

Re-sequencing the Ludic Orchestra

Evaluating the Immersive Effects of Dynamic Music and Situational Context in Video Games

Hans-Peter Gasselseder^(✉)

Department of Communication and Psychology, Aalborg University,
Aalborg, Denmark
hpg@hum.aau.dk

Abstract. A growing number of video games utilize structural and expressive methods to implement music dramaturgy ‘dynamically’ into a non-linear and immersive format. Nevertheless little is known about these musically supported immersive experiences and their processing at the user end. Accordingly, this paper studies the effects of dynamic music on attention allocation and subsequent cognitive-emotional involvement. Self-report data is presented from 60 subjects playing an action-adventure in dynamic as well as static music conditions with low and high arousal expression. The dynamic condition systematically affected imaginary components of immersive experience. Design recommendations for flow and self-location are given with regards to task context and regulatory effects of emotion experience.

Keywords: Dynamic music · Games · Immersion · Presence · Flow · Emotion

1 Introduction

Think of playing a game. You play the game, and the game plays you. The same should apply for its music. However, when trying to create music for video games, developers face the challenge of having to implement dramaturgic conventions of linear storytelling into an interactive non-linear framework. To make matters worse, the complexity of narratives in video games promises to surpass those of traditional audio-visual media in the near future [10]. Factor in the need for supporting challenge-based motives and satisfying aesthetic expectations of the audience, and it becomes apparent: A pre-recorded music loop won’t do. It will either become annoying or make you turn off the game completely. What happened? You certainly were not ‘transported’ to that fictional setting, which was enticingly promised in the game’s marketing material. Nevertheless, if music can destroy that much of your gaming experience, what potential will it hold when it follows your actions, and even more, what if it could express why to perform these actions? These premises form the motivation of the present paper. Dynamic music, while trying to fulfill the above-mentioned shortcomings by adaptively reacting to game events, not always aligns perfectly to the remaining modalities presented in a video game. However, its most important function, as will be argued here, may be seen in the ability to establish

an awareness of situational context; a cognitive representation of how we relate to our surroundings. Knowing what situation one is present in lays out the purpose of an action that will have it appear in a different light when contextualized in the virtual. These non-mediated or immersive experiences arise from a holistic experience of congruent multisensory input [24, 30]. To achieve this congruency in musical terms, standard implementations in the genre action-adventure deploy structural and expressive characteristics of dynamic music. Here, structure refers to the design of horizontal sequencing, which retrieves music cues in a way to match the general, narrative and mood of a scene. In contrast, expression is handed over to the design of vertical re-orchestration that adds and removes individual music tracks according to the portrayed intensity in gameplay. Despite these rather sophisticated techniques, relatively little is known about the involved perceptual processes in playing games with (dynamic) music [10]. The present paper intends to fill this gap by providing an empirical validation of the hypothesized facilitatory effects of dynamic music on attention allocation and emotional involvement. In addition, by covering the aforementioned challenges in the course of the discussion, game designers may gain an empirically founded insight in what ways audio implementations can potentially improve game experience – and what’s more – keep their players in front of the screen.

2 Music Set in Situational Context

Research on non-mediation has seen many descriptions of experiencing situational awareness within a virtual scenario. Terms such as ‘immersion’, ‘presence’, ‘involvement’, ‘absorption’, ‘suspension of disbelief’ as well as ‘self-location’ and ‘flow’ form examples of overlapping constructs that sometimes are used interchangeably and merely denote a specific application, such as virtual reality or challenge-based video games. However divergently defined, it may be argued that common ground is found in the notion of specific motivational states and altered cognitive representations of situational factors. Here, imaginary aspects found in constructs such as absorption and suspension of disbelief are discerned from embodied, sensory-spatial aspects found in flow, self-location, and perceived possible actions. The multiconstruct ‘immersive presence’ [10] aims to consolidate these imaginary and sensory-spatial aspects of non-mediation within a unified framework that incorporates the notion of relational differentials during agency detection. In this view, immersive experiences arise from perceptual processes that juxtapose expected and incoming sensory data as a function of situational demands [5, 10, 23]. Hereby relational differentials operationalize an agent’s current and future state as well as realm of interaction in the environment with regards to expected outcomes for the user. Ascribing purpose and relevance to surrounding events in relation to our own beliefs and desires appears to be a ubiquitous process of perception [32]. Research on theory of mind supports this view in that activity of the mirror neuron system is only observed when actions are attributed to agents but not to non-agents (see [8]). In order to assess a situation, these relational differentials are subsumed to a syntax or reference frame that determines the situational context, which is further projected to subsequent cue juxtapositions and awareness of the range of possible actions. Relational differentials may then be seen in connection to

intrinsic motivation, which in turn is believed to support the experience of flow and the exploration of the environment. It is hypothesized that music achieves immersive experiences by altering relational differentials as a result of directing selective attention and retrieving schemata in function to varying levels of expression-congruency. In doing so, connotations based on prior experiences and cultural codes drive expectations and evaluative functions of music [6, 11]. Applying this information to relational differentials enhances the validity and predictive value of individual cues. The attribution of a reference frame and its associated situation model then emerges from sensations caused by corresponding audio-visual accent structures [3, 6, 21].

The first step of achieving immersive experiences through music in multimedia may be seen in the primal urge of humans to synchronize incoming stimuli [18]. When initializing selective attention and searching for salient cues in the environment, other senses are taken over by the superior temporal resolution of sonic dimensions [18, 27]. At this, a first set of filters directs subsequent hypotheses testing towards congruent percepts [5]. Synchronization ensures the assessing of audio-visual accent structures at contact with visual and other stimuli. If music and the remaining modalities are found to follow similar structural features causing analog sensations, multisensory expectations on emotional congruency towards the situation at focus are formed. If matching combinations of stimuli are found to be congruent to a hypothesis of perception, attention allocation to the media content is intensified [10]. While at this point connotations of music are processed on an extramedial level, that is a conscious integration into relational differentials within the situational context of media reception (e.g. sitting in front of a PC and knowing that music is played back by speakers placed in the room), the emerging reference frame (e.g. defining challenge-based motives) is attributed towards the situational context implied by the media content. This process of 'situational context localization' sets the stage for experiencing imaginary immersion by giving access to portrayed intentions and motivations [9, 10, 30]. At this point only expressive features reach the processing of relational differentials. Previous work by the author suggests that expressive features related to emotional valence may play a dominant role during extramedial processing [10]. This may be due to a basal matching process of synchronization that yet does not fully account for momentary changes. Accordingly, valence is less likely to change spontaneously; suggesting that the potential of music in modulating emotional valence, as for example by the means of minor keys and dissonance, may provide an efficient way of establishing mood and situation. The associated connotations are integrated consciously, meaning that the perceiving subject is still able to discern the presence of music as well as its surface features from the remaining modalities of the media content. This is of relevance for the attribution process of agency, which is negotiated between a subpersonal automatic level for action identification and a more conscious level for sensing agency related representations about intentions, plans and desires [8]. While this hierarchy is asserted for real-life social interactions, it is the contention of the here presented situational context model that in media reception this sequence may be reversed [10]. Hence, the conscious sense of agency pre-exists and is followed by covert automatic processing that couples pre-motor action with a virtual avatar. Compared to automatic bottom-up, the conscious top-down path processes information slower, making it more susceptible to information that is carried by the valence potential of music. For the faster bottom-up

path, however, a more efficient source of information may be seen in the changing levels of arousal potential in music. Having reached extramedial localization, the basal matching process of synchronization may be extended by momentary changes. The gained relevance of the latter allows for lower latency in action identification so that varying levels of music expressed arousal take a dominant role in driving multisensory expectations on emotional congruency. If proportional to the arousal potential of remaining modalities, an increase of arousal in music leads to an intramedial localization of schemata recall. Arising connotations are now unconsciously integrated into relational differentials attributed to the situational context implied in the media content. However, the transition to intramedial localization is gradual insofar as it depends on the latency that music takes to follow the remainder of the accent structure. This latency determines the degree of attributing a particular event or action to an agent, such as the user itself. Note that due to constraints in terms of syntactic structure and scoring conventions, music expression rarely mimics on-screen action directly [6]. Within pre-motor activation, however, music may affect expectations directed towards action readiness or ‘forward models’ at a higher level that encodes global specifications of the action with the controlling and adapting to their goals and underlying motivations [8, 14]. This synchronization of pre-motor activity marks the point when intramedial localization has been reached and the self has become aware of its physical extension towards the possible realm of action and its location. Since an action may become an intrinsic motivator in its own right, it is more likely to be attributed to the self. Within the context of the flow model [7], the additional information provided by music affects the assessment of task demands, while also modulating self-perception of skills [10]. For presence, previous studies have found a correlation between varying levels of induced arousal and self-location [23]. Moreover, forward models may also contribute to discern one’s own thoughts and emotions apart from others, providing the foundation of cognitive empathy [8]. This discrimination may allow relational differentials to become emotionally contagious. Finally, schemata recall and emerging relational differentials are contextualized beyond those motivational ties that were ascribed to the usage situation (e.g. playing for fun in the living room). The situational context model thus operationalizes immersive presence as a mediated perspectivation of situational characteristics that are represented by the media content and its expressed meaning structures [4].

3 Method

60 subjects (23 female, 37 male) aged 18-30 years ($M = 23.72$, $SD = 3.4$) answered self-report questionnaires and rated their emotional state each time after playing a 3rd-person action-adventure video game for 10 min in three randomized conditions accounting for (1) dynamic music, (2) static music/low arousal potential and (3) static music/high arousal potential. Subjects spend on average 2.37 h ($SD = 1.68$) at 2.81 days per week ($SD = 1.78$) with playing digital games. None of the included subjects had played the game ‘Batman: Arkham City’ [25] before. Also note that no subjects identified the experimental manipulation.

3.1 Materials

Because of its guided navigation allowing free movement, the challenge map ‘Penguin Museum’ of the critically acclaimed 3rd person action-adventure ‘Batman: Arkham City’ [25] was set as a stage for investigating immersive experiences. The game’s demands set as per instruction allow the player 10 min of time (as shown on a countdown) to distract enemies from chasing escaping hostages before challenging them in a final battle. The provided timeframe has been shown to be sufficient for immersive experiences to manifest [20]. The orchestral score of ‘Arkham City’, written by Nick Arundel and Ron Fish, makes use of a horizontal sequencing mechanism that reflects calm and confrontational situation changes by musical expression of low and high arousal potential. In addition, the score utilizes a vertical mechanism that reflects dramaturgic aspects ranging from danger to task progress by adding and removing four orchestral stems to the mix relative to the actions and performance of the player. For example, if the avatar subdues henchmen by using stealth strategies, the orchestration dims down to strings only. Additional layers of brass and percussion will be introduced when further enemies are attacked in secret or evolve to a tutti arrangement when the player is being discovered by surprise.

3.2 Instruments

EMuJoy. The emotion software measurement instrument ‘EMuJoy’ [19] operationalizes the circumplex model of emotion [26] in an intuitive visual interface. Here the emotional space is represented as a coordinate system between degree of valence (pleasure-displeasure, X-axis) and arousal (Y-axis). Ratings are given by moving a cursor and pressing a controller. Previous applications have found high re-test and construct correlations of about $r > 0.8$ as well as high consistency between continuous and distinct measures. The present study makes use of distinct measures before and after game presentation as to prevent interference with immersive experience.

iGEQ. Two dimensions taken from the ‘In-Game Experience Questionnaire’ [13] were used to measure subjects’ experience of immersion and flow while playing the game. Each dimension contains a pair of items, which were order-randomized and rated on a Likert-type scale scored from 0–4. Good internal consistencies of about $\alpha = .80$ attest reliable measures for the German translation in use [16]. The dimension ‘imaginative and sensory immersion’ aims to measure narrative elements and associated empathic responses while also considering sensations caused by the audio-visual quality and style of the game. ‘Flow’ describes a holistic sense of absorption and its intrinsic gratification when merged in performing an activity, though it has been found that iGEQ item operationalization primarily address autotelic experiences [22].

MEC-SPQ. Three dimensions taken from the ‘MEC Spatial Presence Questionnaire’ [29] add to the measurement of immersive presence. Each dimension contains four items presented in randomized order and rated on a Likert-type scale scored from 0–4. Prior studies demonstrate good internal consistencies of the questionnaire, $\alpha = .80$ to $\alpha = .92$ [29]. The dimension ‘self location’ refers to a sense of physical projection when interacting with the game. Herein the dimension ‘possible actions’ measures perceived

interactive qualities. ‘Suspension of disbelief’ bears on the cogency of the medium. Here, item operationalization appears to refer on the plausibility of the presentation, in this way accommodating links to absorption and imaginary immersion [10].

3.3 Procedure

The game was displayed on a 15.6" notebook running at 1366×768 pixels, 32-bit, 60 Hz, and second highest graphic settings. Sound was provided on closed stereo headphones (AKG K270 Studio) connected to an audio interface (MOTU 828 mk1) at 30 percent volume. Sound-fx were fed to the monitoring input of a DAW (Apple Logic Pro set at 128 samples buffer) and, for static conditions, mixed with the pre-recorded original music tracks (A-weighted volume matched). Prior testing, subjects pass a 30 min training session involving game mechanics and EMuJoy. Before starting the game, EMuJoy ratings on current emotional state were recorded. Following this, a controller button was pressed to start the game excerpt. The game excerpt is presented in three sessions of 10 min length, each reflecting one out of three music modalities contrasting dynamic/static mechanisms and arousal potential characteristics in randomized order. At the end of each game excerpt an animation of five seconds length signaled successful completion, which marked the point when sound was faded out gradually. Following this, subjects were asked to provide ratings on EMuJoy and to fill out the ‘iGEQ’ and ‘MEC-SPQ’ questionnaires.

4 Results

As shown in Table 1, music accompaniment systematically affected both ratings on imaginary and sensory-spatial aspects of immersive presence in the video game. As against static conditions, Friedman’s test showed significantly higher ratings on ‘imaginary and sensory immersion’ when playing the game with dynamic music, $\chi^2(60) = 6.23$, $p = .04$, while ‘suspension of disbelief’ closely approached significance, $\chi^2(60) = 5.14$, $p = .07$. A more differential result emerged for sensory-spatial components. On the one hand, ‘flow’ saw significantly higher ratings following the low arousal potential versus the high arousal potential and dynamic conditions, $\chi^2(60) = 5.88$, $p = .05$. On the other hand, ratings on ‘possible actions’ approached significance when static music with high arousal potential was presented, $\chi^2(60) = 5.22$, $p = .07$. However, no statistically meaningful differences were observed in ratings on ‘self-location’. Looking at emotion as rated on EMuJoy, reports of valence approached significance after playing the game in static high arousal-potential versus low arousal-potential conditions, $F(59) = 3.83$, $p = .06$. Contrary to expectations, no differences were observed when putting ratings on arousal in low against high arousal-potential conditions, $F(59) = 0.48$, $p = .49$. For the following correlations between latent variables such as emotion and immersive presence, attention is drawn to prior studies in music and social psychology where mean effect sizes range between $r = .21$ and $r = .40$ [15]. Also note that this analysis draws on a local significance level of $p = .017$ (Bonferroni-corrected). Based on Spearman-ranks, a moderate to strong

Table 1. Mean rankings (Friedman’s test statistics), arithmetic mean, standard deviation (sd), median, and interquartile range (iQ) of ratings on dimensions from iGEQ and SPQ. ‘AP’ denotes arousal potential. Asteriks in brackets (*) denote approached statistical significance.

	Dynamic	Non-D. Low AP	Non-D. High AP
Imag. & Sens. Immersion (iGEQ)	2.23 *	1.87	1.90
Mean (sd)	3.59 (.83)	3.22 (.86)	3.18 (.88)
Median (iQ)	3.50 (1.37)	3.50 (1.50)	3.50 (1.50)
Suspension of Disbelief (SPQ)	2.23 (*)	1.89	1.88
Mean (sd)	4.12 (.71)	3.79 (.82)	3.86 (.90)
Median (iQ)	4.00 (1.25)	3.75 (1.25)	4.00 (1.25)
Flow (SPQ)	1.88	2.23 *	1.88
Mean (sd)	3.23 (.98)	3.58 (.83)	3.22 (.91)
Median (iQ)	3.25 (1.50)	3.50 (1.00)	3.50 (1.50)
Self Location (SPQ)	.90	.90	.90
Mean (sd)	2.48 (1.07)	2.45 (1.03)	2.53 (1.07)
Median (iQ)	2.25 (1.63)	2.25 (1.75)	2.38 (1.69)
Possible Actions (SPQ)	1.96	1.83	2.21 (*)
Mean (sd)	3.30 (.79)	3.15 (.87)	3.58 (.83)
Median (iQ)	3.50 (1.25)	3.25 (1.00)	3.75 (1.25)

correlation between post-gameplay arousal and self-location appeared when the game had been played with dynamic music, $r = .37, p = <.01$, and static music with low arousal potential, $r = .32, p = .01$, but was not present in the high arousal potential condition, $r = .22, p = .09$. Further analysis indicates no significant difference between dynamic music and static music in low arousal potential due to the inclusion of zero within the confidence interval, 95 % $CI l = -.27$ and $u = .17$. In contrast to these results, a Spearman-based moderate correlation between the pre-post difference measure of arousal and self-location emerged only in the dynamic music condition, $r = .31, p = .01$, but did not meet significance in static conditions including low arousal-, $r = .22, p = .09$, and high arousal-potential, $r = .23, p = .08$.

5 Discussion

Overall, the results indicate enhanced immersive experiences following gameplay with dynamic as compared to static music with low and high arousal potential. In the dynamic condition, subjects reported higher imaginary and sensory immersion. For suspension of disbelief, however, this effect was only observed when contrasting dynamic music with static low arousal potential music, but not with its high arousal potential variant. In this regard it shall be noted that MEC-SPQ suspension of disbelief item wording does not distinguish between narrative and sensory aspects of immersive experiences. Accordingly, it is to be expected that item responses on suspension of disbelief are more susceptible to expressive characteristics carried over by the music score as well as its interaction effects with other modalities, such as sound-fx and gameplay controls. One of these effects could entail changes in the perceived degree of realism. Differences in arousal potential of music result in differing relative volume of

sound-fx, leading to an increase in saliency of perceptually realistic cues that, amongst others, are held responsible for spatial-sensory states of immersion [31]. Combining the results obtained on imaginary and sensory immersion and suspension of disbelief, it is suggested that dynamic music exerts a strong influence on the narrative-dramaturgic premises during situational context localization. As will be shown below on self-location, such an effect would seem to indicate the presence of a progressive modulation of arousal experience corresponding to gameplay and its integration into dramaturgy, rather than the outcome of stringing together multiple pre-defined cues alone [10]. However, during extramedial localization one may presume the valence potential of music to take the leading role in determining mood and situation. Demonstrating this effect on the current data proves difficult due to the construct discrimination used in the iGEQ. The ambiguous item operationalization of sensory immersion, such as “I found it impressive” confounds both imaginary and sensory aspects of immersion, and may be understood as an overall evaluation of the presentation. Conversely, the scales’ imaginary aspect is clearly marked out when rating “I was interested into the narrative of the game”. The better-defined notion of narrative may explain subjects’ higher ratings following the presentation of dynamic music, suggesting the involvement of other factors apart from arousal potential. Future explorations of this issue may provide a final answer to the interpretation given above. Nevertheless, the differences in the findings on imaginary and sensory immersion as well as suspension of disbelief give reason to presume a dominating role of relatively stable cues, such as implied valence potential in horizontal re-sequencing, during the first steps of extramedial localization.

Moving on to sensory-spatial components of immersive presence, the situational context model stipulates higher sensitivity regarding sensory inputs in contrast to the more abstract nature of imaginary components. For the experience of flow, similar circumstances as found for the ratings on suspension of disbelief seem to have contributed to the obtained results. One aspect of perceptual realism could be seen in the naturalness of environmental feedback [31]. Depending on the perceived latency between an event and an incoming stimulus, subjects are more likely to ascribe an action to agents corresponding to either the avatar or external/extroavatar characters. This effect may have contributed to higher ratings of flow following presentation of static music with low arousal potential as opposed to high arousal and dynamic conditions. In relation to conditions featuring high arousal potential music, sound-fx appear at higher volume when arousal expressed in the music scenery is low, preventing overshadowing from music and thus boosting saliency of immediate feedback. Subsequently, sound-fx operationalize player skills and task difficulty by affording moment-to-moment synchronization of the sense of control over action and environment. Where Csikszentmihalyi’s original model [7] identifies feedback and control as essential components of flow experience, a similar facilitatory effect may be at work in the agent-based action-identification achieved by sound-fx. Here, actions are identified and attributed according to their perceived latency within subpersonal automatic processing structures. The latter may be associated to ‘forward models’ and their subsequent synchronized pre-motor activation, leading to autotelic stages of flow [14]. Having reached an autotelic state, the purpose of the action is fully intrinsically motivated, suggesting that goals have been clearly defined and taken over by the user [7, 10].

As pointed out previously, the role of music is seen in the suggestion of motivational cues and associated goals, but may be limited in terms of following momentary changes within the accent structure of the media content. Thus, in order to achieve emotional congruency towards momentary changes, music would have to mimic several key actions as to underline their meaning within the progress of a subtask. By adding and subtracting orchestral stems in the progress of moving between opponents, typical implementations of dynamic music, like the one used in the present study, rely on a more general abstraction of the task. Consequently one may expect music to influence only top-down processes that relate to corresponding intentions, plans and desires [10]. The goals of these motivational ties must be contextualized in a way that matches narrative and dramaturgic levels, or when put in flow typology, the global and local goals. While the used game excerpt presents the vigilante fight against crime as well as the protecting of the innocent in a grandiose light as global goals, its local goals, as suggested per instruction, involve the not being discovered and distraction of opponents for a 10 min time window. This constellation follows a better match of music cues referring to stealth behavior, which suggests low arousal potential music to be a better fit for mastering local goals as presented in the used game extract. Whenever changes to higher arousal potentials are introduced, a shift from local to global goals may be suggested to the user who subsequently is inclined to readjust the centering of attention so that a fluid course of action is broken. In view of this, special attention must be paid on integrating global and local goals so as to ensure recall of matching musical connotations. Music then can help to contextualize local goals within the main narrative and potentially prevent unwanted sensibility towards weaknesses of game design, such as in striking cases of gamification [10].

A somewhat different pattern of results comes to light when inspecting self-location. Regardless of the music condition used, no differences were found in subjects' ratings. At first sight this would indicate that music does not affect the sense of being physically located into a virtual environment. However, when consulting correlations with responses on experienced arousal, moderate to relatively strong effects appear in conditions featuring dynamic and static music with low arousal potential. Again these effects may be linked to autotelic states arising from pre-motor activation. Forward models facilitate not alone the intrinsic motivation of an action, but also its attribution to the self. By establishing top-down representations of an action from the encoding of global specifications, matching information from bottom up processes are compiled at lower latencies and thus more likely to be attributed to the self [8, 10, 14]. Combined with perceptual realism, which is gained through sound-fx in conditions including low arousal potential, the actual experience of arousal varies with self-location. This result goes in line with predictions made in the situational context model as well as prior studies that identify correlations between felt arousal and presence [24]. Though, it does not explain why subjects experience similar levels of self-location across music conditions. Some of this inconsistency may be due to inhomogeneous qualitative modes of presenting visual and auditory channels [20]. Another viewpoint is provided when looking at the pre-post measure of arousal. Being more sensitive to in-game changes, pre-post arousal is more likely to capture variations of expression in music dramaturgy as opposed to single measurements [10]. In line with this, reports of self-location correlate significantly with pre-post arousal only in

the dynamic music condition. Forward models may account for this insofar as the pre-processing of top-down cognitive empathy affects relational differentials when it is met by the bottom-up processing of emotionally salient cues. If matching arousal expression appears synchronized, spread activation linked to music cognition (e.g. syntactic structure) will be forwarded for integration into the currently active forward model (see [6, 10]). Corresponding to this, the present results indicate a covarying influence of dynamic music on arousal and self-location, albeit with absence of changes in either measure when being rated after gameplay. For the reasons given above as well as the findings obtained from imaginary components, it is presumed that dynamic music holds the potential for protective applications aimed at regulating arousal correlates and cognitive dissonance linked to post gameplay short-term, aggressive tendency without impairing the experience of self-location [for an overview see 1]. While at the time of this writing a follow-up study is in the works, support for this hypothesis may be seen in previous efforts by Grimshaw and colleagues [12] who find combined presentations of sound-fx and music to result in ratings higher on immersion and lower on tension, negative affect as compared to the presentation of either alone. However, it remains to be seen to what extent interactivity in music can contribute to better emotion regulation following gameplay.

Beyond self-location, the present study also asked subjects to report possible actions perceived during gameplay. In line with prior work relating spatial-presence with emotions, the high arousal potential condition received significantly higher ratings compared to low arousal potential and dynamic conditions [24]. Even so, no changes in arousal were observed following the static condition with high potential music. Instead, this condition sees significantly elevated reports of valence, indicating that subjects experience more positive emotions compared to conditions featuring music with low arousal potential. As these measures were taken immediately after having finished gameplay, recency effects may have influenced retrospective reports [2]. Where most prior empirical efforts on the subject fail to ensure consistent feedback to performance across conditions, the present study ensured that subjects completed each game extract successfully. This positive feedback was signaled by a short animation at the end of each trial, but would be translated musically only in the dynamic music condition. Static conditions continued in their pre-defined arousal potential characteristic while fading out slowly (a standard convention in games utilizing static music engines). Though one may expect dynamic music to reinforce positive feedback of the animation, its relieving character and change of tonality suggests lower arousal potential than the continuing music scenery of the high arousal potential condition. Moreover, when consulting expressive parameters reported in the literature and comparing them across conditions, music with high arousal potential shows considerable overlap to the emotion category 'fun', suggesting higher likelihood of experiencing positively valenced emotions when matched with other stimuli [15]. Stevens and Raybould [28] bring up the notion of 'fiero', a high arousal state emerging from the overcoming of obstacles. Being an important source of fun, fiero is likely to appear when high arousal potential characteristics meet the positive reinforcement implied by the visuals [17]. Applying this view on the current findings, it is suggested that the original valence potential in music expression may have given way to congruent percepts of expressed arousal. This could have led to altered schemata recall as a function of situational

context [10]. However, in order to study these interactions in more detail, future work will have to check for effects on valence when game excerpts are not completed successfully and accompanied by congruent music stimuli.

The present study demonstrated dynamic music as an efficient tool to enhance immersive experience in the action-adventure video game genre and has given an outlook in what ways future implementations can profit from strategically utilizing horizontal and vertical mechanisms. This is exemplified in that subjects were unaware of the experimental manipulation, though reports indicated differing game experiences. To these belonged perceived changes in difficulty and appearances of characters as well as objective measures of the number of performed combo moves [see 10]. Consequently, given the growing number of commercial titles featuring suitable audio engines, ludopsychology is in need to gain more knowledge on the emotional experience and semantic processes involved in procedural aesthetics.

References

1. Anderson, C.A., et al.: Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western Countries. *Psychol. Bull.* **136**(2), 151–173 (2010)
2. Atkinson, R.C., Shiffrin, R.M.: Human Memory. In: Spence, K.W., Spence, J.T. (eds.) *The Psychology of Learning and Motivation*, vol. 2. Academic Press, New York (1968)
3. Boltz, M.G., Ebendorf, B., Field, B.: Audiovisual interactions: the impact of visual information on music perception and memory. *Music Percept.* **27**(1), 43–59 (2009)
4. Bruner, J.: *Actual Minds, Possible Worlds*. Harvard University, Cambridge (1986)
5. Bruner, J.S., Postman, L.: On the perception of incongruity. *J. Pers.* **18**, 206–223 (1949)
6. Cohen, A.J.: Music as a source of emotion in film. In: Juslin, P., Sloboda, J. (eds.) *Music and Emotion*. Oxford University Press, Oxford (2001)
7. Csikszentmihalyi, M.: *Flow*. Harper and Row, New York (1991)
8. Decety, J., Grèzes, J.: The power of simulation: imagining one's own and other's behavior. *Brain Res.* **1079**, 4–14 (2006)
9. Ermi, L., Mäyrä, F.: Fundamental components of the gameplay experience. In: de Castell, S., Jenson, J. (eds.) *Proceedings of the 2005 DiGRA Conference*, pp. 15–27 (2005)
10. Gasselseder, H.-P.: *Re-orchestrating game drama*. Diploma Thesis, Department of Psychology, Paris-Lodron University of Salzburg, Austria (2012)
11. Gaver, W.W., Mandler, G.: Play it again. *Sam. Cogn. Emot.* **1**, 259–282 (1987)
12. Grimshaw, M., Lindley, C.A., Nacke, L.: Sound and immersion in the first-person shooter. In: *Proceedings of Audio Mostly*, Piteå, Sweden, 22–23 October 2008
13. IJsselstein, W., et al.: Characterising and measuring user experiences. In: *ACE Conference 2007*, Salzburg, Austria, 15–17 June 2007
14. Jeannerod, M.: *The Cognitive Neuroscience of Action*. Blackwell, Oxford (1997)
15. Juslin, P.N., Lindström, E.: Musical expression of emotions. *Music Anal. Musical Expr. Emot.* **29**(1–3), 334–364 (2011)
16. Klimmt, C., et al.: Identification with video game characters as automatic shift of self-perceptions. *Media Psychol.* **13**(4), 323–338 (2010)
17. Koster, R.: *Theory of Fun for Game Design*. Paraglyph Press, Phoenix (2005)
18. Maasø, A.: Synchronisieren ist unnorwegisch [Dubbing is atypical norwegian]. *Montage AV* **9**(1), 147–171 (2000)

19. Nagel, F., Kopiez, R., Grewe, O., Altenmüller, E.: EMuJoy. *Behav. Res. Methods* **39**(2), 283–290 (2007)
20. Örtqvist, D., Liljedahl, M.: Immersion and gameplay experience. *Int. J. Comput. Games Technol.* **2010**, 1–11 (2010). Article ID 613931
21. Petrini, K., Crabbe, F., Sheridan, C., Pollick, F.E.: The music of your emotions. *PloS One* **6**, 4 (2011). <http://dx.plos.org/10.1371/journal.pone.0019165.pdf>
22. Poels, K., de Kort, Y., IJsselstein, W.: It is always a lot of fun! In: FuturePlay Conference, Toronto, Canada, 15–17 November 2007
23. Popper, A.N., Fay, R.R.: Evolution of the ear and hearing. *Brain Behav. Evol. (Issues and questions)* **50**(4), 213–220 (1997)
24. Robillard, G., Bouchard, S., Fournier, T., Renaud, P.: Anxiety and presence during VR immersion. *CyberPsychol. Behav.* **6**(5), 467–476 (2003)
25. Rocksteady. *Batman: Arkham City*. Warner Bros. Interactive, Burbank, CA (2011)
26. Russell, J.A.: A circumplex model of affect. *J. Pers. Soc. Psy.* **39**, 1161–1178 (1980)
27. Spence, C., Driver, J.: Audiovisual links in exogenous covert spatial attention. *Percept. Psychophys.* **59**, 1–22 (1997)
28. Stevens, R., Raybould, D.: Designing a game for music. Integrated design approaches for ludic music and interactivity. In: Collins, K., Kapralos, B., Tessler, H. (eds.) *Oxford Handbook of Interactive Audio*. Oxford University Press, Oxford (2012)
29. Vorderer, P., et al.: Development of the MEC Spatial Presence Questionnaire, MEC-SPQ. Report for the European Commission. IST Programme ‘Presence Research Activities’ (2004)
30. Wirth, W., et al.: A process model of the formation of spatial presence experiences. *Media Psychol.* **9**, 493–525 (2007)
31. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments. *Presence Teleoperators Virtual Environ.* **7**(3), 225–240 (1998)
32. Zwaan, R.A.: Situation models. *Current Dir. Psy. Sci* **8**, 15–18 (1999)