced when using loudspeakers. However, high frequency cues are more effectively reproduced when using headphones [Gardner, 1999]. In the 2D plane instead, height change should make no difference and without head movement we can not determine elevation or whether the sound source is in front or behind of us. Advantages of the HRTF are that they create a more natural listening environment by reducing listening fatigue, especially when listening through headphones. Once the sounds are spatially separated, a listener can easily follow where the sound sources are located.

Study of Marentakis and Brewster [2005] states that sound localization will not be perfectly accurate neither in real or virtual audio environments. Localization errors range from ± 3.6 degrees in the frontal direction, when listening to sound sources presented by loudspeakers in well controlled, natural environments. Yet again, the localization errors range as much as ± 10 degrees in left/right directions and ± 5.5 degrees to the back of a listener. In addition, sound localization error rates may be decreased by using headphones.

3.3.2 Stereo Panning

A stereo panning technique is used to obtain 2D spatial sound without a need for virtual sound processing. In the stereo panning technique, a set of loudspeakers are used to create a sound field across a listening space [Evans et al., 1997]. The idea of stereo panning, used in audio conferencing, is to steer the voice of each participant to a narrow space immediately in front or to the sides of the user in order to help to distinguish who is speaking. As a

technique to improve the intelligibility and perception of the audio conferencing, the stereo panning technique might be less complex to implement than a full 3D audio processing with HRTF. However, stereo panning would only offer support to a maximum of 3 to 5 participants, as the audio output positioning is restricted to a smaller area. Stereo panning allows positioning of the sounds to far left and right, middle left and right and centre, front of the listener. Positioning of the sounds to the sides, rear, above or below the listener are not supported [Gardner, 1999]. Therefore, the stereo panning technique creates a very unnatural listening environment, as the audio is directly heard in the left or right ear and binaural listening experience would be inaccurate.

3.4 Visual – Interactive Functionality

What do we mean with the *visual – interactive* functionality?

A method of presenting information in graphical or interactive form is known as information visualization. The aim is to apply visual or interactive cues, or combination of both, in order to make information easier to handle by the users. Together, visual and auditory cues emphasize the information content by providing a bigger sense of realism. Especially, visual-interactive functionality can become useful in group applications such as audio conferencing. [Wenzel et al., 1990]

The recent study by Kilgore, [2003] suggests that trust is reduced if visual cues are removed. This would mean that we would reduce our head swivelling for forward and back elevation cues. On the other hand, removing visual cues might be a benefit when limiting a disruption caused by unrelated personal activities. The Thunderwire study [Hindus et al., 1996] found that the social preference of the audio only media space was to know who is present in the audio space and suggesting that the user interface functionality should be improved by making the communication space more intelligible. In addition, Goose et al., [2005] and Yamazaki & Herder [2000] suggest that egocentric view of a virtual meeting room, illustrating each conference participant by colourful spherical avatars could improve the intelligibility and overall perception of the conference call.

In the Conferencing3 model, the conference participants were located around a table in the virtual space as if they were in the real world. Participants were able to control their 3D audio space by dragging the spherical avatars around the screen, relocating them and causing the audio to emanate from the corresponding positions in the audio environment. As the participants joined or left the session, an avatar was reflected or removed. The user had an option to fetch some basic metadata about the fellow participants by clicking the corresponding avatar, e.g. contact number, email address etc. [Goose et al., 2005]

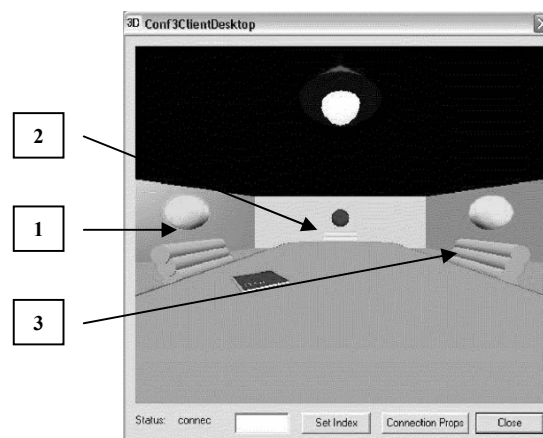


Figure 3-1: Desktop client of the 3D audio conferencing space with three participants and the user [Goose et al., 2005].

Both Goose et al.[2005] and Kilgore et al.[2003] show that moving participants manually around the UI could increase the chance of remembering who said what in an audio conference call. The spatial audio itself was not seen to improve the participant's ability to remember who said what. However, if the conference participants were given an opportunity to move similar sounding voices into separate locations in the audio environment, improvement on the memory was increased [Kilgore et al., 2003]. Contradicting Kilgore's findings, Baldis showed in her study that spatial audio enhanced memory, focal assurance and perception of the sound during the conference call [2001]. We also have to remember that individuals skilled at monaural listening, e.g. partly deaf people, may have differencing preferences. Both Kilgore and Baldis reported that the stereophonic audio format was preferred to non-spatial, monophonic audio format. In addition to above,

spatial audio reduced the difficulty of perceiving and identifying the speakers during conferences. [Kilgore et al., 2003; Baldis, 2001]

British Telecom's study [Mortlock et al., 1997] had however presented more extensive virtual conferencing room visualization. In this egocentric visualization model, participant's movement, expressions and actions are demonstrated by virtual characters, avatars. The visual presentation accuracy and user details were dependent on the number of the conference participants. Less the conference participants present, more detailed the avatars would be (Figure 3-1). A model by Yamazaki & Herder [2000], demonstrated a virtual audio conferencing environment where the user could move and turn around to change his/her position and other participants would hear the changes in sounds.

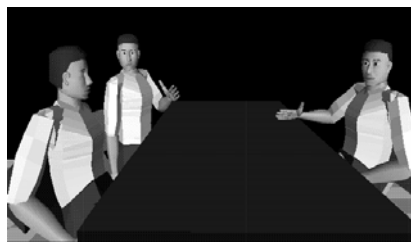


Figure 3-2: Virtual conferencing space by Mortlock et al. [1997].

Hindus et al. [1996] suggested that muting a microphone during an incoming call would reduce the ambient noise and would therefore improve the call performance. The study also found that additional functionality to allow two-way private conversations could be useful. Furthermore, a desktop audio teleconferencing study [Billingshurst et al., 2002] supports the findings of Hindus, adding that the document sharing and chat facilities could aid in group communication.

The ACE project itself researched into audio teleconferencing enhancements on mobile devices. Two major ways of enhancing the mobile audio conferencing intelligibility and perception were investigated: spatial, 3D audio and visual-interactive functionality. The objective of the visual-interactive user interface functionality was to improve the interaction between the user and the application, as well as the interaction between the conference participants. The conference call participants would be visualised by name icons on a user

interface. Once the participant would talk, the name icon would flash or change its colour indicating the active participant. The visual representation could improve the intelligibility and perception, especially in the situations with high number of the participants or when the background noise would become extensive. In order to support the visual-interactive functionality, physical characteristics of the mobile phone user interfaces needed to be further investigated and developed.

The ACE demonstration allowed the conference participants to adjust the volume, to change the participant positioning on a user interface and to mute the microphones during the conference calls.

4. Subjective Audio Testing with the MUSHRA Method

The subjective audio tests were based on the *Multiple Stimulus Hidden Reference and Anchor* (MUSHRA) method. This method was designed for Subjective listening tests of intermediate audio quality and is a recommendation BS 1534, by the International Telecommunication Union – Radio communication (ITU-R). The MUSHRA method was developed by the European Broadcasting Union (EBU) project group in collaboration with the ITU-R in 1999. [TGS-SA WG4, 2004]

The ‘MUS’ part of the word (MUS-HRA) represents a multiple stimuli which means that all the different processed versions of the audio samples were presented to the subject. In practise all the test samples were available to the subject all the time throughout the test, allowing the subject compare them to each other. Therefore, by comparing the audio samples by quickly switching between them allowed subjects to make easy observations and decisions about the relative quality of the different versions of the audio.

The ‘HRA’ (MUS-HRA) instead, represent a hidden reference and anchor which stands for the percentage of the defined multiple stimuli. A reference tone, an unprocessed version of the audio sample, was available to the subjects during the test and was then used to compare and grade the other audio samples against. The reference tone was also included among the test tones and was therefore called as the “hidden reference”. The hidden reference was to achieve the highest score in the test.

Among the test tones, four bandwidth limited samples were included as anchors: 32kHz, 24kHz, 16kHz and 8kHz as suggestions. To decide which listeners achieved an acceptable

level of performance, the ability to identify the difference between the hidden reference and the bandwidth limited anchors were used. Any listeners who could not tell the difference between these two stimuli on most of the test items were rejected and marked as unreliable.

The grading scale for assessing the audio sample quality was introduced by using the following intervals:

- Excellent (5)
- Very Good (4)
- Good (3)
- Satisfactory (2)
- Unacceptable (1)

The MUSHRA method suggests using experienced listeners in the audio testing process, but inexperienced listeners could also be used after undertaking an audio listening training. [Zieliski et al., 2002; EBU-UER, 2003]

The subjective test procedure consisted of three parts: a preliminary hearing test, a test on intelligibility and a test on audio perception. The Preliminary hearing test was carried out by playing audio samples and implementing simple questions about the audio in order to measure the listener's hearing deficiencies and to segment the listeners into reliable, semi-reliable and un-reliable categories. These categories would then be used for filtering the final test data. The tests on intelligibility were to ascertain if the spatial audio increased user's ability to determine the number and origin of speech inputs (participants) in a compressed speech environment. The perception tests instead were to indicate if the listeners preferred spatial audio content to non-spatial mono or stereo audio content in a compressed GSM multi-person conversation.

4.1 Research question

Series of subjective tests using the MUSHRA method were undertaken. The primary objective was the attempt to assess the effect of spatial multiple audio on the intelligibility

and general perception of conversations held across a mobile network. Prior to the tests, some core questions were presented for the research:

- Can the spatial audio increase user's ability to determine the number of speech inputs (participants) in a compressed speech environment?
- Can the spatial audio increase user's ability to determine the origin of speech inputs (i.e. which participant is speaking) in a compressed speech environment?
- And do users subjectively prefer spatial audio to monophonic audio output in the mobile teleconference environment?

A listener's ability to detect, to identify and to monitor multi-person speech signals was measured across monophonic, stereophonic, and spatial audio environments. Factorial combinations of four variables, including *audio condition*, *spatial condition*, the *number of speech signals*, and the *gender of the speaker* were used in test samples throughout the testing procedure.

4.2 Participants

Due to lack of trained listeners, all the test subjects undertook a simple listening practise and a hearing test in order to normalise the spatial test results. Subjective tests were performed 'in-house' using subjects from Vodafone Group R&D-UK. Test subjects consisted of 10 people (7 men, 3 women), ages ranging from 20 to 37 with a mean of ~28 years. Total of 10 people completed the preliminary hearing test resulting that 6 of the subjects qualified without major hearing deficiencies providing reliable test data. 3 subjects provided semi-reliable test data and one data set was unreliable.

4.3 Acoustic Requirements

The acoustic properties of the listening room were extremely important for the test procedure. By defining acoustic properties, we ensured that the test results were

reproducible. If less stringent requirements were allowed then the results obtained from a noisier environment would not have matched those from a quieter one. Different characteristics of inferior loudspeakers, or headphones would have also led to anomalous results, and to a test that would have been hard to replicate. Therefore, all the testing took place in the offices of Vodafone R&D-UK between the hours of 9:00am and 5:00pm, during the week days. The ambient noise environment was consistent, approximately up to 40db for the all periods of testing. Participants were required to use headphones throughout the testing.

4.4 Test Design and Implementation

The subjective tests were completed on a WEB based forms using a standard laptop computer and the subjects were required to wear a set of standardized headphones, Sennheiser HD 477 Open, dynamic hi-fi stereo. Instructions given to the test subjects could significantly affect the way the subject performed the test, therefore written clear instructions were provided with each test case along with a grading table (where applicable) to limit confusion and unnecessary ambiguities. The grading table was a part of MUSHRA audio evaluation method and was used in the audio perception testing phase. The audio items chosen for the testing were representative of the material that could come up in the audio conference environment.

4.5 Data analysis

In the data analysis process of all the parts of the subjective tests, the mean grades and 95% confidence intervals (2 standard deviations) were calculated, thus providing error bars indicating the probability that the true value of a result lies within a given confidence interval.

4.6 Part 1: Preliminary Hearing Test

The aim of the preliminary hearing test was to allow subjective test results to be filtered regarding to listener's hearing deficiencies.

4.6.1 Test Task design

The preliminary hearing test consisted of 11 test tones and each of them represented a processed 'beep' sound in various frequency levels. Known frequencies covering the dynamic range of a human auditory system were used varying from ~45Hz to ~16 kHz.

The test tones used for the testing:

Clip1: A tone was outputted to right ear only at ~0.1 kHz frequency level. The audio volume and the pitch were consistent throughout the test.

Clip2: A tone was outputted to both ears at ~45 Hz frequency level. The audio volume and the pitch were consistent throughout the test.

Clip3: A tone was outputted to left ear only at ~8 kHz frequency level. The audio volume and the pitch were consistent throughout the test.

Clip4: A tone was outputted to both ears at ~1.5 kHz frequency level. The audio volume and the pitch were consistent throughout the test.

Clip5: A tone was outputted to both ears at ~1 kHz frequency level. The audio pitch was consistent but the volume was slightly changed throughout the test.

Clip6: A tone was outputted to left ear only at ~0.1 kHz frequency level. The audio volume and the pitch were consistent throughout the test.

Clip7: A tone was outputted to both ears at ~4 kHz frequency level. The audio volume and the pitch were consistent throughout the test.

Clip8: A tone was outputted to right ear only ~1 kHz frequency level. The audio volume and the pitch were consistent throughout the test.

Clip9: A tone was played to left ear only at ~16 kHz frequency level. The audio volume and the pitch were consistent throughout the test.

Clip10: A tone was played to right ear only at ~8 kHz frequency level. The audio volume and the pitch were consistent throughout the test.

Clip11: A tone was played to both ears at ~0.5 kHz frequency level. The audio volume and the pitch were consistent throughout the test.

4.6.2 Test Procedure

At the beginning of the preliminary test, the subjects were asked to play a 1000 Hz audio reference tone and to reduce the volume of their audio output to a point where the tone was barely audible.⁴ Once the audio output volume was configured, the volume level was left untouched throughout the testing.




			
	1000Hz Reference Tone	clip 1	clip 2
Q1. Is the Tone Audible?	<input type="text" value="-"/>	<input type="text" value="-"/>	<input type="text" value="-"/>
Q2. Which ear(s) does the tone sound?	<input type="text" value="-"/>	<input type="text" value="-"/>	<input type="text" value="-"/>
Q3. Is the tone's volume consistent?	<input type="text" value="-"/>	<input type="text" value="-"/>	<input type="text" value="-"/>
Q4. Is the tone's pitch (frequency) consistent?	<input type="text" value="-"/>	<input type="text" value="-"/>	<input type="text" value="-"/>

Figure 4-1: Preliminary hearing test tasks.

A series of 11 test tones, monophonic (right or left ear) and stereophonic (both ears) were played in turns to the test subjects. Each test tone lasted approximately 5 seconds and the total length of the preliminary hearing test was approximately 10 to 15 minutes, depending on the subject's performance. The subjects were able to replay the audio clip if needed. During the audio listening exercise, subjects were required to answer test tone related questions:

1. **Is the tone audible?** This allowed any frequency related hearing deficits to be flagged up for each test subject.
2. **In which ear does the tone sound?** This allowed any single ear (left or right) hearing deficiencies to be analysed.

3. **Is the tone's volume consistent throughout the duration of the tone?** This allowed any binaural volume balance problems and time dependant volume detection problems to be outlined.

4. **Is the tone's pitch (frequency) consistent throughout the duration of the tone?** This allowed any binaural tonal balance problems and time dependant pitch detection problems to be outlined.

The data was collected through local web server and saved into log files for later analysis.

4.6.3 Results

The test tones with very low and high frequencies were inaudible by all of the test subjects. Therefore, the test tones with frequencies of ~45Hz and ~16 kHz were excluded from the data analysis. The subject's hearing accuracy was then measured through hearing errors made in different audio frequencies, changes of pitch or the changes in volume for right, left or both ears. The results were then categorised into 6 groups which were to measure the hearing deficiencies within common human auditory system.

- 1: Deficient at frequency smaller than 0.1kHz.
- 2: Deficient at frequency smaller than 1kHz.
- 3: Deficient at frequency about 8kHz.
- 4: Right ear is deficient.
- 5: Left ear is deficient.
- 6: Binaural hearing inaccurate.

Three levels of reliability were used in the data analysis. The reliability was measured by the defects found from a subject's listening performance. *Reliable* subjects did not have any inaccuracies when hearing different levels of frequencies, especially the reference and

⁴ Users were requested to reduce the volume until the tone could no longer be heard and then increase the volume until a point was reached where the tone was audible again.

anchor tones. The subjects might have had some problems when hearing the changes in pitch or volume of the tone or when identifying if the sound was binaural (both ears) or outputted only into right or left ear.

Semi reliable test subjects had difficulties to hear one frequency level, but did not have problems with hearing the hidden reference or anchor tones. The semi-reliable subjects had also problems with identifying if the audio clip was binaural or monaural.

Unreliable subject did not manage to hear neither of the hidden reference or the anchor tones during the test. The test subject had also difficulties to hear other frequency levels correctly and had problems to identify binaural and monaural audio samples.

The preliminary hearing test results showed that out of 10 test subjects, 6 subjects qualified as reliable, 3 subjects as semi-reliable and 1 unreliable. Figure 4-2 illustrates the hearing deficiencies by each participant.

Person	Deficient at freq < 0.1 kHz	Deficient at freq < 1 kHz	Deficient at freq ~8 kHz	Right ear is deficient	Left ear is deficient	Binaural hearing inaccurate	Conclusion
A				✓		✓	Reliable
B							Reliable
C						✓	Reliable
D	✓			✓	✓	✓	Semi-reliable
E							Reliable
F	✓	✓	✓	✓		✓	Unreliable
G				✓	✓	✓	Semi-reliable
H	✓			✓	✓	✓	Semi-reliable
I						✓	Reliable
J				✓		✓	Reliable

Figure 4-2: Test subjects categorised by their data reliability.

After completing the preliminary hearing tests, the subjects were required to continue to a test on audio intelligibility. The results from the audio intelligibility tests were then divided into reliable, semi-reliable and unreliable test data.

4.7 Part 2: Subjective Test on intelligibility

The subjective tests on the audio intelligibility were to investigate if the spatial audio would increase the user's ability to determine the number and the origin of the speech inputs (participants) in a compressed speech environment.

4.7.1 Test Task Design

7 audio clips were used in the subjective intelligibility tests. The audio clips were recorded using 3 male and 3 female voices (speakers), age of 20 to 37 years. Each speaker was recorded counting numbers from 1 to 9 respectively. These recorded scripts of 6 speakers were then mixed into 7 different audio clips. The numbers counted were not representation of the identification 'tags' for the participants, but they were imitating a simple speaker outputs. After mixing the audio clips they were modified to simulate a call quality of a GSM connection, performing a high-pass filter at 100 Hz, and a low-pass filter at 4 kHz. In practise, this would mean that the assigned audio filtering would allow passing audio frequencies between 100 Hz and 4 kHz during the recording procedure. Each test clip lasted for 30 seconds at most and they were played once in a random order to the subjects.

The front and the rear hemisphere environments were constructed by positioning the sound signals equidistantly around a 180 degree of arc. However, for the mixed hemisphere environment the speakers were positioned in 6 of the 8 available positions, such that the average difference in *source-midline distance* (SMD) was a maximum for the configuration. The SMD algorithm was used to avoid the front / rear confusion which would create an impression that the rear hemisphere sounds are originating from the front and vice versa. In practice this would mean that the human ears find it difficult to detect the sounds which are positioned directly opposite to each other at the front and rear hemisphere (e.g. 45 degree front right and 45 degree rear left). Therefore, the sound would be perceived as it would be panned to the side and then back to the front instead of panning the sound from the front, around to the side, and back to the rear. Following the SMD scheme used by Nelson et al.

[1998 and 1999], the sound sources were positioned slightly off the direct alignment with the opposing sound sources, in the way that the angular separation between them was maximised. [Nelson et al., 1999; Nelson et al., 1998]

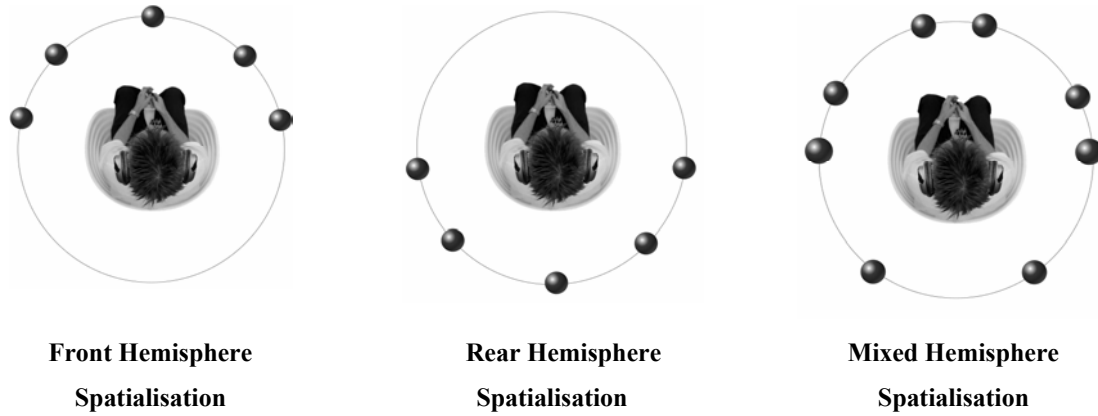


Figure 4-3: Spatial positioning of the sounds around the listener for the spatial audio clips.

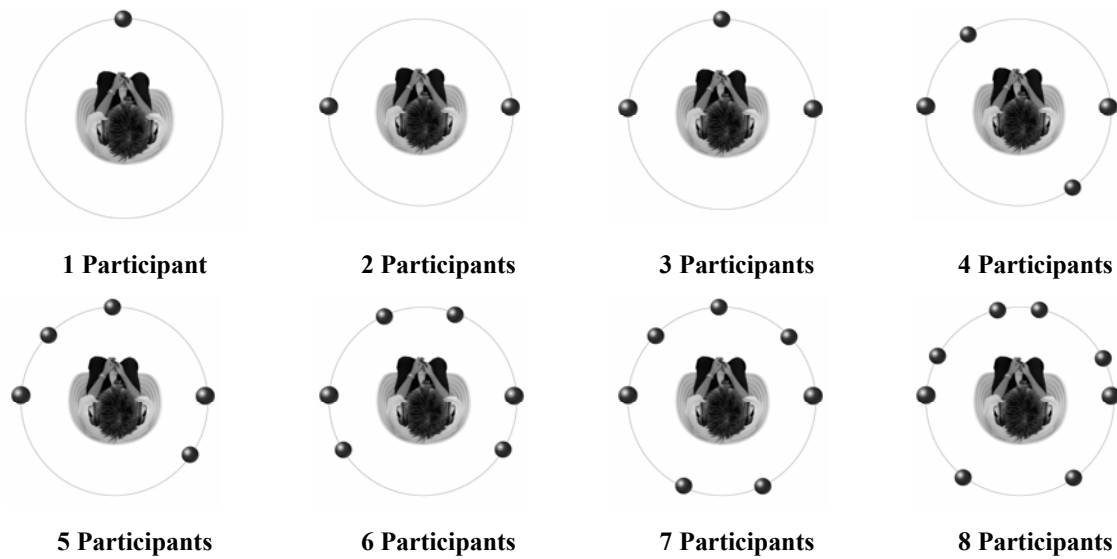


Figure 4-4: Recommended mixed hemisphere SMD positions for 8 sound sources [Nelson et al., 1998].

Each of the test clips contained 6 speakers counting numbers from 1 to 9 in turns. See example of the Clip1 structure in the following.

Speaker, female1: “one”

Speaker, male1: “two”

Speaker, male2: “three”

Speaker, female2: “four”

Speaker, female3: “five”

Speaker, male3: “six”

Speaker, female1: “seven”

Speaker, female3: “eight”

Speaker, male1: “nine”

Clip1: Monophonic audio clip was played through one audio channel and the ‘speaker’ voices were perceived coming from one sound source, from left and right ear simultaneously. Therefore no particular position could be identified. Total of 6 speakers were recorded in the monophonic audio clip.

Clip 2: Mixed hemisphere audio clip was played using ‘speaker’ voices which were virtually positioned in both, front and rear (mixed) hemisphere audio environment. The ‘speakers’ were positioned in 2D horizontal plane, 360° around the listener. Total of 8 potential positions were available for the speaker placement, however 6 speaker positions were selected randomly and occupied in this audio clip recording. The speaker saying number “four” was located in the position indicated with a tag E in the diagram (figure 4-5).

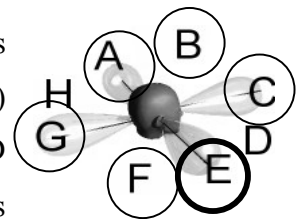


Figure 4-5:
Positions of the
speakers.

Clip 3: Flat stereophonic audio clip however worked on the premise that any sound source located to the left of the user should be 100% heard in the left ear. The differing ‘distances’ from the listener to the participant were then simulated by increasing or decreasing the amplitude of the signal in that ear only. This method allowed participant positioning only to the left or right, closer or further away from the listener. Total of 6 speakers voices were present in the recording. The speaker saying

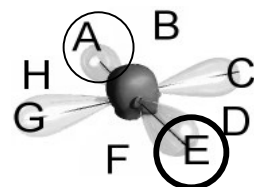


Figure 4-6: The
speaker positions in
the audio space.

number “four” was located in the position indicated with a tag E (figure 4-6).

Clip 4: The speakers were virtually positioned in the front hemisphere (180° view) of the 2D horizontal plane in the spatial environment. Total of 6 speakers were present and 5 different positions were used in the front hemisphere audio clip recording. Therefore, two of the speakers were given same spatial position in the audio clip recording. The speaker saying “four” was located in the position indicated with a tag D (figure 4-7).

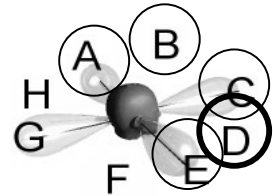


Figure 4-7: The speaker positions in the audio space.

Clip 5: The panned stereophonic audio clip divided the amplitude of the audio output between left and right channels. The ‘spatial like’ audio was reproduced by positioning the sound source middle left from the user, by recording the left signal with 75% of the total amplitude level and right signal with 25% of the total amplitude level (various amplitude levels were used in the design). This created a feeling as if the sound was coming e.g. from middle right of the listener in an audio environment. The stereo panning method enabled positioning of the sounds in middle right, far right, middle left, far left and front. The speaker saying “four” was located in the position indicated with a tag D (figure 4-8).

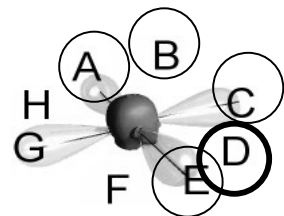


Figure 4-8: The speaker positions in the audio space.

Clip 6: The speakers were virtually positioned in the rear (180° view) hemisphere of a 2D horizontal plane in a spatial environment. 6 speakers were present in the rear hemisphere audio clip recording and 5 unique positions were occupied. The participant saying “four” was located in the position indicated with a tag F (figure 4-9).

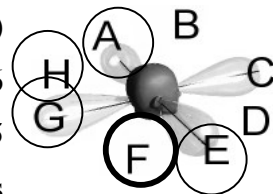


Figure 4-9: The speaker positions in the audio space.

Clip 7: : Mixed hemisphere audio clip was played using speakers voices, which were virtually positioned in both, front and rear (mixed) hemisphere audio environment. The speakers were positioned in 2D horizontal plane, 360° around the listener. Total of 8 potential positions were available for the speaker placement, but 6 speaker positions were selected randomly and occupied for the audio clip recording. The speaker saying number “four” was located in the position indicated with a tag E (figure 4-10).

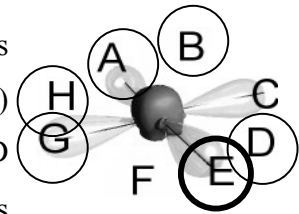


Figure 4-10: The speaker positions in the audio space.

4.7.2 Test Procedure

The test subjects were asked to listen to a randomised selection of 7 audio clips. After listening to each audio clip, subjects were to answer the audio intelligibility related questions. One of the questions required using a separate paper based diagram (see diagrams in figures 4–5 to 4-10).

1. **How many people did you think took part in the counting?** This addressed the issue of intelligibility and whether spatial audio samples could increase the chance of the test subject correctly ascertaining the number of participants within the clip.
2. **Did any one person speak more than once?** Also this question was modelled to learn more about the intelligibility and to ascertain if the spatial audio could help the test subjects to differentiate easier between the speakers.
3. **Indicate where you think the person saying “4” was sitting, using the diagram provided.** Third question was aiming to provide a parity check as an indication of the user’s ability to interpret compressed spatial audio signals.

4.7.3 Results

The results of the audio intelligibility tests were weighted according to the data sets created from the preliminary hearing test into reliable, semi-reliable and unreliable data.

The results show clearly that spatial audio can help to increase the intelligibility of a multi-person conversation in a compressed audio environment compared to that of a standard monophonic output. Data collected from the reliable and semi-reliable subjects reveal that spatial audio is better at allowing listeners accurately deduce the number of participants in a conversation compared to a monophonic output. This might have been explained with the spatial voices being easier to recognise and remember than the non-spatial ones (especially with similar sounding voices).

Usually, front hemisphere placement is more accurate than rear (mainly due to the human construct of turning ones head to face a sound allowing a position to be pinpointed), however these results indicated that the intelligibility of front hemisphere placement was similar to that of rear hemisphere. This might be due to front / rear location confusion that is common with spatial audio.

The mixed hemisphere placement solution appeared to provide the result that users found most intelligible. This is most likely to be due to the increased 'distance' between voices and the use of SMD placement eliminating any front / rear confusion. SMD placement, in mixed hemisphere allowed user to position, and consequently remember voices easier. [Nelson et al., 1999]

The results from intelligibility tests also suggest that stereophonic samples can help to increase the intelligibility of the monophonic audio. Flat stereo samples provide easier interpretation and the panned stereo samples provide similar results to that of the spatial audio samples.

The error bars shown indicate 2 standard deviations of the data set, representing a 95% data confidence level. Due to the small number of participants within this study, the error bars are very large and the differences between the spatial and stereophonic samples are rather small. However, if taking into account these large error bars, it is still apparent that spatial and panned stereophonic audio samples have proven to be more intelligible than standard monophonic samples in a compressed audio environment.

It should also be noted that the monophonic samples used were played back through stereo headphones, producing an identical left and right channel in accordance with the performance of smart phones with a stereophonic output. However, after discussing the results in the user interface team, I believe that if these samples were listened to through single monophonic earphone the intelligibility may have been further reduced.

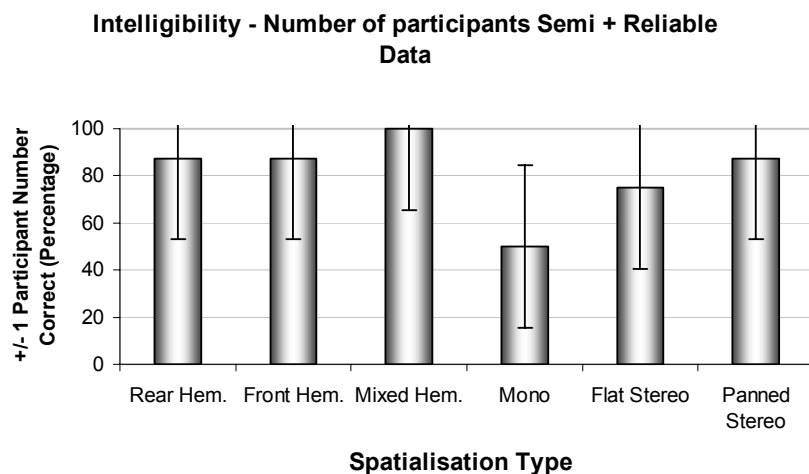


Figure 4-11: The tests show that mixed hemisphere spatial positioning was the most intelligible.

4.8 Part 3: Subjective Test on Perception

The goal of the audio perception test was to ascertain user's preferred sound stage in a

compressed multi-speech environment. By listening monophonic and stereophonic audio samples, the listeners were to evaluate the perception quality of the different audio clips.

4.8.1 Test Task Design

8 individual male and female participant voices were recorded and modified to simulate a call quality of a GSM connection of 4 kHz. The audio scripts were recordings of the ‘conference call like’ conversations. Two different scripts, producing multi-person conversational sound clips were recorded.

The testing was carried out in two test slots, each slot containing 7 audio clips. The duration of each individual test clip varied from 10 to 30 seconds. Monophonic, panned stereophonic, flat stereophonic and spatial audio clips were created giving the individual voices a virtual spatial position. Each test slot was designed and implemented using a different test scripts and changing speaker voices. The purpose of two different test slots was to gather larger amount of data for the analysis process. The nature of the clips was hidden and they were presented in a random order to the test subjects (e.g. if the clip 2 was played first in the Slot A, the clip might have been played fifth in the Slot B).

Clip 1: Monophonic presentation of the multi-person conversation, with 5 speaking participants in an imitated audio teleconference environment.

Clip 2: Flat Stereophonic presentation of the multi-person conversation, with 6 speaking participants in an imitated audio teleconference environment.

Clip 3: Panned stereophonic presentation of the multi-person conversation with 6 speaking participants in an imitated audio teleconference environment.

Clip 4: Spatial, front hemisphere audio presentation, with 5 speaking participants in an imitated audio teleconference environment.

Clip 5: Spatial, rear hemisphere audio presentation, with 5 speaking participants in an imitated audio teleconference environment.

Clip 6: Spatial, mixed hemisphere audio presentation, with 8 speaking participants in an imitated audio teleconference environment.

Clip 7: Spatial, mixed hemisphere audio presentation, with 6 speaking participants in an imitated audio teleconference environment.

4.8.2 Test Procedure

Test subjects listened to a randomised selection of audio clips through headphones. After listening to each audio clip, subjects were asked to evaluate each clip using the MUSHRA mean opinion score (MOS) subjective marking scheme as a guideline [Zieliski et al. 2002]. This marking scheme was used to indicate a subjective assessment of the users 'listening experience' of the audio clips. The subjects were then asked not to mark the clip on encoding quality, but instead on their perception of the clip as a whole. In other words, users listened to several different audio clips and were then asked to mark their listening experience using the marking scheme provided. Audio listening experience was divided into 5 levels:

- Excellent, 5 points
- Very good, 4 points
- Good, 3 points
- Satisfactory, 2 points
- Unacceptable, 1 point

Users were also asked to explain their answers to allow a better understanding of their subjective assessment.

4.8.3 Results

In the data analysis process, reliable and semi-reliable data was combined to form one valid data set. This was due to small number of participants in the test as well as project being commercial by its nature.

The perception test results suggest that spatial, mixed hemisphere audio output provided the most pleasing listening experience of a multi-person conversation in a compressed audio environment. As also seen with the previous intelligibility tests, mixed hemisphere, virtualised samples appeared to have provided the best listening experience. The reason for mixed hemisphere providing the most pleasant listening experience might be due to wider range of participant positions in the virtual environment in small (3 to 4) to medium (5 to 7) sized audio teleconferences. Front and rear hemispheric audio clips appeared to be similarly well accepted. This might again be due to the relatively common front / rear placement errors associated with human hearing.

The panned stereophonic samples appeared to have produced an experience better than that of the monophonic, however reduced in comparison to that of spatial audio, lacking the inherent ‘surround’ sound feel associated with this. However, flat stereo samples proved to be the most unpopular sample, providing a slightly worse user perception experience than that of a monophonic sample. This is probably due to the extremely unnatural sound field that single ear audio provides and the listener fatigue this can create. In addition the intelligibility test transitioned at a comfortably slow pace whereas in the real-life a standard conversation tend to take place at a faster rate, people often interrupting and talking over each other. This increased pace in perception tests will probably have served to exaggerate the effect that characterises the flat stereo.

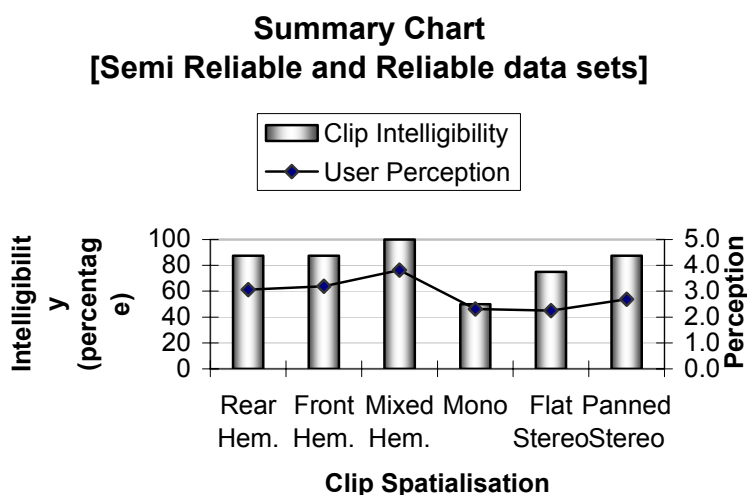


Figure 4-12: Summary chart of the results from the audio intelligibility and perception tests.

5. WEB Based User Survey

A WEB survey can be a very efficient way of collecting data. The major advantages with web surveys might be that it is a fairly cheap and fast way of collecting large amount of data within limited time scale. However, there are some disadvantages with web surveys, such as collecting the recipient contact details can be time consuming and getting people to respond to the survey is a different matter.

The primary purpose of the audio conferencing user survey was to identify user profiles, usage and the user experience with the current audio conferencing systems. Specifically, we were interested in the mobile phone users and the ways the mobile based audio conferencing could be enhanced.

5.1 Survey Design

The survey was designed using Vodafone web survey tool, WebAppFarm, which allowed easy implementation and maintenance of the survey as well as fast data collection. The survey questions were formed using multiple choice and open ended questions. The multiple choice questions were chosen for the survey implementation because they were quick and easy to answer to. All of the questions were specified as compulsory, which guaranteed a larger set of valid data feedback. The open ended questions were optional, in order to avoid invalid or 'made up' responses. The survey took approximately 11 minutes to complete and it was divided into four major sections:

1. User profiles

The first part of the survey was to identify the audio teleconferencing user profiles. Respondents were grouped by age, gender and their department or profession.

2. Usage of audio conferencing

In order to gain deeper understanding of the usage of the current audio conferencing, respondents were asked several usage related questions (See appendix A for survey details).

3. User experience

User experience section of the survey was to investigate the problems that users experience during the audio conferencing. The questions asked were targeted to learn more about the issues with intelligibility and perception during the audio conference calls.

4. Improvement feedback

In the final part of the survey, respondents were given multiple choice questions about possible ways of improving the audio conferencing. An open ended question was provided to encourage the respondents to use their imagination and creative thinking to come up with usable solution ideas for enhancing the audio conferences.

5.2 Respondents

The web survey was distributed to 860 people who were Vodafone Europe employees or partners. All of the contacts on a distribution list were property of Vodafone R&D-UK database.

5.3 Survey procedure

Respondents received an email invitation which provided an URL link to the web survey page. Respondents were asked to fill in the web based form by following the instruction given on the page.

5.4 Results

The time spent on the survey varied from 6 to 15 minutes, mean average of 11 minutes. Total of 177 people took part in the web survey and 17 of them had not used the audio teleconferencing and were therefore excluded from data analysis leaving filtered population of 160 responses. 62% (99) of the survey respondents were 18 to 34 years of age, 35% (56) were age between 35 and 55 and rest 3% (5) were over 55 years. Majority of the respondents (61%) were professionals in technical, marketing or sales fields. 72% (116) of the population were men and 28% (84) were female. Overall survey response rate was 20%.

5.4.1 User Profile and Usage

37% of the respondents participated in the audio teleconferences *every week or more*. The majority, 67% of the respondents participated *at least once in two weeks*.

68% of the conference calls lasted approximately 1 hour.

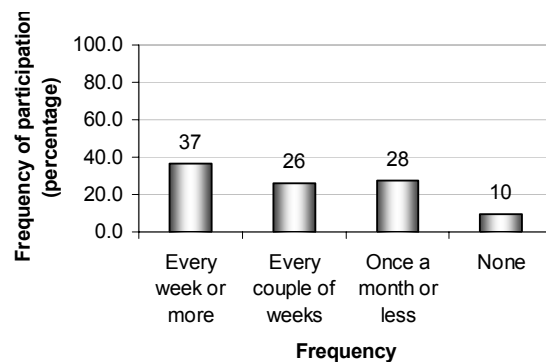


Figure 5 – 1: Frequency of participation in audio teleconference

When saying ‘approximately’ 1 hour, we mean the approximate values given in the survey as the options for the call lengths: 30 minutes, 1 hour, 2 hours or more than two hours. By providing approximate values for the options for the call lengths, our belief was to avoid problems the respondents might have faced when stating accurate values for the call lengths.

48% of the survey respondents stated that they usually participated in the audio conferences with 5 to 6 people. 31% of the respondents were participating in the conferences with more than 7 respondents. User survey also revealed that majority (66%) of the audio conference users participated in the 2/3 of their conference calls through mobile phone. The reason for

large number of mobile users might be because the survey was distributed to Vodafone employees and partners. Therefore, the validity of the survey data must be evaluated bearing in mind that these results can be very dependent on the respondents of a large global mobile phone operator.

41% of the survey respondents used a mobile phone and a single earpiece to connect to the audio conference calls. 24% of the respondents used a mobile handset only. The dual earpiece was used by 21% of the respondents.

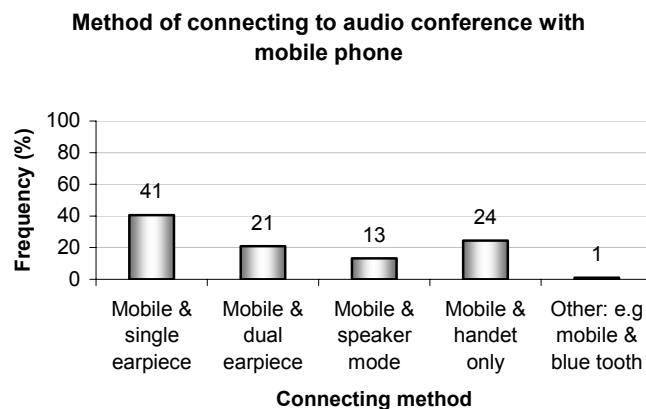


Figure 5-2: Method of connecting with mobile phone to audio teleconferences

These results showed that monophonic audio output (65%) seemed more popular than the stereophonic. This might have been due to the fact that the mobile phone did not support stereophonic audio output or it was a personal preference to use single earpiece or mobile handset only.

5.4.2 User Experience

One of the major reasons for collecting data through user survey was to identify issues appearing with the user experience of the audio conferencing. Issues identified were in connection with conference call intelligibility and participant identification.

Statistical analysis showed that 91% of the respondents stated that background noise made individual participants within the audio conference hard to hear and 81% said that noise made the identification of the participants difficult. 83% of respondents therefore commented that inconsistent voice volume between participants made the conference

irritating and hard to follow. Conference participant identification due to similar sounding voices was found to be a common issue by 74% of the respondents and 62% of the respondents said that high number of participants made identifying individual participants difficult

Open ended answers revealed that majority of the respondents believed that differing user interfaces across mobile platforms made tasks such as muting the phone very difficult or even impossible. This was mainly due to difficult muting functionality on the current mobile devices.

According to the findings of the user survey, clear issues can be identified with the intelligibility and participant identification during the audio conferencing. The major reasons for the problems being background noise, inconsistent voice volume, similar sounding voices and high number of participants. All of these issues could possibly be enhanced by the spatial audio and the visual-interactive user interface functionality, as discussed later.

5.4.3 Improvement suggestions

The survey respondents were asked how they would improve the current audio conference systems. 80% of the respondents commented that providing automatic volume control over conference participants would be beneficial. This would mean that participant's voice volumes would be equalised at the beginning of the call in order to provide more pleasant audio conference experience. In addition to this, each conference participant would have a manual control over the volumes of the conference participants. As many as 63% of the respondents suggested that functionality for providing easy invite of new conference participants would be useful. This would allow instant invitation of new participants into the conference call and could become useful, e.g. in a situation when one's expert opinion is instantly needed.

The feedback gained in the open ended questions provided improvement suggestions which contained three major findings:

1. A visual indication of the participants would help identifying conference participants. This would mean that conference participants would have a visual indication (e.g. text or picture icon) on the user interface. A flashing participant icon could indicate when and which participant is speaking.
2. The background noise reduction algorithms should be applied in the design. However, this requirement is not further considered in this research study.
3. An easy method of muting oneself within the conference is required. The mobile users suggested that muting oneself should be just a simple button press, instead of the current menu option. Easy, fast muting functionality would improve the conference call quality by decreasing the background noise during the mobile audio conferencing, e.g. traffic.

6. Product Demonstration and Focus Groups

The product presentation, demonstrations and focus groups were conducted to several small groups of potential users of the enhanced audio conference systems. The participants were employees from Vodafone UK and from the other European offices.

The product demonstration was carried out at the premises of Vodafone R&D in UK. The demonstration started with a short presentation, moving onto the audio conference system demonstration. Total of 6 demonstration sessions were carried out during a period of 3 weeks. 6 groups of 4 carried out a conference call with the ACE prototype application.

6.1 ACE Demonstrator

A portable demonstrator was set up, containing a wireless local area network (LAN) with two servers and 4 network clients. Portable Sony VAIO U70 devices with 512MB of Ram and a 20GB hard drive were used as network clients. The device had 800x600 pixels touch screen with an integrated pointing stick and mouse buttons. Each device had 802.11 wireless network port connected to wireless router which enabled the wireless network connection to the server. The call quality followed the 8 kHz GSM quality.

Demonstrator had a simple, 'easy to use' visual-interactive user interface, illustrating the audio conference space. The graphical user interface indicated the active conference participants at their positions in a mixed hemisphere, spatial audio environment. The numbered items (1 to 3) were indicating the conference participants (figure 6-1).



Figure 6-1: Portable, Sony VAIO U70 used as a network client.

The fourth participant was an image of a head in the middle of the conference room which demonstrated the listening participant itself.

The functionality of the demonstrator was limited. Participants were able to swap between the monophonic and spatial audio during the demonstration call by simply clicking the option provided on a screen. This enabled the participants to evaluate and make a distinction between the monophonic and spatial audio conferencing performance and quality. The participants had also an option to mute their microphone temporarily. By dragging the conference participants (circles with numbers) closer to the listener, the voice volume of this particular voice would increase, and if the participant was dragged further away from the listener the voice volume would decrease. Listener had a control to move the participant around on the screen, for example in order to separate similar sounding voices from each other.

6.2 Participants

Six combinations of demonstrations and focus group sessions were conducted. Each group involved 4 participants, a mixture of female and male participants.

6.3 Procedure

Participants were asked to discuss about the demonstrator performance with each other during the conference call. This motivated the participants to make easy discussion with each other. Each demonstration session lasted approximately 10 to 20 minutes depending on the level of enthusiasm. During the demonstration, participants were asked to switch between the monophonic and spatial stereophonic audio modes in order to identify differences in conference performance and intelligibility. Participants were also encouraged to test the visual-interactive functionality, by moving participants around the GUI.

After the demonstration, participants joined a focus group session to discuss the performance differences between the traditional audio conference system and demonstrated enhanced version. The focus group discussion involved topics in spatial 3D audio and visual-interactive functionality. Focus group leader was to keep the conversation up as well as interactive, giving each participant a chance to talk and bring up ideas for the potential enhanced solution. Focus group feedback was recorded and later analysed.

Some basic questions were presented to the participants:

1. Your first impression on the enhanced audio conferencing system?
2. What was good about it? How about what is it lacking?
3. How does the enhanced version compare to the traditional system?
4. Would you be prepared to use dual earpiece?

6.4 Feedback Results

A large set of feedback data was gained from the focus groups. The findings from the focus groups were very similar as the ones from the web based user survey which meant that the feedback gained was common among the conference users. The aim of the focus groups was to collect ideas and suggestions for the potential solution of the mobile based audio conferencing: graphical user interface visualisation and interactive functionality. The results

were considering only the qualitative data analysis. The most frequently suggested ideas formed a set of user requirements. The user requirements were then documented and delivered to the solution developer as a guideline for the future development.

R1: The conference participants should be visualised in name or text labels on a GUI in order to aid in participant identification (demonstrated in number labels).

R2: The name label of the speaking participant should get highlighted or resized when participant would speak (demonstrated in changing colour).

R3: The participant voice volumes should be equalised at the beginning of the conference call to create more pleasant listening environment. System should automatically measure the output voice frequency of each conference participant and equalise the volume regarding to that value.

R4: The participants should have a manual control over the voice volumes by moving participant (icons on a GUI) closer and further away from the listener. Further the participant would be from the listener's icon, quieter the voice volume would get. Closer the participant would be to the listener's icon, louder the voice volume would get (this functionality is already demonstrated).

R5: Simple, manual muting by clicking own name icon on a GUI or pressing a button on a keypad, would improve the conference call quality by reducing the amount of background noise.

R6: The focus group participants agreed that positioning or grouping the conference members by e.g. department, company or project could aid in participant identification. Also this functionality could be useful for separating similar sounding voices from each other (demonstrated).

R7: Quickly inviting or adding new participants to the conference call was also as a very useful function. This would allow instant invitation of the new participants into the conference call by sending an invitation message or alert.

R8: The participants suggested that the ability to send instant messages ('private chat') within conference participants would be useful while on a conference call. This would allow private discussion and commenting with the chosen participants.

7. Discussion

This thesis presents a framework, called Audio Conferencing Enhancements, which offers visual and interactive user interface functionality in combination with 3D spatial audio to support audio conferencing on the mobile phones. These enhancement techniques are to improve the current conferencing systems by eliminating the issues with speech intelligibility and speaker identification.

Firstly, our research study was to find answers to our research question about the effectiveness of the 3D audio:

1: Can the spatial, 3D audio improve the speech intelligibility and audio perception of the audio conference systems?

When introducing the reproduction of the sound to create spatial audio space by means of HRTF, we faced a question about the dual and single earpiece usage among the users. This study has proved that the HRTF are ideally suited to headphones. However, the answer to the preferred way of connecting to mobile audio conferencing remains unclear. This study showed that mobile users connected most frequently to the audio conference calls through single earpiece. In order to gain the full potential benefit from the enhanced audio conferencing solution, dual earpiece should be applied. Therefore, a core question remains unanswered:

Would the ACE users be prepared to use dual earpiece during the conference calls?

Our study findings support Marentakis and Brewster [2005], indicating that the use of headphones or dual earpiece might isolate the conference user from their real world audio environment. However, our focus group study showed that the majority of the participants were positive about using dual earpiece set during the conference calls. The reasons for the positive feedback being that the dual earpiece would help blocking out the noise around the listener and would therefore help concentrating better.

The participants who were not willing to use the dual earpiece for connecting to audio conference calls commented that they were willing to hear what was happening in the real audio environment during the conference call. They were also not particularly happy to carry dual earpieces around with them. In this particular argument we are unable to reach to the clear final conclusion about the earpiece usage within this study. Most of the respondents did not have the real life experience with the dual earpiece usage and therefore their answers were based on the ‘feeling’ they had in that particular moment. Surprisingly, some of the committed single earpiece users changed their strict opinion after the enhanced audio conferencing demonstration session.

Overall, the results from the intelligibility and perception tests showed that subjective listening performance was improved when 3D, spatial audio was used. The speaker identification, speech intelligibility and perception were remarkably improved when using mixed hemisphere audio placements. The test results showed that the spatial, 3D audio can help to increase the intelligibility of the multi-person conversation in a compressed audio environment. The spatial audio was also experienced to be more natural and effective compared to that of a standard monophonic audio output. The spatial positioning of the conference participants during the call provided additional memory cues creating a more efficient use of our working memory. Therefore, identifying the conference participants became easier.

Naturally, the mixed hemisphere placement of the audio sources appeared to be the most intelligible. The front hemisphere placement was expected to be more accurate than the rear

hemisphere placement. However, the results gained from the reliable and semi-reliable test subjects indicated that the intelligibility of the front hemisphere was similar of that of rear hemisphere. This might have been due to the front / rear confusion [Gardner, 1999] which is common with the spatial audio space.

Surprisingly, the panned stereo samples provided similar results to that of the front and rear hemisphere spatial samples. However, the performance was noticeably reduced compared to the mixed hemisphere.

The perception test results supported the intelligibility test results, indicating that the spatial, mixed hemisphere audio offered most pleasing listening experience in a multi-person audio environment. The panned stereophonic samples appeared to be experienced better than the monophonic audio. However spatial audio was again preferred due to 'surround' sound feel associated with this.

The second question presented at the beginning of the research:

What are the user requirements for the visual-interactive functionality on a mobile based audio conferencing application?

The user survey and the focus groups provided valuable information for the user requirement 'specification'. The major findings were focused on the user control over the volume and muting as well as visualisation of the participants. The participants were most interested in improving the conference call quality in order to identify the conference call participants. They also find it important to flexibly interact with the other conference participants, such as sending private messages during the conference call.

Finally, we can conclude that in order to bring an enhanced audio conferencing solution to the markets, larger scale market research should be carried out within the consumers. The functionality should be kept simple and the service and device costs should be kept to minimum. The ACE project is still on-going and is currently further investigating the market place and the technical details for the potential product.

8. Summary

As discussed in this thesis, the audio conferencing can be very efficient way of collaborating between people. The enhanced audio conferencing solution using 3D, spatial audio in addition to visual-interactive user interface functionality, shows the great prospective for the mobile phone users. The ACE research study showed that 3D, spatial audio and visual-interactive functionality can improve the audio conference intelligibility and audio perception. The core idea of the audio conferencing enhancements was to improve the quality of the distance collaboration. However, the ACE system is still experimental and further user studies are to be carried out in order to evaluate the effect on the communication and social processes and later determine how the system should be implemented to provide the best collaborative experience. In addition, further marketing research studies and business cases are to be carried out in order to understand the market place and the real customer requirements. The findings of the ACE study are considered to be valuable for the future development of the mobile audio conferencing services.

References

- Aoki, P.M., Romaine, M., Szymanski, M.H., Thornton, J.D., Wilson, D. and Woodruff, A. (2003), The mad hatter's cocktail party: a social mobile audio space supporting multiple simultaneous conversations, *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 425–432, Ft. Lauderdale, USA, 2003.
- Arons, B. (1992). A Review of the cocktail party effect, *Journal of the American voice I/O Society*, 12, pp. 35-50, 1992.
- Baldis, J. J. (2001). Effect of Spatial Audio on Memory, Comprehension, and Preference during Desktop Conferences, *Proceedings of the ACM Computer Human Interaction, Human Factors in Computing Systems Conference*, pp.166-173, Washington, USA, 2001.
- Billinghurst, M., Bowskill, J., Jessop, M. and Morphett, J. (1998). A Wearable Spatial Conferencing Space, *Second International Symposium on Wearable Computers (ISWC'98)*, pp.76, 1998.
- Billinghurst, M., Cheok, A., Prince, S. and Kato, H. (2002). Real World Teleconferencing, *Proceedings of IEEE Computer Graphics and Applications*, vol. 22, 6, pp. 11-13, December, 2002.
- Brungart, S.D., Ericson, M.A. and Simpson, B.D. (2002). Design Considerations for Improving the Effectiveness of Multitalker Speech Displays, *Proceedings of the International Conference on Auditory Display*, Japan, July 2-5, 2002.
- Burgess, D.A. (1992). Techniques for Low Cost Spatial Audio, *Proceedings of the Fifth Annual Symposium on User Interface Software and Technology (UIST '92)*, ACM, , pp. 53-59, New York, 1992.
- EBU-UER. (2003). EBU Subjective Listening Tests on Low – Bitrate Audio Condecs.
- Erickson T. and Kellogg, W. A. (2000). Social Translucence: An Approach to Designing Systems that Support Social Processes, *ACM Transactions on Computer Human Interaction in the New Millennium, (TOCHI)*, vol.1, 1, pp. 59–83, 2000.
- Evans, M.J., Tew, A.I. and Angus J.A.S. (1997). Spatial audio Teleconferencing – Which way is better?, *Proceedings of 4th International Conference on Auditory Displays*, Palo Alto, California, November 2-5, 1997.

- Gardner, W.G. (1999). 3D Audio and Acoustic Environment Modelling, <http://www.harmony-central.com/Computer/Programming/3d-audio.pdf>.
- Goose, S., Riedlinger, J. and Kodlahalli, S. (2005). Conferencing3: 3D audio conferencing and archiving services for handheld wireless devices, *Wireless and Mobile Computing*, vol.1, 1, 2005.
- Greenhalgh, C. and Benford, S. (1995). MASSIVE: A Collaborative Virtual Environment for Teleconferencing, *ACM Transactions on Computer-Human Interaction*, Vol 2, 3, pp. 239–261, 1995.
- Hindus, D., Ackerman, M.S., Mainwaring, S. and Starr, M. (1996). Thunderwire: A field study of an audio-only media space, *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'96)*, pp. 238–247, ACM, 1996.
- Johnson, C.W. and Dell, W. (2003). The Limitations of 3D Audio to Improve Auditory Cues in Aircraft Cockpits, *Proceedings of International Systems Safety Conference*, International Systems Safety Society, Unionville, USA, pp. 990-999, 2003.
- Kan, A., Pope, Graeme., Jin, C. and Van Schaik, A. (2004). Mobile Spatial Audio Communication System, *Proceedings of ICAD'04 10th international conference on Auditory Display*, Sydney, Australia, July 6-9, 2004.
- Kilgore, R., Chignell, M. and Smith, P. (2003). Spatialized Audio Conferencing: What are the benefits? *IBM Conference Proceedings of the Centre for Advanced Studies on Collaborative research*, pp. 135–144, Toronto, Canada, 2003.
- Marentakis, G.N. and Brewster, S.A. (2005). Effects of reproduction equipment on interaction with a spatial audio interface, *Conference on Human Factors in Computing Systems CHI '05*, pp. 1625–1628, Portland, USA, 2005.
- Mortlock, A., Machin, D., McConnell, S. and Sheppard, P. (1997). Virtual conferencing, *BT Technology Journal*, Vol. 14, No. 4, pp. 120–129, October, 1997.
- Nelson W. T., Bolia R.S., Ericson M.A. and McKinley R.L. (1998). Monitoring the simultaneous presentation of spatialized speech signals in a virtual acoustic environment, *Proceedings of the 1998 IMAGE Conference*, pp.159-166, June, 1998.
- Nelson W. T., Bolia R.S., Ericson M.A. and McKinley R.L. (1999). Spatial Audio Displays for Speech Communications: A Comparison of Free Field and Virtual Acoustic

- Environments, *Proceedings of the Human Factors and Ergonomics Society*, 43rd annual meeting, pp.1202-1205, 1999.
- Olson, G.M. and Olson, J.S. (2000). Distance Matters, *Proceedings of Human-Computer Interaction 2000*, Vol. 15, pp. 139–178, 2000.
- Olson, J.S. and Teasley, S. (1996). Groupware in the Wild: Lessons Learned from a Year of Virtual Collocation, *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, Boston, US, pp. 419-427, 1996.
- Shaw, S.J. (1996). Directional Perception in the Human Auditory System, *Organismic and Evolutionary Biology Journal*, vol.3, pp.135-140, 1996.
- Stifelman, L.J. (1994). The Cocktail Party Effect in Auditory interfaces – A Study of Simultaneous Presentation, *Technical Report, MIT Media Lab*, September 1994.
- TGS-SA Wg4. (2004) Audio codec selection tests: Reports from the Subjective Testing Labs, Phoenix, USA, 15–18 March, 2004.
- Vause, N. L. and Grantham, D.W. (1998). Speech Intelligibility in Adverse Conditions in Recorded Virtual Environments, *International Conference on Auditory Display*, November, 1998.
- Wenzel, E.M., Stone, P.K., Fisher, S.S. and Foster, S.H. (1990). A System for Three-Dimensional Acoustic “Visualization” in a Virtual Environment Workstation, *IEEE Proceedings of the First Conference on Visualization '90*, pp. 329–337, San Francisco, USA, 1990.
- Williams, E. (1997). Experimental Comparisons of Face-to-Face and Mediated Communication, *Psychological Bulletin*, Vol. 16, pp. 963-976, 1997.
- Yamazaki, Y. and Herder, J. (2000). Collaborative Virtual Environments, *Proceedings of the 3rd international conference on Collaborative virtual environments*, pp. 207–208, San Francisco, California, USA, 2000.
- Zielinski, S., Rumsey, F. and Bech, S. (2002). Subjective audio quality trade-offs in consumer multichannel audio-visual delivery systems. Part 2: Effects of low frequency limitation, *Proceedings of the AES 22nd International Conference*, Espoo, Finland, 15–17 June, 2002.

APPENDIX

WEB Survey (WebAppFarm survey tool format)

vodafone™

Vodafone Survey Tool

Surveys

Reports

Libraries

Invitations

Styles

Logout

Surveys > ACE - R&D

1. Properties

2. Edit

3. Style

4. Permissions

5. Activation

Edit Survey

View: All Pages

Hidden Items

New Page

Preview

Print

Page 1

Add Item

Conditions

Branching

Copy

Move

Delete

Branching rules:
If answer to "How often do you participate in audio teleconferences?" is equal to "None" then go to Page 4

Edit

Move

Copy

Insert

Delete

Audio Teleconferencing User Survey

Welcome!

This audio teleconferencing survey is targeted at people like **you**, who participate in audio teleconferences. We are interested in your previous experience, personal preference and future expectations of audio teleconferencing systems. The survey forms part of a larger project that aims to investigate potential audio teleconferencing enhancements. The Project is being carried out by the User Interface Technologies team of Vodafone Group Research and Development (R&D). By participating in this survey you will provide us, and the business, with valuable information which could help us enhance audio teleconferencing solutions.

This questionnaire takes approximately **5 minutes** to complete.

Everyone participating in this survey will automatically be entered into a prize draw, to win a bottle of Champagne.

If you have any questions on this survey, please email leena.vesterinen@vodafone.com or call +44 (0)776 647 8542.

5. Age group:

☐ 18 - 34
☐ 35 - 55
☐ 55+

Edit

Move

Copy

Insert

Delete

6. Gender:

☐ Male
☐ Female

Edit

Move

Copy

Insert

Delete

7. How often do you participate in audio teleconferences?

☐ None
☐ Once a month or less
☐ Every couple of weeks
☐ Every week or more

Edit

Move

Copy

Insert

Delete

Page 2

There are no conditions. This page will always be displayed.

Edit Move Copy Insert Delete

* **8. Including yourself, what is the average number of participants in the conference calls you attend?**

☐ 3 - 4
☐ 5 - 6
☐ 7 - 10
☐ More than 10

Conditions There are no conditions. This item will always be displayed.

Edit Move Copy Insert Delete

* **9. How long on average do your audio teleconference calls last?**

☐ Approx. 30 minutes
☐ Approx. 1 hour
☐ Approx. 2 hours
☐ More than 2 hours

Conditions There are no conditions. This item will always be displayed.

Edit Move Copy Insert Delete

* **10. In general what percentage of your audio teleconferences take place using a mobile device (as opposed to a fixed line)?**

☐ none, all teleconferences are from fixed line
☐ Less than 33% from mobile
☐ Between 34 - 66% from mobile
☐ Most of the calls are from mobile

Conditions There are no conditions. This item will always be displayed.

Edit Move Copy Insert Delete

* **11. Are the majority of your audio teleconference calls for business or personal purposes?**

☐ Personal use only
☐ Mainly personal, less than 30% for business
☐ Both, approx. 50% each
☐ Mainly business use, less than 30% for personal use
☐ Business use only

Conditions There are no conditions. This item will always be displayed.

Edit Move Copy Insert Delete

12. How do you usually connect to audio teleconferences?
Select all that apply.

☐ Fixed line phone in speaker mode
☐ Fixed line phone & handset only
☐ Fixed line phone & headset
☐ Mobile phone & handsfree with single earpiece
☐ Mobile phone & handsfree with two earpieces
☐ Mobile phone in speaker mode
☐ Mobile phone handset only
☐ Other, Please give details:

Conditions There are no conditions. This item will always be displayed.

Edit Move Copy Insert Delete

13. User Experience of audio teleconferencing:

Select appropriate options for the following claims.

	Strongly disagree	Disagree	Do not know	Agree	Strongly agree
It is often difficult to distinguish when a different person starts talking.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
High number of participants on a call makes it difficult to follow the conversation.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Similar sounding voices make it difficult to identify individual speakers.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Background noise makes it hard to identify the participants on a call	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Inconsistent voice volume between participants adversely affects the communication.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Background noise can make the conference hard to hear and follow.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Voice echo and other acoustic effects adversely affects the call performance.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Conditions There are no conditions. This item will always be displayed.

Edit Move Copy Insert Delete

14. Please describe any other problems you have experienced with audio teleconferencing? How does it affect the call performance?

Conditions There are no conditions. This item will always be displayed.

Edit Move Copy Insert Delete

15. Which, if any, of the following options do you think could improve your experience of audio teleconferences?

Select all that apply.

- ☐ Send instant messages to selected participant (private chat).
- ☐ Dynamically controlling the volume of individual participants so that the same volume level is maintained across the call.
- ☐ Document sharing while on a call.
- ☐ Possibility to conduct presentations.
- ☐ Invite new participants to join the conference (quick invite).

Conditions There are no conditions. This item will always be displayed.

Edit Move Copy Insert Delete

16. Please provide any suggestions that you feel could improve the user experience of audio teleconferencing:

Conditions There are no conditions. This item will always be displayed.

	VIDEO conferencing:	
	There are no conditions. This item will always be displayed.	
	* 17. Have you tried video conferencing from your desk?	
	<input type="radio"/> Yes <input type="radio"/> No	
	There are no conditions. This item will always be displayed.	
	* 18. Would you like the idea of video conferencing from your desk?	
	<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Maybe	
	There are no conditions. This item will always be displayed.	
	* 19. Would you like the idea of video conferencing on your 3G device?	
	<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Maybe	
	There are no conditions. This item will always be displayed.	
	20. Any other comments?	
	<input type="text"/>	
	There are no conditions. This item will always be displayed.	
Page 4		
There are no conditions. This page will always be displayed.		
<p>Thank you for your interest.</p> <p>Your details have been sent to our prize draw.</p> <p>(please click finish to complete the form)</p>		
	There are no conditions. This item will always be displayed.	