

The Power of Sound Design in a Moving Picture: an Empirical Study with *emoTouch* for iPad

MAXIMILIAN KOCK[1]

Ostbayerische Technische Hochschule Amberg-Weiden, Germany

CHRISTOPH LOUVEN

Universität Osnabrück, Germany

ABSTRACT: The art of sound design for a moving picture rests basically on the work experience of pragmatists. This study tries to establish some guidelines on sound design: In an experiment 240 participants gave feedback about their emotions while watching two videos, each combined with four different audio tracks – music, sound effects, full sound design (music and sound effects) and no audio (as the comparative "null" version). Each participant viewed an audiovisual combination once to prevent habituation. The lead author employed a tablet-computer with the *emoTouch*-application serving as a mapping tool to provide information about the emotional responses. The participants moved a marker on the tablet's touch screen in a two-dimensional rating scale describing their felt immersion and suspense. A 3-factor-ANOVA showed significant increases of the median (and maximum) values of immersion and suspense when the participants listened to music and/ or sound effects. These values were always compared to the induced emotions of the participants who watched the videos with no audio at all. The video with full sound design audio tracks increased the median immersion values up to four times and the median suspense values up to 1.4 times. The median suspense values of the video with either music or sound effects dropped by 40 percent compared to the median suspense values of the null version. In contrast, the median immersion values were increased up to 3.6 times. The findings point to the importance of sound effects in an appropriate mix with music to enhance the viewers induced immersion and suspense.

Submitted 2018 September 20; accepted 2018 October 17.

KEYWORDS: *influence of sound design on perception, moving picture, perceived emotion*

INTRODUCTION

SINCE the start of talking movies in the 1920's, the audio track and its influence on film perception has been a strong topic in research activity. However, in the past one hundred years the impact of music on the viewer's perception has mostly been at the center of interest. The possible power of sound design (i.e. a balanced mix of music and sound effects) has only been investigated in a few publications in scientific literature (Brauch, 2012; Cohen, 2006, 2010, 2015; Flückiger, 2007; Hoekstra, 2012). Additionally, filmmakers and sound designers have summarized their practical implicit experiences about the viewer's emotions and reactions, which can be evoked by sound (Chion, 2012; Görne, 2017; Holman, 2010; Lynch, 1998; Raffaseder, 2010; Steinitz, 2015; Sonnenschein, 2001). Until now, neither music psychology, film psychology nor neuroscience has profoundly explored the original power of sound design in scientific research.

In principle, audio induced emotions highly depend on the film plot and film genre (film type): For example, in most cases, viewers of animated movies might need additional sound information compared to others watching a live-action movie, because silent animated movies are not always self-explanatory per se (Goldmark, 2013). Animated films often consist of computer-generated pictures, whereas the dialogue and sound tracks are analogously recorded afterwards with actors and sound designers. In contrast, live-action movies are staged with actors, shot on a movie set with motion-picture cameras and often with simultaneous dialogue and sound recordings. Viewers of animated films unconditionally might need music and sound



effects to get more background information in order to understand the story line, whereas live-action movies might do their job merely on the visual level (like silent movies before the 1920's). Thus, different film types could cause different results, depending on their film layout.

Former studies about the emotional impact of audiovisual combinations on viewers (e.g. Cohen, 2006, 2010, 2015; Hoekstra, 2012) were designed in such a way that participants watched the same video repeatedly and successively with different audio versions. After the test or after each single test run, each participant filled out questionnaires about their felt emotions. This type of test procedure nevertheless brings at least two important aspects into focus: 1. The participant's emotional reaction is retrieved **after** they have seen the film or video. Therefore, the amount of the noted emotional intensity might have conceivably changed compared to a instantaneously measured affect. 2. Participants might change their felt emotion when they are viewing the same video a second, third or fourth time with different audio tracks. It could be inapt to compare the reaction data of test participant A watching a video the first time with the reaction data of test participant B watching the identical video a third time. The possible differences in the resulting data might be caused by the potential habituation of participant B to the film plot. In consequence, a test procedure had to be developed, which relied only on one 'untouched' test run for each participant with immediate measuring of his or her emotional reaction. This new approach led to the key questions of this study (see paragraph 'Experimental Questions') and thus gave us insights into the use of music and sound effects might influence the viewer's perceived emotions.

This study implemented therefore an experimental design, which corresponded to the participant's everyday situation: watching a professionally produced television event or movie on a tablet screen with headphones in a relaxing surrounding. Bullerjahn stated that this experimental design could result in more sensible outcomes (Bullerjahn, 2008) than other test designs. Tracking of eye-movements (Auer et al., 2012), measurement of heart rates/ skin conductance or fMRI-methods (functional magnetic resonance imaging) imaging the participant's brain activity might generate less valid results caused by potentially exerting stress to the participants during their testing procedure (Donnelly, 2015; Eklund, Nichols, Knutsson, 2016). In addition, compared to other research tools like *EMuJoy* (Nagel et al., 2007), the *emoTouch*-application (Scholle & Louven, 2013, 2015) does not need technical interfaces like a mouse or joystick and allows the participant's nearly simultaneous and direct reaction to be measured. Therefore, the *emoTouch*-application was employed as a mapping tool in this study (see the next 'Method' section).

For this exploratory study, it was decided by the lead author to choose two key audio-visual affects (i.e. suspense and immersion) to assess the emotional reaction of the participants while they were watching the test videos. According to the work experience of pragmatists in audio production (Brauch 2012; Chion, 2012; Flückiger, 2007; Görne, 2017; Sonnenschein, 2001; Steinitz, 2015), the viewer's aural envelopment is mostly based upon these two key 'media' affects. Future researchers can discuss how far these two affects might be interconnected: The perceived suspense, initiated by the film plot, can be the precondition to feel immersion; If there is no suspense there might be no immersion of the participant. Admittedly, there is a mutual interaction between these two affects: Without efficient immersion there would be no perceived suspense. Cohen used in her own studies the terms "absorption" or "involvement" (Cohen, 2010). The latter term is very close in meaning to 'perceived immersion'. Nevertheless 'immersion' comprises both, the technical-acoustic envelopment (e.g. by surround sound or headphones) and the emotional involvement (e.g. by the film plot). The other dependent variable 'suspense' is similar to the term "arousal" (Russell, 1980; see also the next paragraph). 'Perceived suspense' nevertheless is more a 'media' affect, presumably representing together with 'perceived immersion', the potential impact of the independent variables music, sound effects and film type. It might be considered that a film plot combined with sound design could initiate a very high immersion, which may cause participants to respond inaccurately on a mapping tool. However, in this study each participant experienced the very same experimental design and, aside from that, was far less stressed than the participants in previous studies using fMRI or skin conductance measuring methods.

The lead author disregarded the audio type 'speech' in this study, because this audio element is special in its functional purpose: Speech has to be understood. The immersive impact of speech is on principle lower compared to music and sound effects. Speech is typically the loudest element in an audio mix and pushes the two others into the aural background. The main purpose of speech is mostly to pass information, less to be immersive.

METHOD

A Mapping Tool to Provide Information about Emotional Responses

In 2013, the *emoTouch*-application (Scholle & Louven, 2013, 2015, 2017), a free scientific research tool operating with Apple's iOS on an iPad-tablet, was published by the department of Systemic Musicology at Osnabrück University. It allows participants to continuously mark their reaction in a two-dimensional x-/y-rating scale to media content (music or video) played by the device. Originally, the *emoTouch*-application was designed as a tool for research on musical emotions, based on the two-dimensional emotion space model with its valence (negative - positive) and arousal (active - passive) dimensions (see Russell, 1980).

The *emoTouch*-application was the ideal mapping tool for this research study, because it provided videos with an additional scaling background and thus enabled a spontaneous user feedback: The participants were asked to mark their felt emotions on a tablet's touchscreen while watching a video on the *same* screen and listening simultaneously to the audio variables (music and sound effects) via headphones (see Figure 1). After each test run, the resulting x-/y-data stream could be easily evaluated with statistical software.



Fig. 1. A participant during a test run

Stimuli and Procedure

Due to the presumed difference in the perception of the sound design in animated and live-action movies (Goldmark, 2013), two films were produced: a computer-generated animated film and a live-action movie, staged with actors, shot with a camera. It is obvious that these videos are only two samples in audiovisual design and do not represent the vast design options in movie production. Therefore, the results of this study can only be related to these two stimuli. Further research is necessary to estimate the extent to which these results could apply to other videos (film types).

Both test videos were presented to the participants in combination with one of four different audio types: without audio (silent), with music, with sound effects, and with complete sound design (music and sound effects).

Only one video with a specific combination of film type and audio track was presented to each participant in order to eliminate any learning or habituation effects. In a pretest, the participants could practice handling the device and the application with the first (animated or live-action) video. The main experiment, which immediately followed, generated the test data using the second video. For example, if a participant had to rate the animated film in the main test run the live-action film was used for training and vice versa.

The plot of both films reflected the inherent story of the music that was chosen for the soundtrack. This music was selected from two pieces of Modest Mussorgsky's 'Pictures at an Exhibition' in the original piano version: 'Samuel Goldenberg and Schmuyle' (Length 1'47", used with the animated film, see Appendix

and Figure 9) and 'The Catacombs' (Length 1'12", used with the live-action movie, see Appendix and Figure 10). These musical pieces turned out to be ideal for the purpose of this study. First, the visual plot of the films could be easily derived as Mussorgsky's compositions tell a story of their own. Second, the piano music provided enough auditory space to optional additional sound effects because of the special acoustic envelope structure of a piano tone.

Two students studying media production and media technology at Technical University Amberg-Weiden (OTH) produced both films as part of a student research project. Previously, the two pieces for piano were performed by the lead author and were recorded in the university's music studio. The lead author later prepared the sound design for this study using sound libraries and new specially designed sound effects (see Appendix).

While watching the videos on the tablet's screen in an office space of the university and listening to the soundtrack via headphones individually (always adjusted with the same sound pressure level), participants were asked to continuously mark their current emotional state on the tablet's touchscreen in two dimensions. The application's coordinate system was aligned by the author: in the x-scale from -1 to +1 describing the amount of the perceived immersion; in the y-scale from -1 to +1 describing the amount of perceived suspense. Both scales were without units and represented only a relative measurement (see Figures 2 and 9). Thus, the study design consisted of three independent, dichotomous variables: music (yes/no), sound effects (yes/no), and film version (animated/live action). The continuously measured dependent variables were 'perceived immersion' and 'perceived suspense' (Figure 2). The user study yielded a continuous x-/y-data stream in steps of seconds. Based upon this data stream, the means, the maxima, the medians, the bandwidth and the evaluating graphics were generated with the free statistical computing software 'R'.

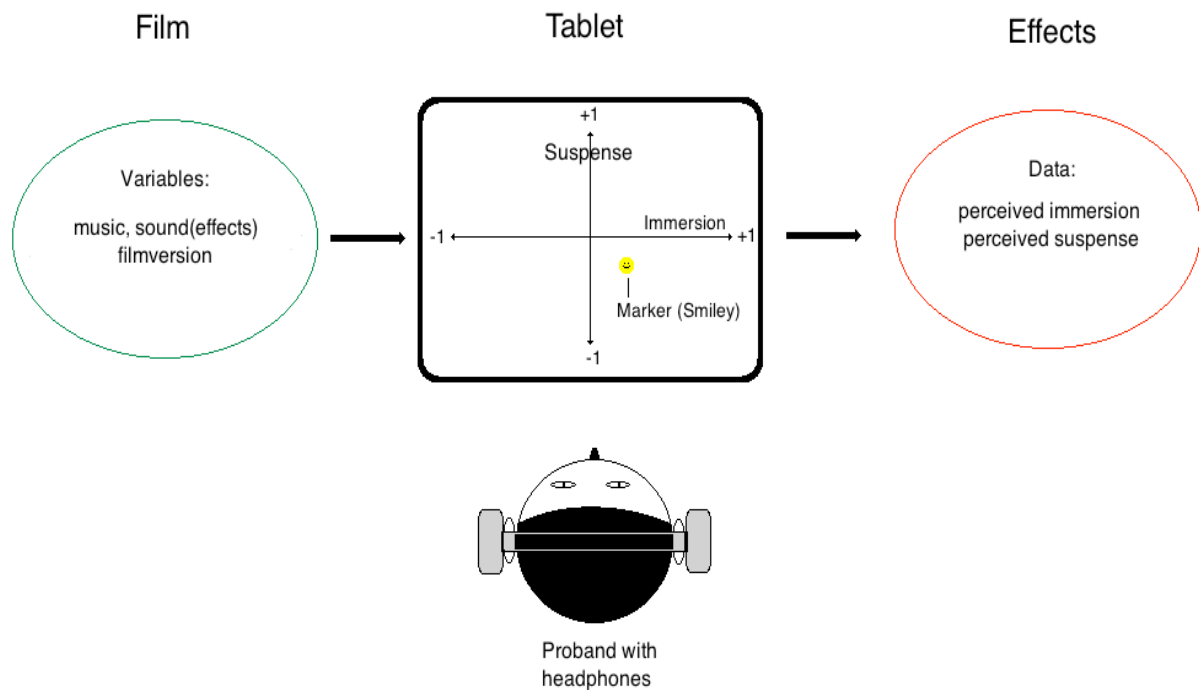


Fig. 2. Scheme of the test procedure

The participants answered immediately after the experiment a short questionnaire comprising eight questions about their usage of recorded music and motion picture, about their musical experience (e.g. playing an instrument) and about their opinions and experience of the impact of audio on audio-visual perception (see Appendix). With these questions, the lead author aimed to acquire an insight into each participant's aural history. These insights were thought to potentially interact with the recorded two-dimensional emotional data of the experiments.

Experimental Questions

How much did the stimuli with sound design increase (or decrease) the participant's perceived immersion and suspense compared to the silent stimulus? To what extent was this change due to the music and/ or the sound effects? Did a certain combination of a film type (animated or live-action-movie) and audio type have a special impact on the viewers?

Participants

In total, 240 participants participated in the experiment (30 cases per condition x 2 videos x 4 audio conditions). The average age was 25.74 years ($SD = 7.58$). 82.5% of the participants were students enrolled in the study course Media Production and Media Technology at Technical University Amberg-Weiden (OTH). Although this implied a high affinity to audio-visual media as a producer and a consumer, only very few test participants were acquainted with the music of Mussorgsky beforehand. The examiner randomly inquired about this after the participant's test run. One third (33.3 %) of the test participants were female. There was no compensation for participation.

RESULTS

The analysis of the data with a 3-factor-ANOVA (factors: film version, music, and sound effects) led to the following results (see Tables 1 and 2).

Table 1. Results of the significance test 'immersion' (***) = $p < .001$, ** = $p < .01$, * = $p < .05$, . = $p < .1$, - = $p > .1$)

<u>Immersion</u>	Medians (p)	Maxima (p)	Bandwidth (p) (Δ Max- Min)	Highest Median in
Music	*	.	-	Animated movie
Sound effects (FX)	**	***	.	Live-action- movie
Film version	-	**	**	
Interaction	No interaction	No interaction	Film version: FX	

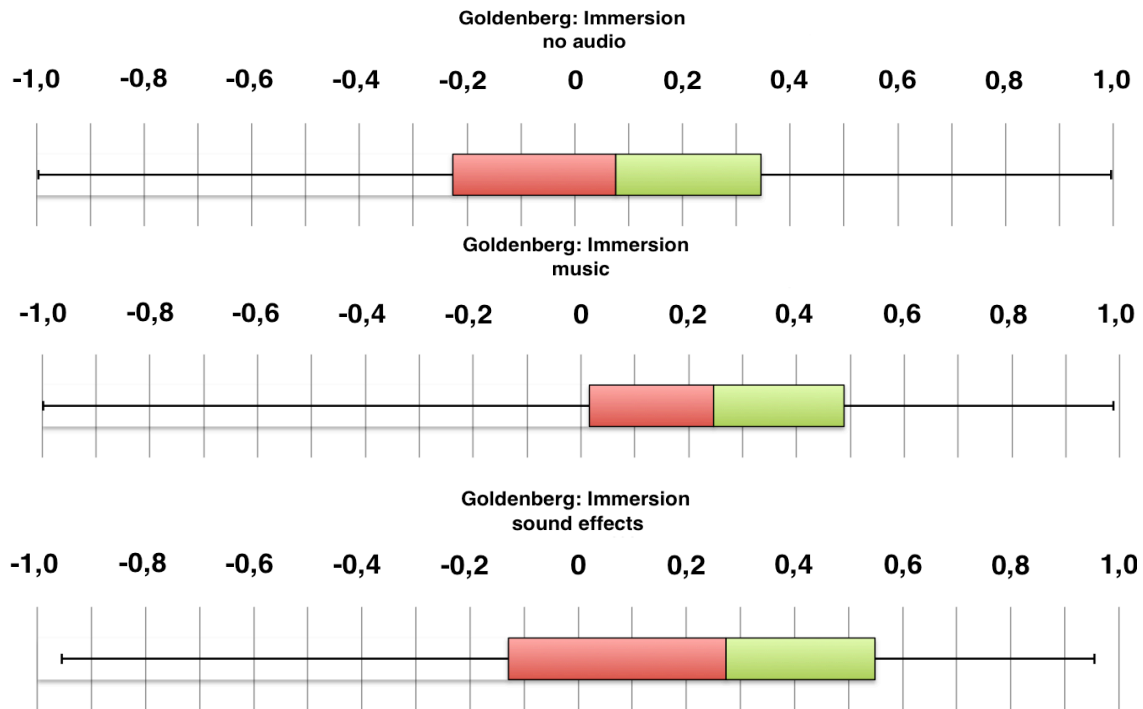
Median values of the perceived immersion were significantly influenced by two factors, the music, and the sound effects. No significant effect was found for the film type.

Table 2. Results of the significance test ‘suspense’ (*** = $p < .001$, ** = $p < .01$, * = $p < .05$, . = $p < .1$, - = $p > .1$)

Suspense	Medians (p)	Maxima (p)	Bandwidth (p) (Δ Max- Min)	Highest Maxima in
Music	-	.	.	Live-action- movie
Sound effects (FX)	-	*	-	Animated movie
Film version	***	-	***	
Interaction	No interaction	No interaction	Film version: FX	

Median values of perceived suspense were significantly influenced by the film type. A significant difference for the maxima was found in the sound effects.

Figures 3 and 4 provide an overview of the findings.



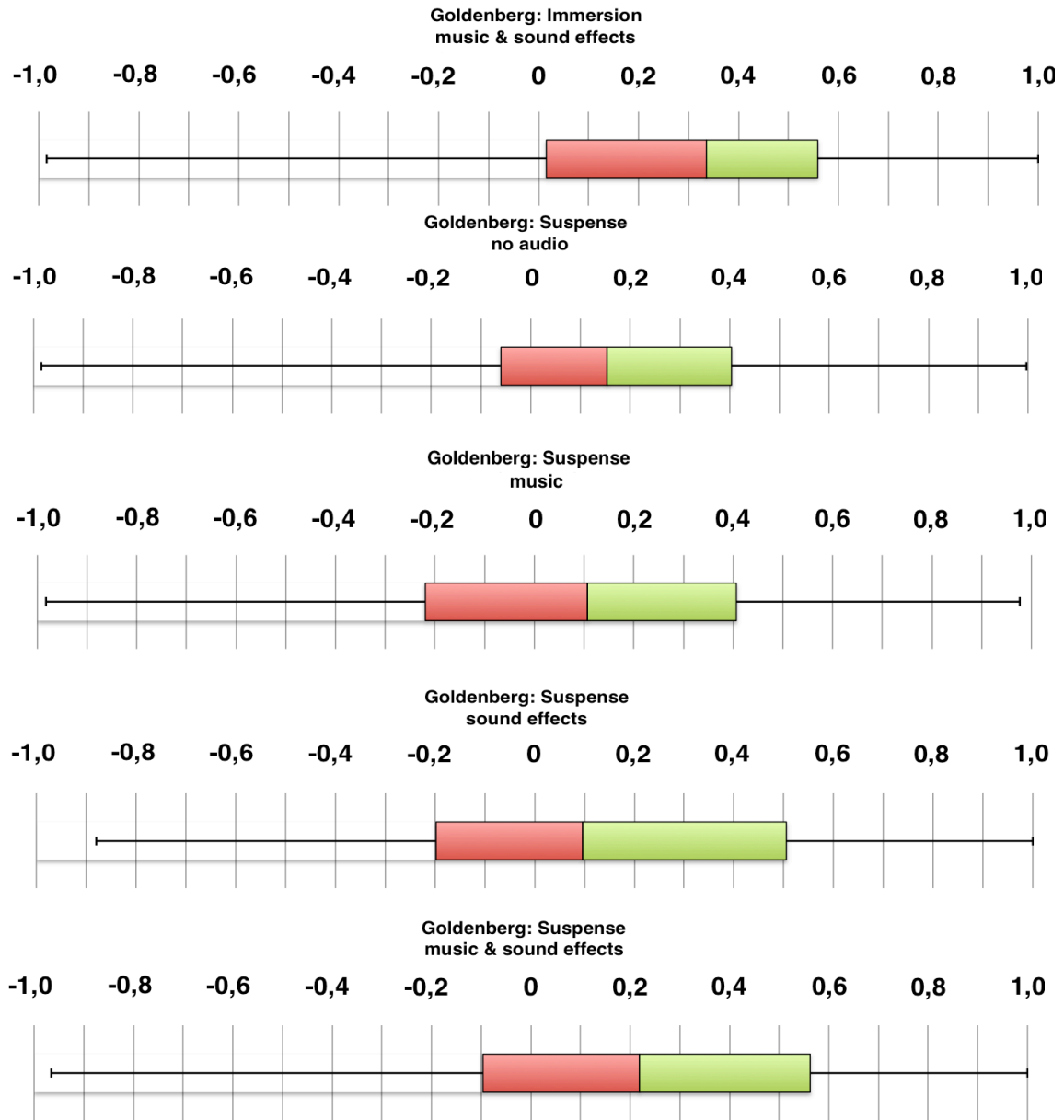
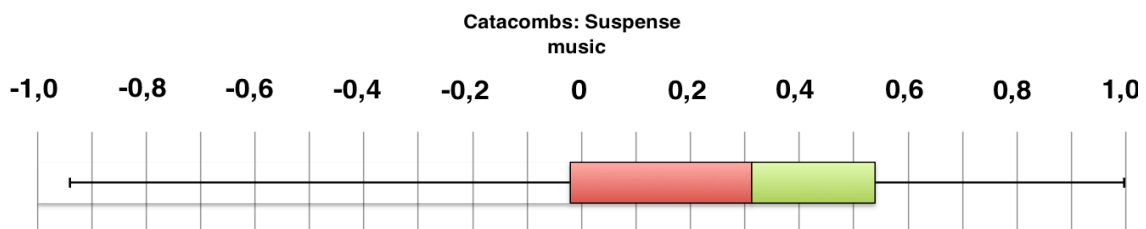
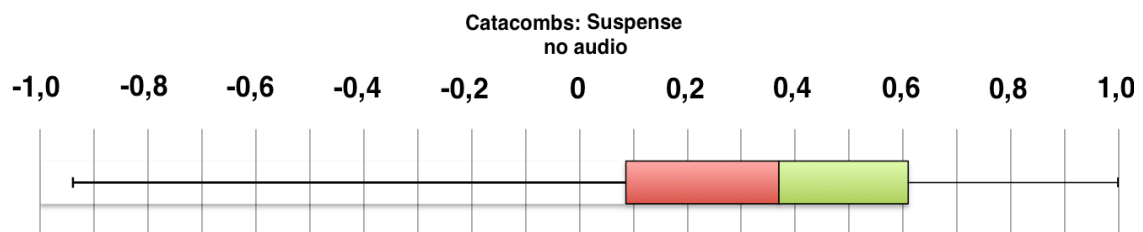
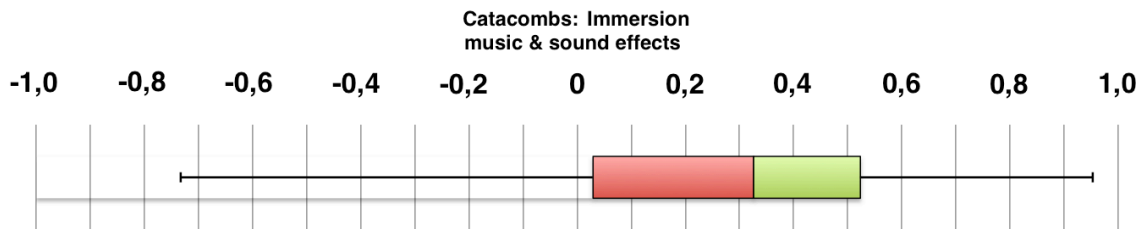
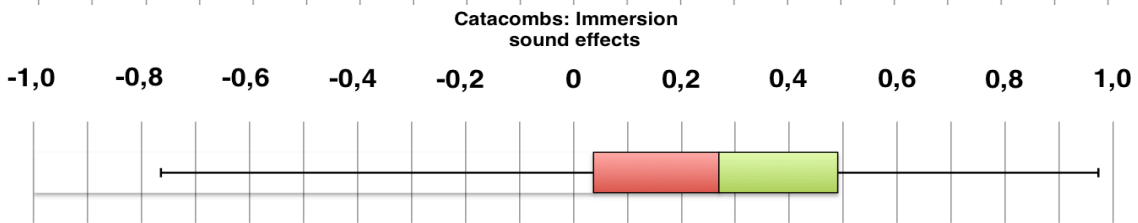
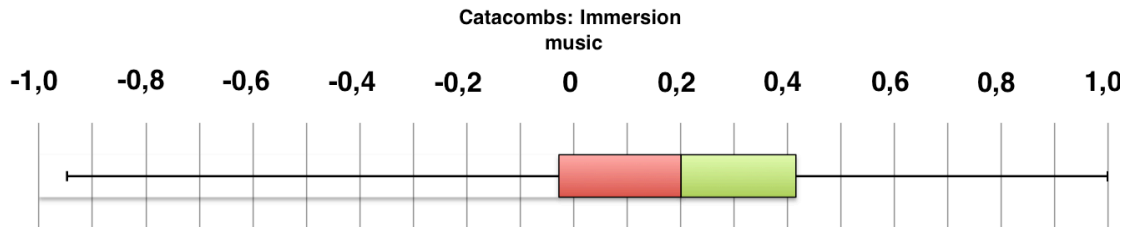
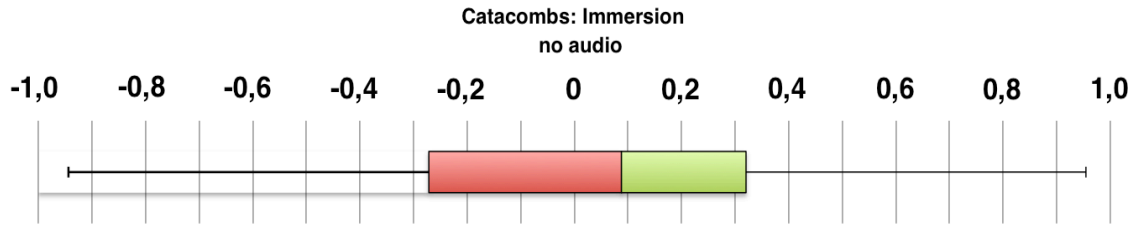


Fig. 3. Boxplots of the participant ratings for the animated movie ‘Goldenberg’. The vertical black line between the green and red boxes indicates the median. The left side of the red box depicts the 25%-quartile, the right side of the green box depicts the 75%-quartile. The whiskers left and right represent the minimum or the maximum of the participant ratings.



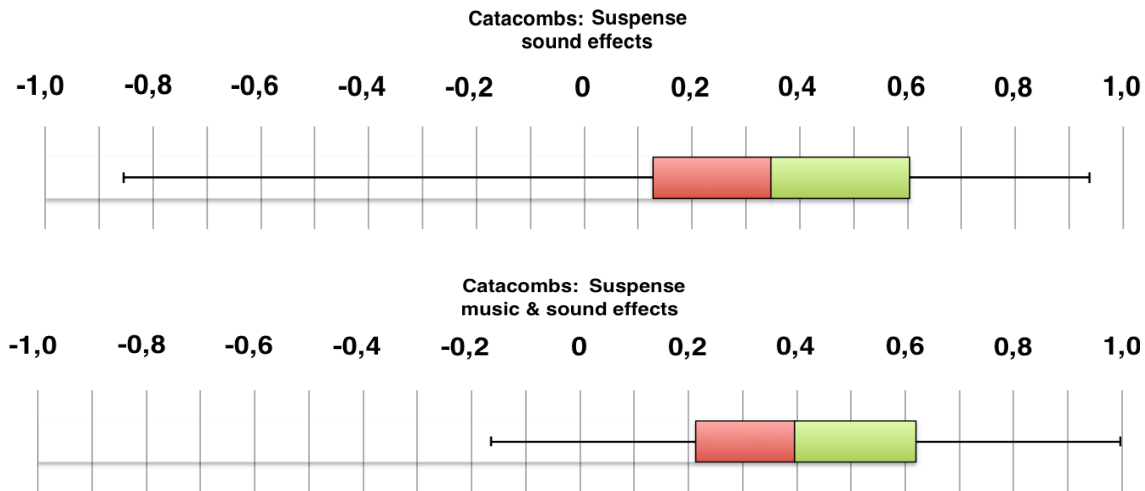


Fig. 4. Boxplots of the participants' ratings for the live action movie 'Catacombs'. The vertical black line between the green and red boxes indicates the median. The left side of the red box depicts the 25%-quartile, the right side of the green box depicts the 75%-quartile. The whiskers left and right represent the minimum or the maximum of the participant ratings.

Compared to the median values of the 30 null version cases (with no sound), the median value of the perceived immersion increased more than threefold when the participants were watching both film types just with sound effects, whereas the median value of perceived suspense dropped by 40% showing the animated film and by 10% when showing the live-action video (see Figures 3 and 4).

The median value of the perceived immersion rose threefold for the animated film combined just with music and more than twofold for the live-action movie and music only. Suspense values were lower (see Figures 3 and 4), but these changes were not significant.

Music and sound effects in an audio mix increased the median value of the perceived immersion 3.7 times after showing the live-action movie and 4.4-times after displaying the animated film. Suspense values were higher (see Figures 3 and 4), but these changes were only significant regarding the maxima of the sound effects.

An audiovisual combination with sound effects alone had a stronger impact on the participants' perceived immersion compared to an audiovisual combination just with music (no matter which film type was shown). An audiovisual combination with sound effects and music had an even stronger impact on the participant's perceived immersion and suspense compared to an audiovisual combination just with music (no matter which film type was shown).

Every variable of the post-test questionnaire was checked with the statistical computing software 'R' whether a variable interacted with another. Eventually, there were no findings of significant interactions between the participants' answers in the post-test-questionnaire (i.e. his or her personal aural history) and the values of perceived immersion and suspense.

Figures 5 to 8 on the following pages depict on the y-axis the mean immersion/ suspense values of both film types, each combined with the four different audio conditions. As mentioned before, both are relative units. The scale is defined from -1 to +1. The following four graphics sometimes show extreme high or low levels of perceived immersion or suspense at the beginning and end of the video timeline (x-axis). These swings are caused by the actual low number of active participants in the first or last two seconds of the videos. These extreme values were omitted in the final discussion of the test results.

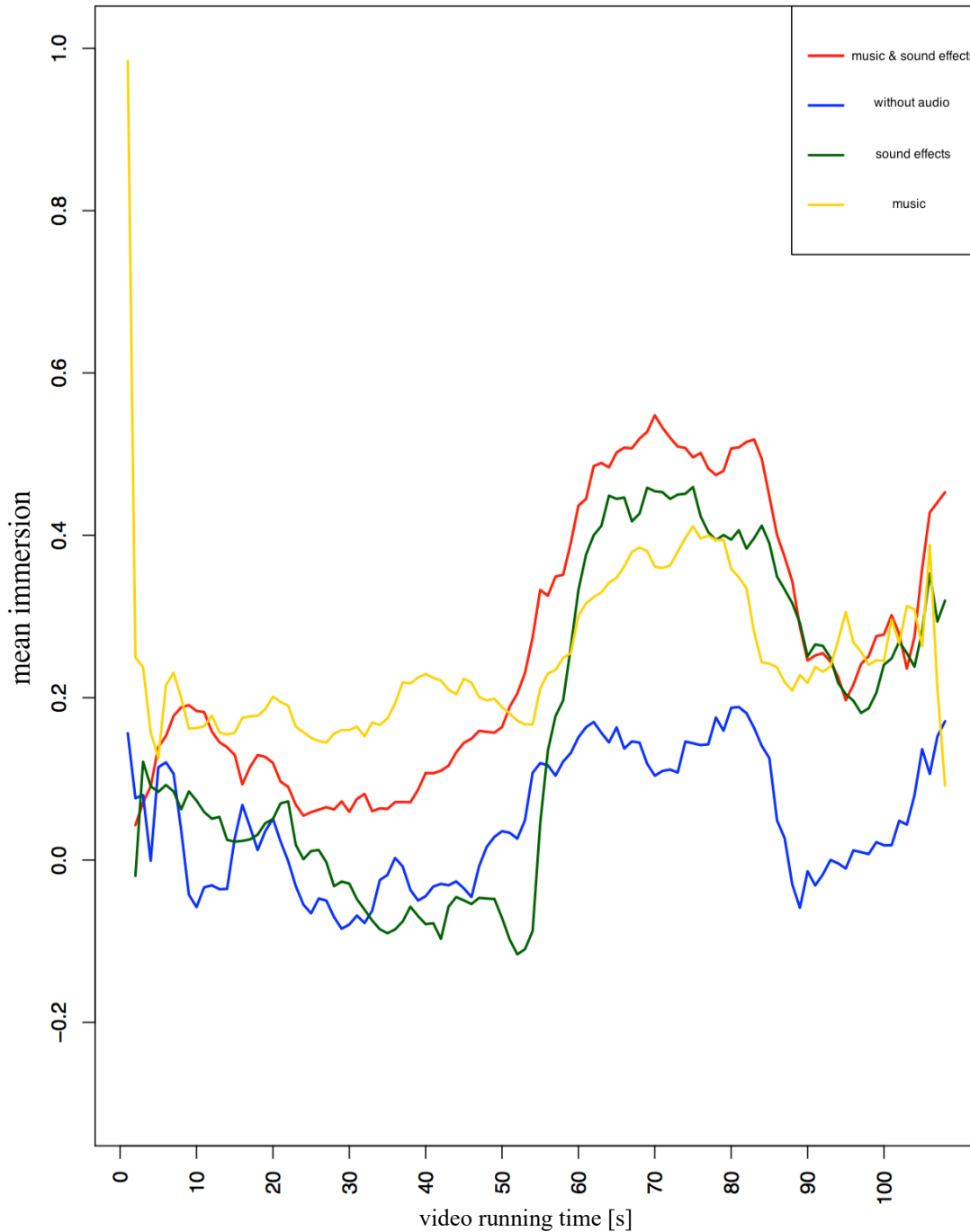


Fig. 5. Mean (average) immersion animated film ‘Goldenberg’

The mean immersion values of the animated film (Figure 5) indicate a higher average immersion when the participants listened to the audio versions, especially when the film plot is displaying visual action (i.e. at 55s running time). The blue line representing the silent version shows lower values. The red (music and sound effects) and the green (sound effects) lines indicate much more dynamic immersion ratings than the yellow (music) and the blue (without audio) lines. In general, music (yellow line) seems to balance the induced immersion, with a dynamic range smaller than the other conditions.

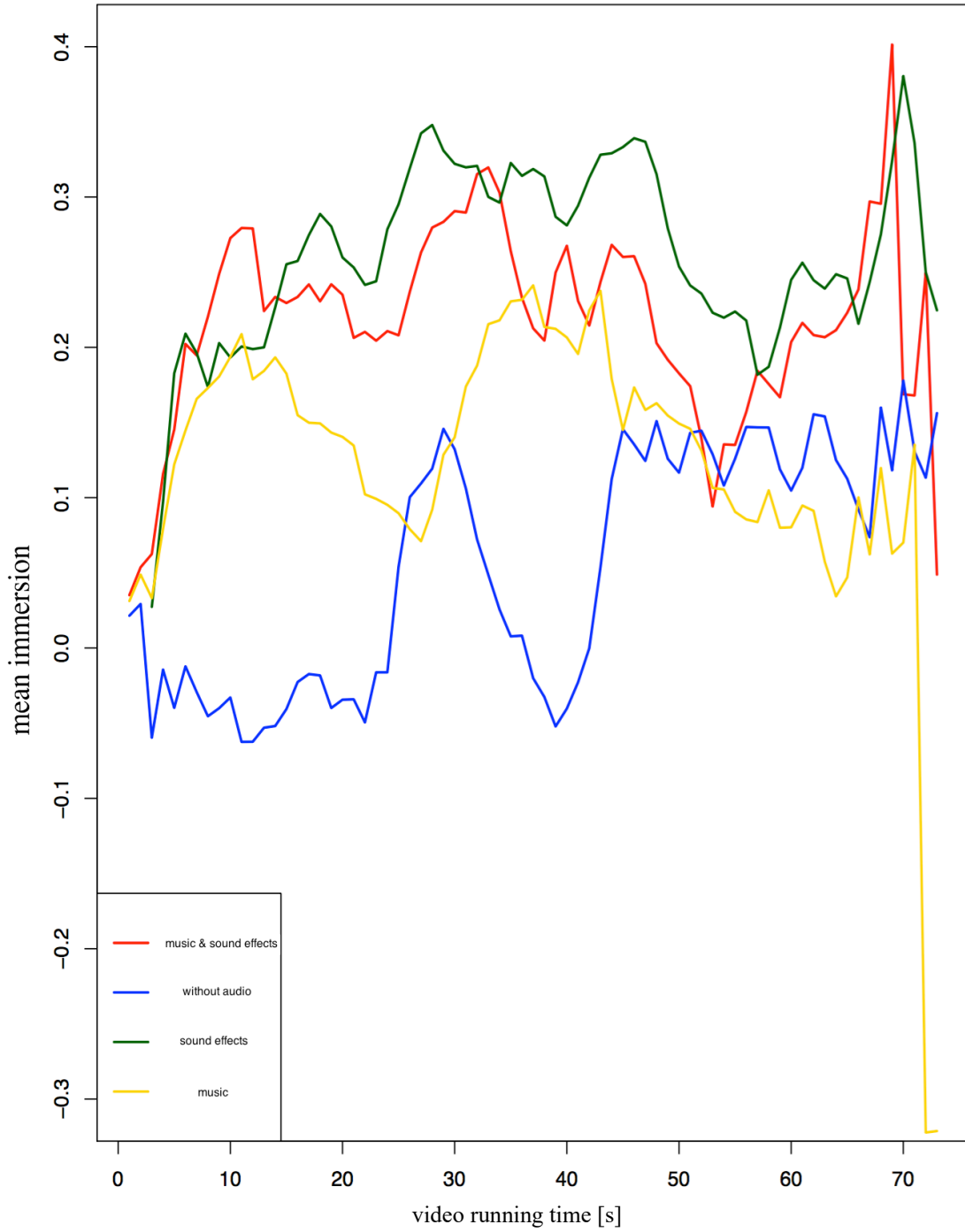


Fig. 6. Mean (average) immersion, live action movie ‘Catacombs’

The mean immersion values of the live action movie (Figure 6) also indicate a higher immersion in the audio versions compared to the silent version. The time curves of all four lines in Figure 6 are obviously less in sync compared to the curves in Figure 5 (animated film). The versions without audio (blue line) and with music (yellow line) produce the lowest immersion.

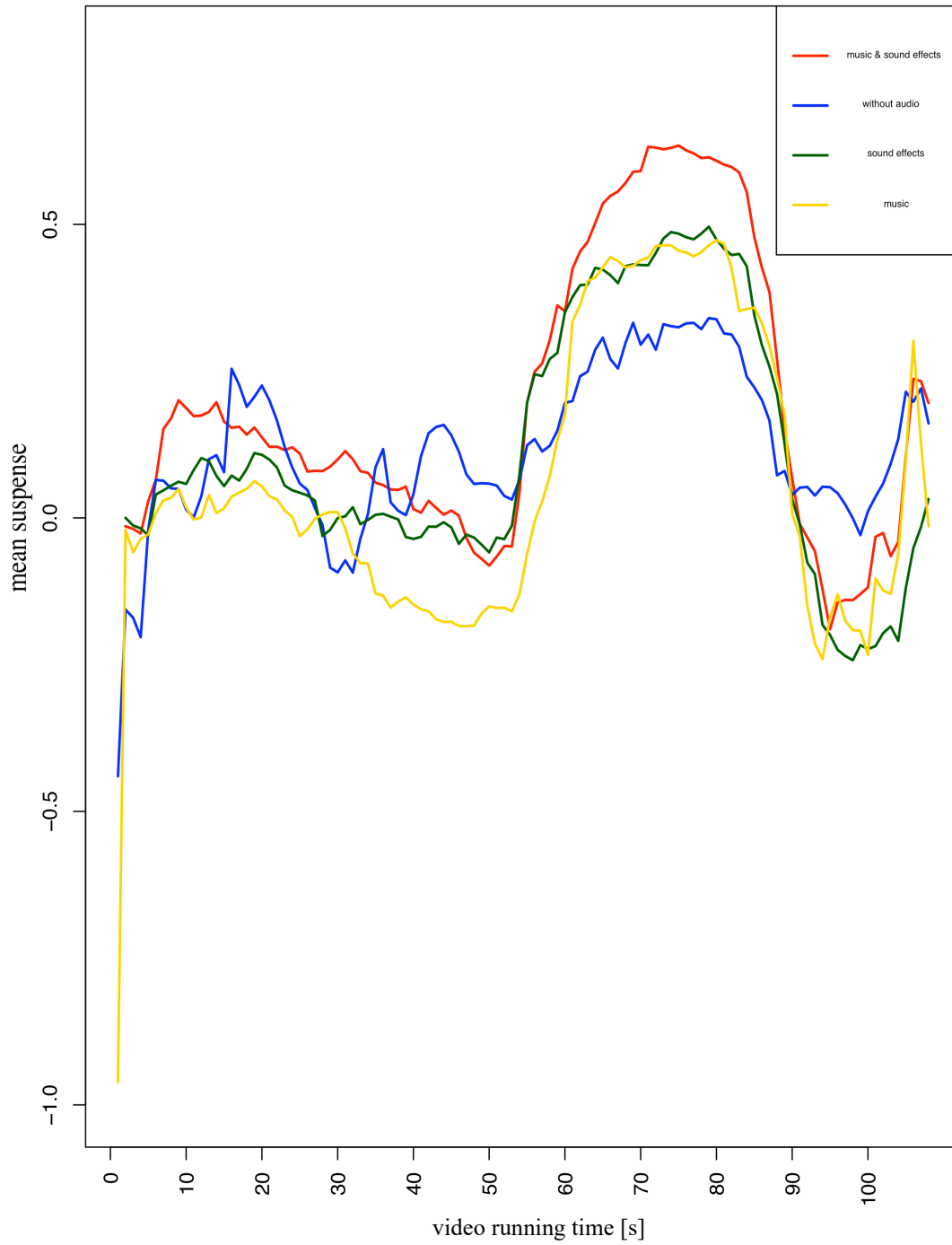


Fig. 7. Mean (average) suspense, animatic ‘Goldenberg’

Figure 7 represents the mean suspense values of the animated film ‘Goldenberg’. It also depicts a similar perceived suspense in the audio versions compared to the silent version. The time curves of all four lines in Figure 7 (animated film) are more in sync compared to the curves in Figure 8 (live action movie). The video with music (yellow line) produces, on average, the lowest suspense values during the video.

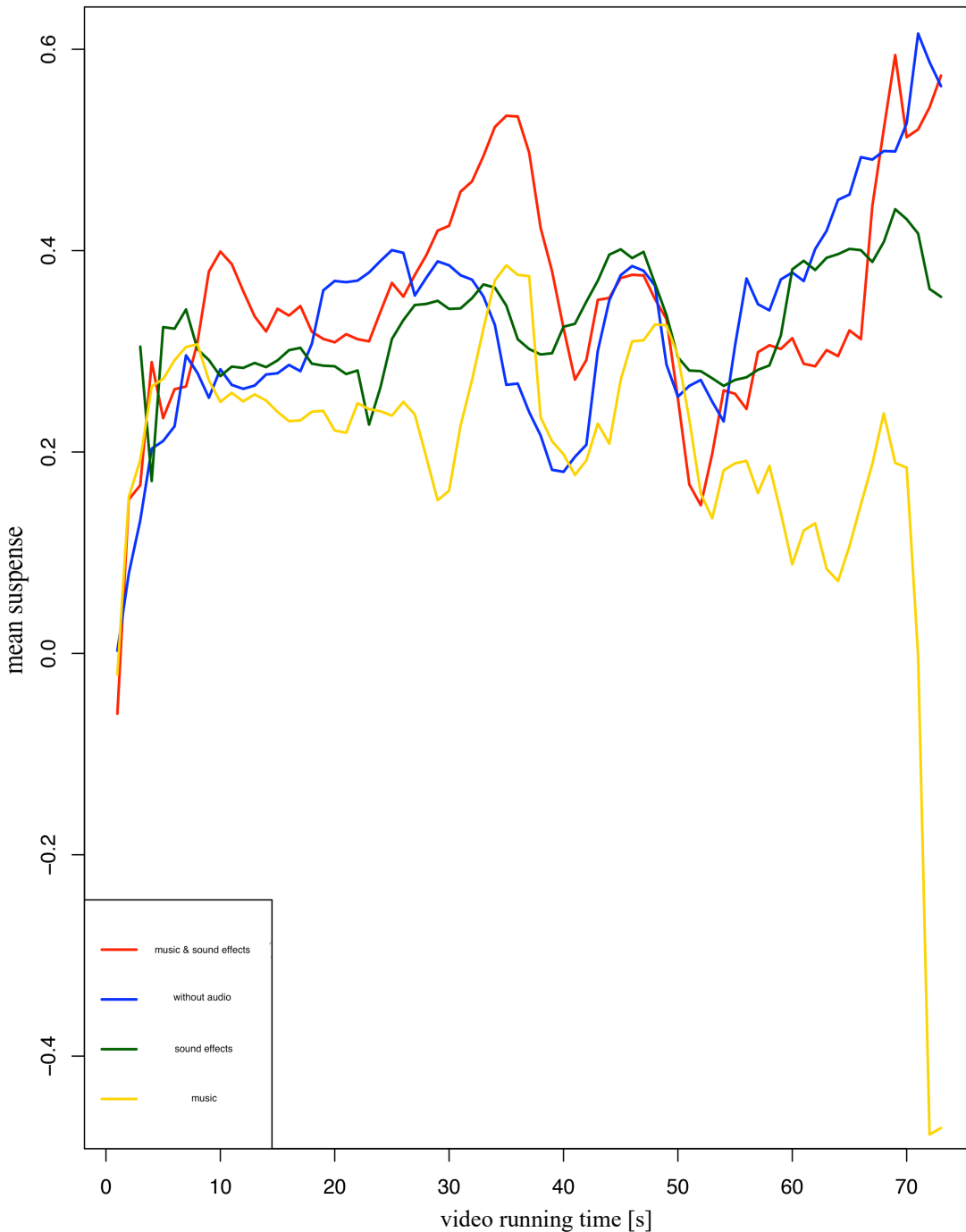


Fig. 8. Mean (average) suspense, live action movie ‘Catacombs’

The mean suspense values of the live action movie ‘Catacombs’ (Figure 8) illustrate that the yellow line (music) has the lowest level of perceived suspense on average, whereas the blue line (no audio) depicts nearly always higher suspense values. Sometimes these values are even higher than in the version with sound effects and the full audio version with sound design. This is noticeable especially towards the end of the video because there is a ‘cliffhanger’ in the plot (see Appendix).

DISCUSSION

Nearly all participants perceived enhanced levels of immersion when they listened to an audio track. The extent of this enhancement nevertheless depended on the visual information and the type of audio. A mere music track tended to flatten the felt dynamic range of immersion. Sound effects alone and in a mix with music strengthened the effect.

In contrast, the 'perceived suspense' correlated more with the film plot and the visual information. Sound effects alone did not enhance the perceived suspense in general. Only the participants who listened to an audio mix of sound effects and music experienced more suspense. This deduces the following conclusions concerning the practice of sound design.

Apparently, the immersive power of music alone and its induced suspense, triggered by audio, seems to be overestimated (see also 'Stimuli and Procedure'). These are the principal findings displayed in Figure 6 (immersion) and Figure 8 (suspense) regarding the live-action movie 'Catacombs'. An explanation could be that the participants might have perceived incongruence between their visual and musical perception channels (see Cohen's Congruence-Association-Model CAM, Cohen et al., 2006). This incongruent sensation led perhaps to a lower emotional effect (Boltz, 2004).

Therefore, the impact of sound effects should be considered more by sound designers. In conclusion, it probably was the well-balanced and congruent audio mix of music and sound effects, which led to higher values of the participants' immersion and suspense in this study.

The sound design of a movie should apparently follow more closely the concept of a (visually independent) radio play (Flückiger, 2007). This, for example, could be realized by non-diegetic sound effects like sound layers of reverb and drones in the video 'Catacombs', or fighting noises caused by actions, which were not seen on screen in the video 'Goldenberg'. In doing so, there was a higher chance of enveloping the participant and inducing him or her to feel the emotion (originated by the film plot) more intensely, because he or she got more immersed through the sound design. Due to the immersive power of the elaborated sound design, the participants could perhaps more easily match their perceived emotions to their own "neural representations" (Levitin, 2007) or "blocks of cognition" (Rumelhart, 1980). This induced the participants to reactivate more vividly the neural pictures stored in their brain. Therefore, by watching an audiovisual movie, the participants felt the emotional impact in two ways: through the continuous actual perception, and also through the reactivation of their stored memories. Levitin asserted that there are two cerebral processes taking place in our brain simultaneously when we are watching audiovisual media: the analysis of audiovisual events and the concurrent and immediate comparison with our neural representations (Levitin, 2007). The findings of this study confirm Levitin's assertion.

The results of this study raise another topic: To what extent does a participant's aural pre-experience change or manipulate his or her own current audiovisual perception? Although the post-test questionnaire did not show any significant effects, film experts presumably react in a worldlier-wise and more experienced manner to the different audio versions than film novices (Filimowicz, 2012; Donnelly, 2015). The vast majority of the participants in this study (82,5% were students of media production and media technology) already had considerable media expertise. Hence, the majority of the 240 participants formed a rather homogenous test group. This could be a possible explanation for the missing significant interaction between their answers in the post-test questionnaire and the measured perception. Furthermore, the questionnaire comprising only eight dichotomous points (yes/no) might not have been detailed enough.

ACKNOWLEDGEMENTS

This article was copyedited by Tanshuree Agrawal and layout edited by Diana Kayser.

NOTES

[1] Correspondence can be addressed to: Maximilian Kock, Ostbayerische Technische Hochschule Amberg-Weiden (OTH), Kaiser-Wilhelm-Ring 23, D-92224 Amberg, Germany, m.kock@oth-aw.de

REFERENCES

- Auer, K., Vitouch, O., Koreimann, S., Pesjak, G., Leitner, G., Hitz, M. (2012). When Music drives Vision: Influences of Film on Viewers' Eye Movements. *Proceedings of the 12th International Conference on Music Perception and Cognition*, 73-76. Retrieved from: <http://icmpc-escom2012.web.auth.gr/proceedings.html>
- Boltz, M. G. (2004). The cognitive processing of film and musical soundtracks. *Memory & Cognition*, 32 (7), 1194-1205. <https://doi.org/10.3758/BF03196892>
- Brauch, M. (2012). *Das Sounddesign im deutschen Spielfilm*. Marburg, Germany: Tectum.
- Bullerjahn, C. (2008). Musik und Bild. In H. Bruhn, R. Kopiez, A.C. Lehmann (Eds.). *Musikpsychologie - das neue Handbuch*, (pp. 205-222), Hamburg, Germany: Rowohlt.
- Chion, M. (2012). *Audio-Vision*. Berlin. Germany: Schiele & Schön.
- Cohen, A. J., MacMillan, K., Drew, R. (2006). The role of music, sound effects & speech on absorption in a film: The congruence-associationist model of media cognition. *Canadian Acoustics*, 34(3), 40-41. Retrieved from: <https://jcaa.caa-aca.ca/index.php/jcaa/article/view/1812>
- Cohen, A. J. (2010). Music as a source of emotion in film. In P. Juslin & J. Sloboda (Eds.). *Music and Emotion* (pp. 879-908), Oxford, Oxford University Press.
- Cohen, A. J. (2015). Congruence-Association Model and Experiments in Film Music: Toward Inter-disciplinary Collaboration. *Music and the Moving Image*, 8(2), 5-24. <https://doi.org/10.5406/musimoviimag.8.2.0005>
- Donnelly, K. J. (2015). Accessing the Mind's Eye and Ear: What might Lab Experiments tell us about Film Music? *Music and the Moving Image*, 8(2), 25-34. <https://doi.org/10.5406/musimoviimag.8.2.0025>
- Eklund, A., Nichols, T., Knutsson, H. (2016). Cluster failure: Why fMRI interferences for spatial extent have inflated false positive rates. In *Proceedings of National Academy of Sciences*, 113(28), 7900-7905. <https://doi.org/10.1073/pnas.1602413113>
- Filimowicz, M. (2012). The audio affect image: Five hermeneutic modalities of sound design. *The Soundtrack*, 5(1), 29-36. https://doi.org/10.1386/st.5.1.29_1
- Flückiger, B. (2007). *Sound Design* (3rd ed.). Marburg, Germany: Schüren.
- Goldmark, D. (2013). *Pixar and the animated Soundtrack*. In J. Richardson, C. Gorbman, C. Vernallis, (Eds.). *The Oxford Handbook of New Audiovisual Aesthetics* (pp. 213-226), Oxford, Oxford University Press.
- Görne, T. (2017). *Sound Design - Klang, Wahrnehmung, Emotion*. Munich. Germany: Carl Hanser Verlag. <https://doi.org/10.3139/9783446449046>
- Hoekstra, N. (2012). How to engineer a mood: A study of sound in audiovisual contexts. *Bachelor Thesis* Karlstad University, Sweden. Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:kau:diva-14120>
- Holman, T. (2010). *Sound for Film and Television*. Burlington, MA: Focal Press (Elsevier).
- Kock, M. (2018). *Der Einfluss unterschiedlicher Audiogestaltung bei gleichem Bewegtbild - Der Weg zu einem effektiven Sounddesign*. Berlin, Germany: Schiele & Schön.
- Levitin, D. (2007). *This is your brain on music - Understanding a Human Obsession*. London, UK: Atlantic Books.

Lynch, D. (1998). The monster meets.....David Lynch. Interview in *Home Theater Buyer's Guide*. Retrieved from: <http://www.lynchnet.com/monster.html>

Nagel, F., Kopiez, R., Grewe O., Altenmüller, E. (2007). EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods*, 39(2), 283-290. <https://doi.org/10.3758/BF03193159>

Raffaseder, H. (2010). *Audiodesign*. Munich, Germany: Carl Hanser-Verlag. <https://doi.org/10.3139/9783446423251>

Rumelhart, D. E. (2018). Schemata: The Building Blocks of Cognition. In R.J, Spiro, B.C, Bruce, W.F, Brewer *Theoretical Issues on Reading Comprehension* (pp. 33-58), New York, NY: Routledge.

Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178. <https://doi.org/10.1037/h0077714>

Scholle, C. & Louven, C. (2013). emoTouch for iPad: A New Multitouch Tool for Real Time Emotion Space Research. *Proceedings of the 3rd International Conference on Music and Emotion*, 70, Retrieved from: <http://urn.fi/URN:ISBN:978-951-39-5250-1>

Scholle, C. & Louven, C. (2015). The consistency of continuous ratings and retrospective overall judgements for live performances. *International Conference of Students of Systematic Musicology (SysMus15)*. <https://doi.org/10.1037/pmu0000153>

Scholle, C. & Louven, C. (2017). *emoTouch application*. Retrieved from: https://www.musik.uni-osnabrück.de/forschung/musikpsychologie_und_soziologie/forschungsprojekte/emotouch_en.html

Sonnenschein, D. (2001). *Sound Design*. Studio City, CA: Michael Wiese Productions.

Steinitz, D. (2015, 10 23) Robert Zemeckis über Hollywood *Süddeutsche Zeitung*, weekend edition, Retrieved from: <http://www.sueddeutsche.de/leben/robert-zemeckis-ueber-hollywood-1.2702043>

APPENDIX

Post-test Questionnaire

Immediately after the experiment the participants were asked eight dichotomous (yes/no) questions which they could answer spontaneously without thinking about too long:

- Do you listen to music (e.g. radio, streaming audio) more than one hour per day?
- Do you go to the cinema more than once a month?
- Did or do you play a musical instrument?
- Does an audio track generally help to understand the movie plot?
- Which audio type helps more to understand the film plot: sound effects or music?
- Which audio type lets you feel more deeply and more emotionally?
- Can you describe the impact of the audio track on your audio-visual perception with one of the following four terms? Immersive - a new spatial awareness - 3D - emotional.
- Does an audio track change the perception of time? Does a video with audio seem to be faster in perceived time than the same video without sound?

The Film Plot of the Stimuli

The animated film ‘Samuel Goldenberg and Schmuyle’ describes the struggle of a rich (yellow) and a poor (blue) character, fighting over money and alcohol in a pub. This fight is staged visually as a stop trick animation and aurally as a radio play. The viewer or listener perceives the fight more by ear than by eye. You hear the splintering of glass, crashing of chairs, the shouting during the fight and finally the weeping of the loser: the poor character. The piano composition of Mussorgsky symbolizes each character with a leitmotiv: the rich character is represented by a droning bass melody, the poor one by a repetitive descant leitmotiv that sounds unconfident and miserable. First, both guiding themes are presented consecutively. Later, Mussorgsky lets the two leitmotifs become ‘controversial’: they are played simultaneously, sounding dissonant.

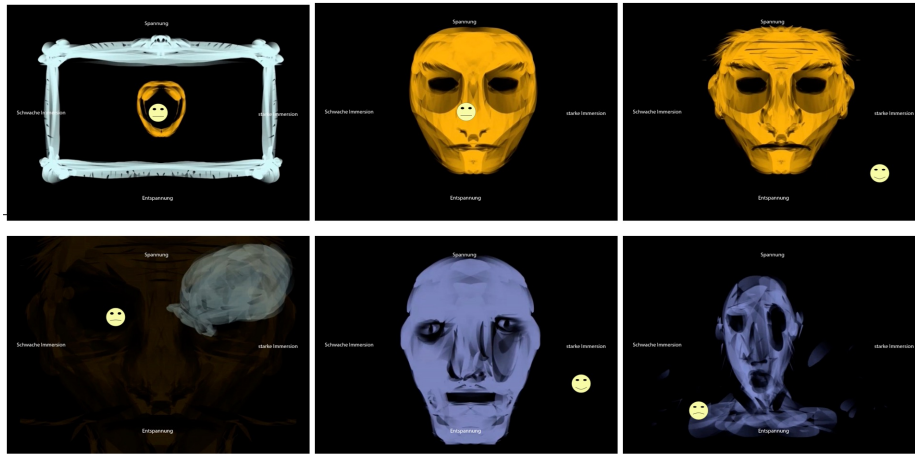


Fig. 9. Animated film ‘Samuel Goldenberg and Schmuyle’ within the *emoTouch* application with the smiley marker in different positions, moved by the participant.

The live-action movie ‘The Catacombs’ is staged with actors in subterranean hallways. These are representing the catacombs of Paris. In Mussorgsky’s programmatic piano composition the catacombs are musically figured as droning and gloomy bass chords; they are played very slowly. The stimulus video enacts a new film plot, taking place in these Parisian hallways: a man accidentally observes a murder.

He is detected by the murderer and tries to escape. The video ends with a cliffhanger. The viewer or listener does not know whether the escape is successful or not. The arranged (diegetic) sound effects support the story visualized in the picture. You hear the steps of all actors and the moaning of the strangled victim. Much reverb acoustically illustrates the hugeness of the catacombs.



Fig. 10. Live-action movie ‘The Catacombs’