

Talk to me: The Influence of Audio Quality on the Perception of Social Presence

Christina Dicke
Human Interface Technology Lab
University of Canterbury
Christchurch, New Zealand
christina.dicke@hitlabnz.org

Viljakaisa Aaltonen, Anssi Rämö,
Miikka Vilermo
Nokia Research Center
Tampere, Finland
{firstname.lastname@nokia.com}

ABSTRACT

In this paper, we compare the impact of monophonic, stereophonic, and binaural human speech recordings in terms of their ability to induce the feeling of presence and influence the understanding of the emotional state the speakers were in. These factors are generally important in entertainment applications, for example when conversing with a non-player character or in mediated synchronous human-to-human communication. Our results show a significant advantage of binaural over mono and stereo sound for inducing the sense of being present in an (virtual) environment. Furthermore, we found that listening to a stereophonic recording of a conversation leads to a significantly stronger understanding of the emotional state of speakers than listening to a mono or binaural recording.

Categories and Subject Descriptors

H.1.2 [Information Systems]: User/Machine Systems – Human Factors

General Terms

Human Factors, Experimentation, Measurement

Keywords

Spatial, Binaural, Stereo, Mono, Dialogue, Speech, Social Presence, Presence, Emotional Understanding

1. INTRODUCTION

3D technology is more and more adopted by the gaming and movie industry and has become widely accepted by end-users. Succeeding the emergence of advanced 3D visual effects, spatial audio has become a standard in recent versions of games like World of WarCraft, Bioshock or Halo as well as in virtual worlds like Second Life. Following this trend, mobile handheld communication devices recently started supporting 3D audio by implementing spatial sound libraries like OpenAL or fmod.

Immersion and the perception of social presence are considered core components of the overall experience not only in gaming and virtual environments but also in mediated collaborative work environments and mediated real-time social communication.

Biocca & Harms [1] define social presence as a sense of being with another in a mediated environment. Both Biocca & Harms and Lombard & Ditton [2] point out that the sense of accessibility of the emotional and intentional state of the other and the emotional interdependence contribute to the perception of social presence. Although mechanisms leading to the perception of social presence vary between different dialogue based applications, we believe the underlying sensory immersion, i.e. immersion induced by the quality of the sound itself, not the content or application, is comparable.

While listeners of mono or stereo sound often perceive the sound source to be positioned inside of their heads, binaurally recorded or synthesized speech seems more realistic as it offers externalization cues. It stands to reason that acoustic realism, including spatiality, might be a highly influential factor to the perception of social presence, but only few researchers have investigated this assumption.

The study presented in this paper explores if there is a difference between spatial and non-spatial sound in terms of the perception of presence and the understanding of the emotional state of several speakers. We aim to understand whether the quality of sound sources utilizing human speech hinders or facilitates the experience of presence and the capacity to know emotionally what another is experiencing and hence influence the perception of being socially present in the (mediated) environment.

2. Related Work

There is a comprehensive and very diverse body of research on presence and social presence from the early 1980s onwards. Various definitions of social presence have been suggested, from the most straightforward sense of ‘being with others’ [3] to the definition by Biocca & Harms [1] where social presence is the ‘sense of being with another in a mediated environment [...] the moment-to-moment awareness of co-presence of a mediated body and the sense of accessibility of the other being’s psychological, emotional, and intentional states’.

(Tele)presence most commonly refers to the physical and spatial sense of ‘being there’, having the perception of being physically present in a remote environment. Some authors make the distinction between ‘telepresence’ and ‘virtual presence’, where the former denotes the feeling of being present at a remote location, and the latter indicates the experience of being immersed in a virtual environment [4]. A detailed discussion of this topic would go beyond the scope of this paper (we recommend [5], [6], [7], and [8] for an overview on the topic). However, for the purpose of our research, we use the definition of social presence by Biocca & Harms [1] and the definition of presence proposed by [3], [9], i.e. the sense and feeling of ‘being there’ in a mediated scene or virtual environment.

The impact of 3D technologies on the perception of (social) presence has been comprehensively studied for visual experiences, but not quite as thoroughly for multimodal or audio only experiences.

Lombard & Ditton [2] include a broad review of research conducted prior to 1997 on presence and include a more specific review of the effects of sound on the sense of presence. Despite rather mixed findings they summarize in their overview, they assume it likely that spatial audio increases the sense of presence. This assumption has later been supported by experiments of Loomis [10] and Hendrix & Barfield [11] finding a positive

influence of spatial realism and externalization cues on the perception of presence. Hendrix and Barfield compared the effects of spatialized and non-spatialized sound in a virtual environment on the user's perception of presence. They found that the addition of spatialized sound significantly increases the reported level of presence but does not have an impact on the perceived realism of the virtual environment [11].

Freeman & Lessiter [12], however, could not verify increased presence ratings for multi-channel audio when presenting participants with a rally car video sequence with accompanying synchronised audio. They argue, though, that these findings may be due to no perceivable advantage of the multi-channel presentation over the stereo presentation. Nevertheless, they found that enhancing the bass content and sound pressure level increases presence ratings.

Västfjäll [13] presented participants with music comparing the effect of mono, stereo, and six-channel sound on the participants' emotional reactions and ratings of presence. Music was chosen to convey either strong positive or strong negative emotions. Västfjäll found that both stereo and six-loudspeaker conditions were significantly more efficient than the mono condition in inducing emotional reactions. Ratings for presence were significantly higher in the six-loudspeaker condition compared to the two other conditions. He concludes that presence is linked to spatial sound reproduction and emotional reactions vary as a function of immersivity of the sound field.

Despite the comprehensive amount of research on presence in mixed and audio only (virtual) environments, surprisingly, the effects of spatialized and non-spatialized human speech on the perception of presence and the understanding of the emotional state of speakers have not been actively studied. The study reported in this paper is designed to gain insights into whether similar effects as found for musical and environmental sounds can be found for human speech.

3. Experiment

3.1. Design Rationale

We designed the experiment primarily to investigate whether there are differences in the human perception of monophonic, stereophonic, and binaural sound in regard to how humans perceive verbally communicated emotions. Furthermore, we were interested in identifying differences – and if so of which kind – in feelings of presence, i.e. having the perception of sitting among the people conversing, not in the listening booth.

To emotionalize participants and create an experience that induced the perception of being somewhere else, we chose a scenario most participants were familiar with either from literature, film, and television or from personal experience. Together with a scriptwriter we created a typical “confession-scene”, in which Paula, Heikki's long-term partner, confesses having an affair with Esa, in reaction to Heikki's proposal of marriage. The emotionality of the scene is amplified by Esa's presence at the table. The script was in Finnish, as the subjects were native Finnish speakers.

To support a smooth transition from the unfamiliarity of the test environment and procedure into the scene, the play started as a regular dinner invitation, with Paula and Heikki as hosts and Antti and Esa as their guests. During the first third of the play all characters are introduced. The atmosphere is friendly, relaxed and cheerful. The positive emotional climax is reached when Heikki proposes to Paula, immediately followed by the

turning point of Paula's confession. The second third is dominated by Heikki's feelings of utter surprise, incredulousness, and later anger as well as Paula's feelings of shame and guilt – expressed in a heated discussion. Emotions calm down during the last third but are not resolved. The play ends with Esa and Antti being asked to leave and them complying with Heikki's request.

The experiment compared three conditions:

Condition 1: A monophonic recording

Condition 2: A stereophonic recording

Condition 3: A binaural recording

As repeated exposure to the content would have diminished the element of surprise and presumably the level of emotionality, we chose a between-subjects experimental design.

The study was designed to test the following hypotheses:

H1: The stereophonic and binaural conditions differ significantly in terms of perceived presence from the mono condition. The perceived presence is strongest in the binaural condition. We assumed that feeling present in a “virtual” environment requires a sense of spatiality and layout of this environment. Hence we surmised the binaural condition, offering more spatial information, would outweigh both other conditions in terms of perceived presence.

H2: The stereophonic and binaural conditions differ significantly in terms of the understanding of the emotions acted out in the play. The understanding and alignment is stronger in the stereo and binaural conditions than in the mono condition. Given Västfjäll's [13] findings, we assumed that, similar to music, speech based audio induces stronger emotional reactions in both stereo and binaural playback than in mono.

The questionnaire used to evaluate the listening experience comprised basic demographic questions, questions about the sound quality and the participants' emotional state, as well as several items taken from the questionnaire on Mediated Communication Experience (ComXQ) [14]. We also asked participants to represent their perception of the scene through a sketch.

3.2. Audio Material and Recording Technique

To ensure maximum quality of the audio play, four professional actors and actresses from the Tampere Komediateatteri performed the play. As can be seen in figure 1, the actors/actresses were seated at a table. Figure 1 and 2 show Heikki as the host at the head site of the table (far left), to his left Antti and Esa, to his right the (imaginary) listener represented by a HEAD acoustics HMS manikin and several microphones and Paula (far right).

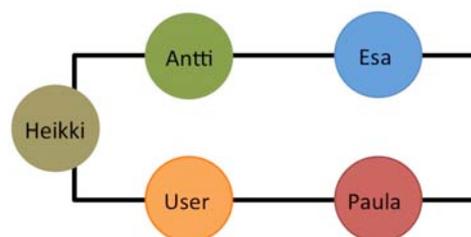


Figure 1: Illustration of the characters' seating order.



Figure 2: Actors and actresses during the recordings of the play.

Several props were used during the recording, for example a bottle of wine, wine glasses, plates, cutlery, and music (“easy listening”, which faded out by the end of the first third).

The play was recorded with multiple microphones in a recording room fulfilling the requirements set in ITU-R BS.1116 [15]. Background noise level was minimal and reverberation times were typical for a large living room.

Five of the recorded signals were used for the experiment. The audio capture was done on a computer located in the next room running Adobe Audition 3.0 in multitrack mode. The audio card used was RME Hammerfall DSP Multiface II. Presonus Firestudio was used as an additional ADAT A/D converter. The Presonus’s internal microphone pre-amplifiers were used for the five main recordings.

Mono recording was captured with a RØDE NT2-A microphone located in front of the manikin (see figure 3). The microphone was set to use an omnidirectional polar pattern. It was located slightly below the manikin ear level in order not to distort the binaural recording.

The stereo recording was done with an ORTF stereo capture configuration also with RØDE NT2-A microphones, which were set to a cardioid polar pattern. In ORTF capture the microphone capsules are located 17 centimeters from each other and spread at a 110° angle. ORTF provides both volume difference (with signals arriving at cardioid microphones at different angles) and timing difference as the sound arrives at the separated microphones with different delay. The binaural recording was done with a HEAD acoustics HMS II.3 artificial head and torso simulator. A type 3.4 artificial ear according to ITU-T Rec. P.57 [16] was used for recording.

During the experiment participants listened to the recordings in silent isolation booths [17] with Sennheiser HD-580 headphones. The listening level was set to be same for all recorded configurations.

3.3. Participants

Eighty-two participants volunteered for the experiment ranging in age from 15 to 54 years ($M = 33$ years), and were recruited within the community of a large company and several sport clubs. Forty-nine participants were male, thirty-three female. All participants were native Finnish speakers. Participants were randomly allocated to the three conditions. Three participants reported having minor hearing problems. They were not excluded from the experiment due to the negligibility of their hearing problems.



Figure 3: HEAD acoustics HMS II.3 artificial head and torso simulator used for the recording. Mono RØDE NT2-A in front of the mouth. Stereo RØDE NT2-A ORTF-pair just above the head pointing outwards. An additional XY microphone (not used for listening evaluation) is also visible at the top.

3.4. Procedure

Before their trial, participants were asked to sign a consent form and were briefly informed about the nature of the experiment. They were then familiarized with the listening booths and were instructed on how to put on and adjust the headphones. Subsequently they were asked to sit down in their assigned booth, relax, and focus on what they were about to hear. After the trial, participants were asked to fill out a questionnaire.

3.5. Experimental Design

A between-subjects design was used for this experiment. The eighty-two participants were randomly assigned to one of three groups. Participants in group one listened to the monophonic recording, group two to the stereophonic recording, and group three to the binaural recording. Group one comprised twenty-five participants, group two twenty-nine, and group three twenty-six participants. No explicit task was given to the participants other than to relax and focus on the conversation.

4. Results

A seven-point Likert scale has been used in the questionnaire (1=I totally agree and 7=I totally disagree). All tests were run using a one-way analysis of variance (ANOVA) with a fixed confidence level ($p\text{-value} = .05$) unless otherwise stated.

As one of the larger subsets of questions was designed to evaluate the listening experience in terms of perceived presence and emotional alignment to the content of the audio play, we examined this subset of eighteen questions for underlying dimensions reflecting these constructs. We ran a Principal Component Analysis (PCA) over a subset of eighteen variables and stepwise excluded three variables with communalities $< .6$. We finally ran the PCA again on fifteen variables ($KMO = .79$, Bartlett's Test of Sphericity $< .001$, Varimax Rotation). As there were less than thirty variables and communalities after extraction were greater than $.6$ we retained all factors with Eigenvalues above 1 (Kaiser's criterion). The PCA result indicated four factors explaining 68.5 percent of the total variance.

We could identify and name these factors as:

1. *Presence (five variables)*
2. *Emotional Understanding/Involvement (five variables)*
3. *Focus (two variables)*
4. *Authenticity (three variables)*

We ran another PCA on a second subset of the dataset. From the initial 18 variables we excluded four variables using the same criteria as stated above. The PCA ($KMO = .74$, Bartlett's Test of Sphericity $< .001$, Varimax Rotation) resulted in indicating three factors explaining 63.6 percent of the total variance.

These factors could be summarized as:

1. *Negative Emotions (containing six variables)*
2. *Positive Emotions (containing five variables)*
3. *Alertness (containing three variables).*

Due to the small sample size and the rather exploratory nature of the items chosen for the experiment we decided to use the sum score method to further analyze the data [18]. Summed factor scores preserve the variation in the original data, which is useful for the further analysis. We used only items suggested by the PCA (with a loading of $> .6$) and assigned cross-loading items to the score it loads higher on. All items on a factor were given equal weight, regardless of the loading value.

4.1. Presence

The *Presence* construct combines the following questions:

1. I felt like the participants in the conversation surrounded me.
2. I felt like I could reach out and touch the participants in the conversation.
3. I felt I was face-to-face with the participants in the conversation.
4. I would have liked to actively participate in the conversation.
5. I felt more like a participant than an observer of the conversation.

We found a significant difference ($F(2,77) = 3.74$, $p = .028$) confirmed by a post hoc Bonferroni test (with $p = .034$) between the mono ($N = 25$, Mean = 4.94, SD = 1.26) and the binaural ($N = 26$, Mean = 3.95, SD = 1.31) condition.

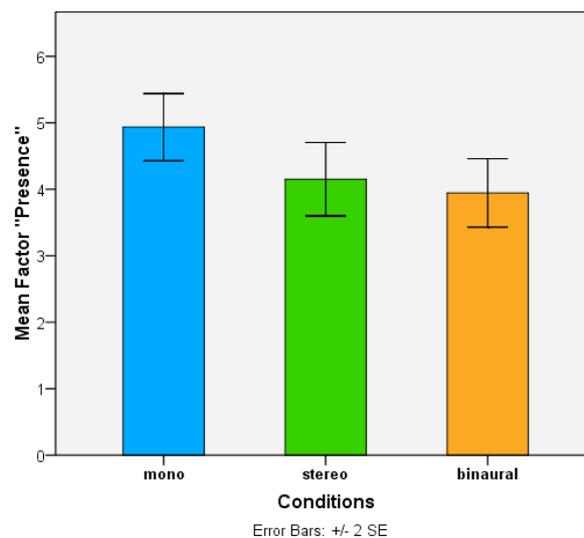


Figure 4: Mean values for the factor Presence over all three conditions.

As illustrated by figure 4, participants in the binaural group agreed significantly stronger to the questions listed above and hence had a stronger sense of presence than participants from the mono group.

As well as the responses to questions, participants were asked to sketch the situation they just listened to. Figures 5 to 10 show a representative sample of the drawings. The drawings are particularly useful for showing whether participants saw themselves as part of the situation (as exemplified by figures 5, 6, and 9) or as observers (exemplified by figures 7, 8, and 10).



Figure 5: Example of a drawing showing the participant (Number 88) sitting at the table among the characters of the play. The characters are all facing the participant.

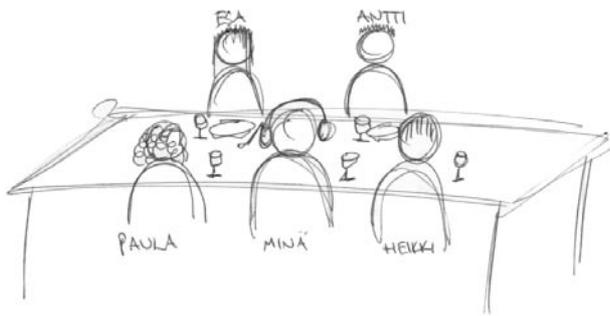


Figure 6: Example of a drawing showing the participant (Minä or “me”) sitting among the characters of the play. All characters are facing each other.



Figure 7: Example of a drawing in which the participant (Nerja) is not sitting among the characters but is standing in the distance. The drawing might also indicate a front-back confusion as the characters are located behind the participant.



Figure 8: Example of a drawing showing the audio scene as an actual play performed on a stage. The participant (Minä or “me”) is sitting among the

audience. A very clear physical and conceptual distinction between the participant and the characters of the play is made.

These drawings are not only interesting in terms of the participants’ sense of membership, but also, in some cases, as indicators for front-back-confusions (items that are located in front are mistaken for being located behind the participant). Figures 7 and 9 give examples of drawings of perceived front-back-confusions.

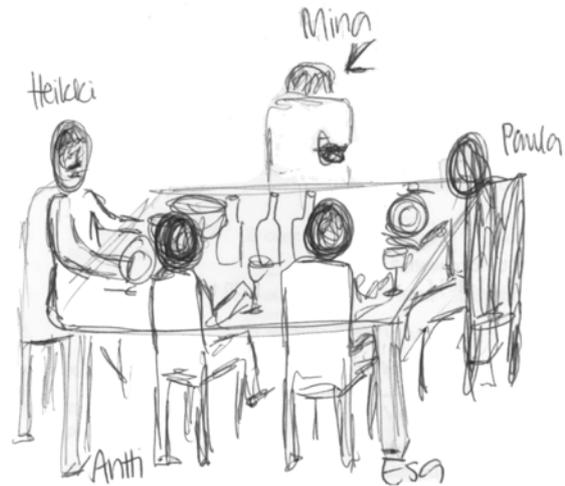


Figure 9: Example of a drawing showing the participant (Minä) as member of the party, but also indicating a front-back-confusion with the participant turning their back towards the table, i.e. perceiving the conversation to be coming from behind.

Some of the sketches even give an insight into how the emotional content has been perceived, as for example figure 10, in which Heikki, who just had to learn that his partner Paula is having an affair with Esa, is shown angry and in an agitated pose, whether Paula, who had to reject Heikki’s proposal and confess her affair, is shown in tears.



Figure 10: Example of a drawing showing the participant (Minä) to be physically separated from the characters. The drawing also depicts the characters emotions, showing Heikki enraged and Paula in tears.

In our analysis we solely focused on the position of the participants (at the table or in the distance) and refrained from further interpretation. As we did not specifically ask participants to draw the emotional state of characters or where exactly they were seated we cannot use these data in a comparative analysis.

As mentioned above, our analysis of the sketches support the results from the analysis of variance conducted on the factor Presence: As illustrated in figure 11, the sketches show a significant difference ($F(2,67) = 5.55, p = .006$) confirmed by a post hoc Bonferroni test (with $p = .026$) between the mono ($N = 20, \text{Mean} = 1.65, \text{SD} = 0.489$) and the stereo ($N = 25, \text{Mean} = 1.28, \text{SD} = .458$) condition. Likewise there is a significant difference (confirmed by a post hoc Bonferroni test with $p = .008$) between the mono and the binaural ($N = 23, \text{Mean} = 1.22, \text{SD} = .422$) condition.

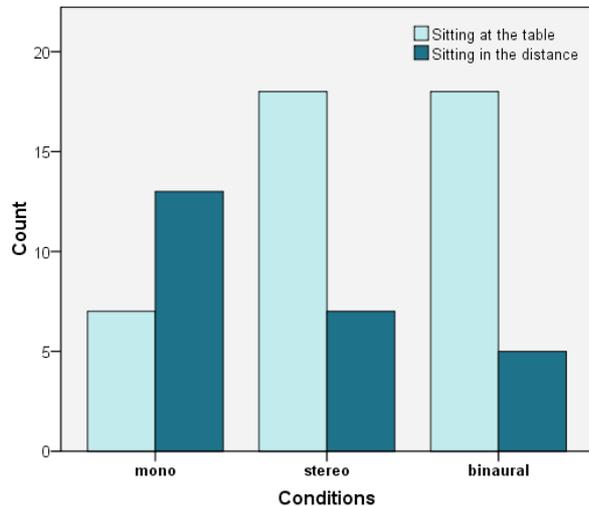


Figure 11: Counts over all three conditions for sketches depicting the participants as part of the group (“Sitting at the table”) or as observers (“Sitting in the distance”).

In conclusion, both the analysis of variance of the factor Presence and the interpretation and analysis of the sketches suggest that the sense of presence in a virtual scene or environment is significantly stronger when participants listen to binaurally recorded sound compared to monophonic sound. The difference between binaural and stereophonic sound is not as distinct, but still verifiable.

4.2. Emotional Involvement/ Understanding

We combined the following questions under the factor *Emotional Involvement/Understanding*:

1. The mood of the participants affected me.
2. I identified myself with one or more of the participants.
3. I knew how the participants felt.
4. I was emotionally moved by the conversation.
5. I was immersed in the situation.

An analysis of variance showed a significant difference ($F(2,77) = 3.582, p = .033$) between the mono ($N = 25, \text{Mean} = 3.48, \text{SD} = 1.194$) and stereo ($N = 29, \text{Mean} = 2.69, \text{SD} = 1.009$) condition confirmed by a post hoc Bonferroni test ($p = .029$). As can be seen in figure 12, participants tended to agree more strongly with the above listed questions in the stereo condition than in the mono condition. The binaural condition ($N = 26, \text{Mean} = 2.97, \text{SD} = 1.073$) produced no significant differences.

Participants in the stereo condition showed the highest emotional involvement/understanding. On average they reported to have a good understanding of how the characters felt. They felt more immersed and more affected than in the other two conditions.

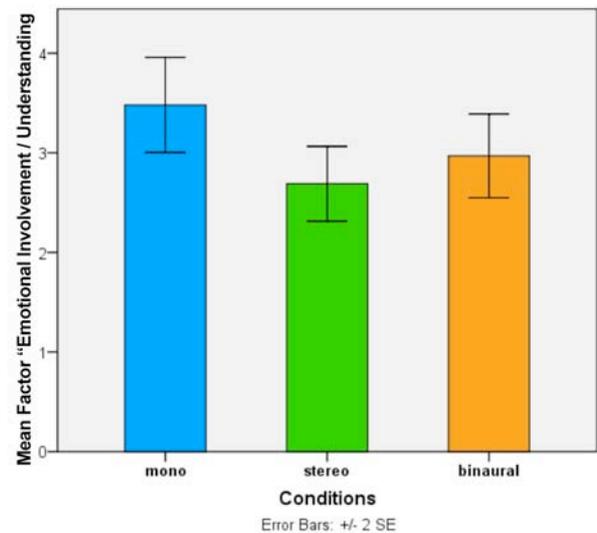


Figure 12: Mean scores for the factor Emotional Alignment/Involvement by condition.

4.3. Focus

The factor *Focus* only comprises two questions, namely:

1. I was focused on the conversation and did not pay attention to the surroundings or the equipment.
2. The test environment did not diminish the listening experience.

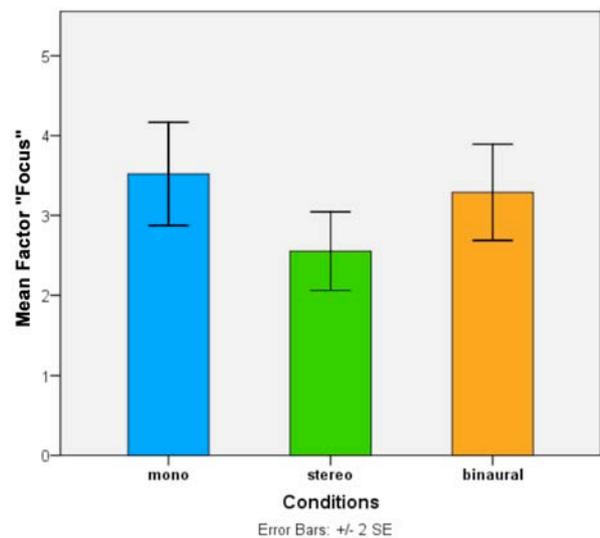


Figure 13: Mean scores for the factor Focus by condition.

An analysis of variance revealed a significant difference ($F(2,77) = 3.163, p = .048$) between the mono ($N = 25, \text{Mean} = 3.52, \text{SD} = 1.617$) and stereo ($N = 29, \text{Mean} = 2.552, \text{SD} = 1.325$) condition confirmed by a post hoc Bonferroni test ($p =$

.05). The binaural condition (N = 26, Mean = 3.289, SD = 1.537) produced no significant differences.

As illustrated by figure 13, generally participants found the test environment did not have a strong impact on their listening experience or distracted them from focusing their attention on the audio play. However, the stereo condition had significantly lower means compared to the mono condition, indicating a positive impact of the stereo condition on the ability to focus on the play.

4.4. Authenticity

Gathered in this factor are the questions:

1. I enjoyed listening to the conversation.
2. The conversation was convincing.
3. The scene felt alive and vivid.

There were no significant differences between the three conditions in respect to how authentic the participants regarded the conversation to be. Although there is trend towards a lower mean value for perceived authenticity in the stereo (N = 29, Mean = 2.598, SD = 1.448) and binaural (N = 26, Mean = 2.69, SD = 1.073) conditions compared to the mono (N = 25, Mean = 3.48, SD = 1.194) condition. In general participants believed the conversation to be quite authentic.

4.5. Emotions

In the post-study questionnaire we asked participants to agree or disagree (on a 7-point Likert scale) to eighteen questions about their emotional state. A PCA suggested a clustering into three constructs:

1. *Negative Emotions* (“I feel: unhappy/shocked/ angry/frustrated/confused/ aggressive.”)
2. *Positive Emotions* (“I feel: relaxed/happy/ satisfied/cosy/chatty.”)
3. *Alertness* (“I feel: active/lively/curious.”)

An analysis of variance showed a significant difference ($F(2,77) = 5.269, p = .007$) between the mono (N = 25, Mean = 5.433, SD = 1.057) and the stereo condition (N = 29, Mean = 4.47, SD = 1.174) confirmed by a post hoc Bonferroni test ($p = .008$).

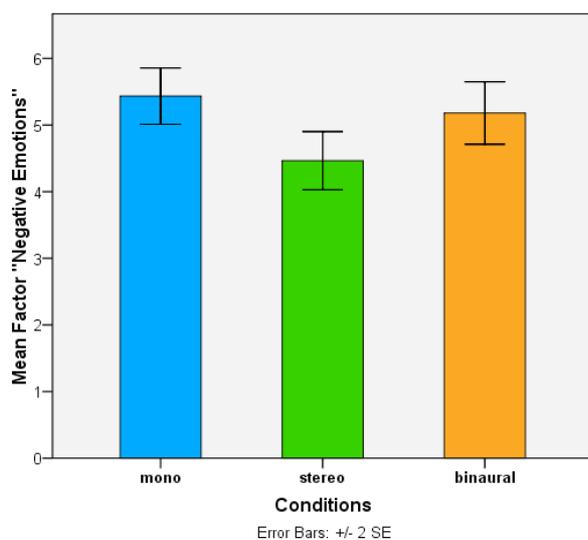


Figure 14: Mean scores for the factor Negative Emotions by condition.

As illustrated in figure 14 participants generally tended to disagree when asked if they felt negative emotions. However, in the stereo condition participants disagreed less strongly when asked about their negative emotions.

There was no significant difference found between the conditions in respect to the factors Positive Emotions and Alertness. Participants generally felt indifferent (means of 3.5 for binaural and stereo and 3.8 for mono) when asked about their positive emotions. Participants tended to feel rather alert with mean values around 3.0 for the stereo and binaural conditions, and around 3.4 for the mono condition.

5. DISCUSSION

Our study was designed to provide insights into the effect of mono, stereo, and binaural sound on the perceived social presence as indicated by the perception of presence in a virtual scene and the understanding of the emotional state of speakers. Our results indicate that, as assumed in H1, there are significant differences between the conditions. Both methods used, the questionnaire and the visualization of the scene by means of drawing, showed a higher perceived presence in the binaural condition compared to the mono condition. Especially the visualization method proved to be a very rich source not only for insights related to how participants perceived themselves in relation to the characters, but also for subtle signs of what they thought was the predominant emotional content. Additionally, the drawings could be used as indicators for localization accuracy. We refrained from doing so as we did not explicitly request participants to draw characters according to the seating plan.

Participants in the stereo condition showed the highest emotional understanding with a mean value of 2.69. On average they reported to have a good understanding of how the characters felt. They felt more immersed and more affected than in the other two conditions. We could only partly accept H2 as we indeed found a significant difference between mono and stereo, but no statistically relevant difference between the binaural and mono conditions. Given Västfjäll’s findings [13] we assumed no difference between stereo and binaural sound in terms of emotional involvement/understanding. Similarly, we found only a significant difference for the factor Negative Emotions between the mono and the stereo condition, not between the mono and the binaural condition. As the play was written to emotionalize and as the emotions displayed in the play were predominately ‘negative’ emotions of anger, shame and fear it is not surprising that we only see an influence of the conditions on negative emotions (as only these had been manipulated).

In contrast, it was unexpected that the binaural condition did not show a significant difference from the mono condition. It seems that the externalization of the sound source has no impact on the perception of other humans’ emotions. There are several possible explanations for why the participants in the binaural condition showed weaker emotional alignments. Firstly, there might have been more participants with lower accuracy in their localization ability (“bad localizers”) in the binaural condition. This seems unlikely, though, as it would have affected the analysis on the factor Presence, which it did not. Secondly, there might have been unknown psychological effects involved. For example, as participants in the binaural condition had a stronger sense of presence, they may have dissociated themselves from the display of negative emotions as a defence mechanism and therefore may have shown lower emotional alignments. Thirdly, as the binaural listening experience through headphones was

new to the participants, they may have been focussing more strongly on the medium than on the message. As we can only speculate about explanations, an experimental examination is necessary.

The stereo condition showed to have a positive impact on the ability to focus on the play. An explanation for this might be that as participants in the stereo condition showed stronger emotions and were more emotionally involved, hence they may have found it easier and more interesting to follow the conversation.

In line with Hendrix & Barfield [11] we found no effect of the conditions on the perceived authenticity of the conversation.

6. Conclusion

In conclusion, using the data presented and some reasonable assumptions, we found strong evidence that stereo provides a significant improvement, at least when there are multiple participants conversing, for enhancing the understanding of the emotional state of speakers. Unlike gaming environments, most electronic communication does not make use of spatial sound but is monophonic only. Adding a second channel could significantly improve the communication experience.

If a strong sense of presence or social presence in a (virtual) scene/play or game (using dialogue with human characters) is desired, our research suggests preferring spatial sound representation to stereo or mono representations.

7. ACKNOWLEDGMENTS

We thank Tim Bell, Andy Cockburn, Mark Billingham, Raphael Grasset, Andreas Dünser, and Julian Looser for early discussion and feedback, Henri Toukoma and Julia Turku for helping with the recordings, Martin Schrader, Monika Polonen, Markus Virta, and Perttu Kivelä for their help running the experiment.

8. REFERENCES

- [1] Biocca, F. and Harms, C. 2002. Defining and measuring social presence: Contribution to the Networked Minds Theory and Measure. In Proceedings of the Fifth Annual Workshop: Presence 2002. Universidade Fernando Pessoa, Porto, Portugal.
- [2] Lombard, M. and Ditton, T. 1997. At the Heart of It All: The Concept of Presence. *JCMC*, 3, 2.
- [3] Heeter, C. 1992. Being there: the subjective experience of presence". *Presence: Teleoperators and Virtual Environments*, 1, 2, 262-271.
- [4] Zhao, S. 2003. Toward a Taxonomy of Copresence. *Presence: Teleoperators and Virtual Environments*, 12, 5, 445-455.
- [5] IJsselsteijn, W.A., de Ridder, H., Freeman, J., and Avons, S.E. 2000. Presence: Concept, determinants and measurement. In Proceedings of the SPIE, 3959, 520-529.
- [6] Lombard, M. and Jones, M.T. 2007. Identifying the (Tele)Presence Literature. *PsychNology Journal*, 5,2, 197-206.
- [7] <http://ispr.info/>
- [8] Schuemie, M.J., van der Straaten, P. Krijn, M., and van der Mast, C.A.P.G. 2001. Research on Presence in Virtual Reality: A Survey. *Cyberpsychology & Behaviour*, 4, 2, 183-201.
- [9] Barfield, W., Zeltzer, D., Sheridan, T.B., and Slater, M. 1995. Presence and performance within virtual environments. In Barfield, W. and Furness, T.A. (Eds), *Virtual environments and advanced interface design*. Oxford: Oxford University Press.
- [10] Loomis, J. 1992. Distal attribution and presence. *Presence: Teleoperators and Virtual Environments*, 1,1, 113-119.
- [11] Hendrix, C. and Barfield, W. 1996. The Sense of Presence within Auditory Virtual Environments. *Presence: Teleoperators and Virtual. Environments*, 5, 3, 290-301.
- [12] Freeman, J. and Lessiter, J. 2001. Here, there and everywhere: the effects of multichannel audio on presence. In Hiipakka, J., Zacharov, N., and Takala, T. (Eds.), *Proceedings of the 7th International Conference on Auditory Display (ICAD)*, Espoo, Finland, 231-234.
- [13] Västfjäll, D. 2003. The Subjective Sense of Presence, Emotion Recognition, and Experienced Emotions in Auditory Virtual Environments. *Cyberpsychology & Behaviour*, 6, 2, 181-188.
- [14] Aaltonen, V., Takatalo, J., Häkkinen, J. Lehtonen, M., Nyman, G., and Schrader, M. 2009. Measuring mediated communication experience, *International Workshop on Quality of Multimedia Experience (QoMEX'09)*, San Diego, CA, USA, 104-109.
- [15] ITU-R BS.1116-1, 1997. Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems." *International Telecommunication Union*, Geneva, Switzerland.
- [16] <http://www.itu.int/rec/T-REC-P.57/en>
- [17] Kylliäinen, M., Helimäki, H., Zacharov, N., and Cozens, J. 2003. Compact high performance listening spaces. In *Proceedings of Euronoise*, Naples, Italy.
- [18] DiStefano, C., Zhu, M., and Míndrilă, D. 2009. Understanding and Using Factor Scores: Considerations for the Applied Researcher. In *Practical Assessment, Research & Evaluation*, 14, 20.