An Interface for Emotional Expression in Audio-Visuals

Kamer Ali Yuksel
Faculty of Engineering and Natural Sciences, Sabanci University, Orhanli – Tuzla, 34956 Istanbul
kamer@sabanciuniv.edu

Sinan Buyukbas
Faculty of Arts and Social Sciences, Sabanci University, Orhanli - Tuzla, 34956 Istanbul, Turkey
sinanbuyukbas@sabanciuniv.edu

Elif Ayiter
Faculty of Arts and Social Sciences, Sabanci University, Orhanli - Tuzla, 34956 Istanbul, Turkey
ayiter@sabanciuniv.edu

ABSTRACT
In this work, a comprehensive study is performed on the relationship between audio, visual and emotion by applying the principles of cognitive emotion theory into digital creation. The study is driven by an audiovisual emotion library project that is named AVIEM, which provides an interactive interface for experimentation and evaluation of the perception and creation processes of audiovisuals. AVIEM primarily consists of separate audio and visual libraries and grows with user contribution as users explore different combinations between them. The library provides a wide range of experimentation possibilities by allowing users to create audiovisual relations and logging their emotional responses through its interface. Besides being a resourceful tool of experimentation, AVIEM aims to become a source of inspiration, where digitally created abstract virtual environments and soundscapes can elicit target emotions at a preconscious level, by building genuine audiovisual relations that would engage the viewer on a strong emotional stage. Lastly, various schemes are proposed to visualize information extracted through AVIEM, to improve the navigation and designate the trends and dependencies among audiovisual relations.

Keywords
Designing emotive audiovisuals, cognitive emotion theory, audiovisual perception and interaction, synaesthesia

1. INTRODUCTION
As the technology evolves, humans can take one step further to express their-selves free from creative limitations. Finding better ways to express their feelings provides them the genuine transmission and sharing of their experiences between each other. In the end, the question comes to what Howard Rheingold insightfully stated, "If our technology ever allows us to create any experience we might want, what kinds of experience we should create?" [9]. Emotions play highly important role on their mental state as they represent a synthesis of subjective experience, expressive behavior and neurochemical activity to help uniquely define their experience of reality. For that reason, we began by investigating how the human emotion works, in order to create successful emotional signals.

In his work Rhetoric, Aristotle suggests that an emotional response must be triggered by a certain pleasure or pain to be qualified as an "emotion". However, "that is not to say that every person will feel the same pleasure or the same pain with any particular emotion, but if a feeling is to qualify as an emotion it must be attended by some physiological sensation of pleasure or pain" [13]. The emotional response can be thought as a fingerprint of individual's unique emotional perception. In other words, there are several similarities in human's emotional response when it comes to aggregate analysis of behavioral trends. For instance, a group's laugh in a cinema or grief in a funeral clearly points out the emotive trends that people do respond similarly.

Cognitive psychologists emphasize the role of comparison, matching, appraisal, memory, and attribution in the forming of emotions. In his theory of cognitive metaphors, Mark Johnson points out how much of human perception and thinking is affected by repeatable patterns of experience that include: motion, directness of action, degree of intensity and structure of causal interaction; what refers to namely "image schemata" [6]. Johnson states that the meaning of balance –"the bodily experience in which we orient our selves within our environment"- creates a pre-conceptual data in our memory. This bodily experienced knowledge of balance also refers to the feeling of harmony; whereas, the state of "out of balance" corresponds to "disharmony", as we feel fear when we lose balance. Thus, the theory of cognitive metaphors demonstrates that we share subconsciously emotional responses that are rooted in our physically experienced knowledge on reality. In order to create strong emotional stimulus on the viewer, these mentally rooted visual and the audiovisual media effectively uses aural experiences, as we will example in later sections.

Judging the emotional stimuli of a situation is a daily action that humans perform unintentionally. Emotional evaluation of these moments handled mainly by their auditory and visual perception, since they are the essential senses that of communication. Instant decisions are made according to feeling of these audiovisual signals even while zapping fast between TV channels. Within a split second, the emotional perception evaluates the subjective attraction of the received multi-sensory messages, which differentiate between staying or keep zapping until an "emotional hook" [1]. This mysterious equation between audio, visual and emotion finds new platforms to be investigated while technology gives artists new mediums for creative expression. At the latest century, the production of affordable hardware for sound and moving image recording carried the earlier experimentations such as Color Organs at the 18th century one step further [8]. Avant-garde animators, such as Oskar Fischinger, was among the first ones who experimented on the audiovisual synesthesia with the advantage of analog editing, that enabled artists to make temporal connections between sound and moving image.

Not long before, computers set the new rules and gave users the ability to digitally create and edit audio and visual data. A new generation of artists learned to use audio-vision as a tool of artistic expression. Recent audiovisual works cover a wide variety of genres ranging from 3d projection mappings to virtual environments. In order to engage people into unique emotional experiences, music videos by directors such as Edouard Salier and Alex Rutherford, explore different approaches on how to meaningfully connect virtual reality with computer-generated music. A new form of underground art,

which is called VJing, appeared on the club scenes, focusing on live performance by projecting computer generated visual materials to reflect the interactive emotional nature of the venue [3].
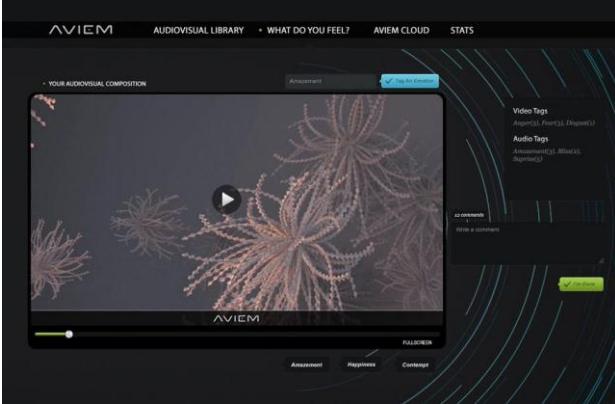

Figure 1. AVIEM interface allowing users to create and compare multisensory compositions and tag their emotions.

In this work, we have developed the project AVIEM (Audio-Visual Emotion) with an inspiration from VJing tools for experimenting on audiovisual relations with an interface that enables users to compare, arrange and create multisensory compositions (Fig. 1). With the integration of a tag database, AVIEM forms an audiovisual library where users can sort all single and multisensory components according to emotion tags added by others. All the visual elements are 3d generated abstract animations describing virtual environments along with digitally created soundscapes, which in the end form all together diverse and engaging emotionscapes. AVIEM puts emphasis on the creation process and strategic use of sound and image in order to catalyze strong emotional response on the viewer. AVIEM aims to employ immersion and interactivity with its audio and visual elements designed to engage participant on an emotional stage, thereby enhances the overall experience of virtual environments. AVIEM library grows and evolves through user contribution through allowing participants to choose, watch and tag the previously created audiovisual compositions by selecting any desired emotion from the interface menu.

In the creation process, AVIEM investigates how to elicit a target emotion using digital abstraction in order to transmit audiovisual signals at a preconscious level. It gives examples on how to trigger an emotion by building associative audiovisual relations that are derived from repeatable patterns of experiences such as motion, directness of action, degree of intensity and so forth. For that reason, digital creation tools are essential for audiovisual design, since they are based on the very same principals. In the evaluation process, AVIEM introduces an interactive audiovisual interface where users are able evaluate audio, visual and audiovisual relations separately. The statistical data of the user evaluation are stored in each libraries database and visualized in order to reflect various aspects of AVIEM library. Such information visualizations are essential to analyze emotional trends and dependencies of the libraries. The interface is designed user friendly in order to reflect the comfortable venue of a library and also guide user through an uninfluenced evaluation.

2. BACKGROUND

Recent theories on cognition and emotion points out that sensorial, cognitive and emotional process of the stimuli are strongly related to each other [1]. According to them, sensorial processing is directly related to audiovisual perception and

consists of parallel and simultaneous synthesis of two different sensorial data: cross-modal or inter-modal processing [10]. Intermodal processing provides the essential data for cognitive and emotional processing by interpreting diverse stimuli from audio and visual channels. Hence, it plays a crucial role on the emotional response emphasizing that the successful reception might rely on the harmonic perception of the moving image and sound [1].

Cognitive evaluation relies on the interpreted stimuli from diverse channels. Recent theories point out that the evaluation process starts with the cognitive attention; meanwhile the divided attention tries to use attentional resources on two or more stimuli. However, the attention has limited capacity and the instant processing of the data from the two different sensory channels is specifically reinforced by the capacity of the brain [1]. Besides, the attention is related with individual's physical and mental well being at moments of intermodal reception and processing. For example, tiredness, sickness or mental and environmental distractions can cause a state of emotional numbness and weaken the impact of the emotional response. Grodal states that cognitive evaluation instantly processes the audiovisual data and starts to build "a web-like structure of associations" by producing semantic relations [4]. He categorizes three mechanisms in this process: Establishing connections; Chunking, grouping (making gestalts); Labeling.

Upon the reception of audiovisual signals, the brain starts to mentally construct the meaning of what we see and hear; some elements are recognized consciously and create links to previous events of memory; at the same time the viewing and hearing activate many associative connections at a preconscious level. Therefore, the meaning that is processed as a result of this complex cognitive evaluation seems to be highly flexible and fluid as it depends on subjective and unconsciously predefined relations [2]. The functional relationships between cognition and emotion are bidirectional on the level of emotional experience. According to Richard Lazarus, the processed meaning of the audiovisual signal generates the emotional response, which "is always a response to cognitive activity that generates meaning regardless of how this meaning is achieved" [7]. Triggering an emotion can cause the activation of a subsequent thought or meaning, which would trigger other emotional responses. According to Carroll Izard's differential emotion theory, "emotion included cognitions may trigger complex memory clusters and cognitions may act as a positive feedback loop and amplify the ongoing emotional state"; as in anger, shame and fear that mostly trigger each other [5].

Another important aspect of the emotional processing is the experienced density of the stimuli, which is primarily dependent on the interpreted data by intermodal processing, regarding how effective the audiovisual signal is designed. It is also dependent on the individual's coping potential: how much he or she can handle an emotional experience. Like in the TV channel-zapping example, one makes instant decisions according to subjectively experienced density of the stimuli. If the individual is getting too much information that he or she can process instantly at that moment, it is highly possible that he or she feels exhausted or uninterested in the audiovisual after a while. To sum up, the construction of the audiovisual meaning and the emotional response as a result are highly subjective matters depending on the individual's personal luggage. However, researches on human behavior have indicated that a target emotive response can be produced through strategic employment of sound and imagery [11, 12]. For that reason, we have focused on the audiovisual perception and examined the techniques for connecting sound and image in more abstract terms in the following subsection.

3. THE PROJECT AVIEM

AVIEM project began by focusing on the creation of audio and visual components in order to investigate what makes visual or aural stimuli intense and effective. The goal was to construct an effective and rich audiovisual library that would provide users a wide range of experimentation through several possible matching combinations among the visual and audio libraries. Creation of such library in the context of academic research purposes requires a few ground rules. First of all, all moving image and sound should be abstract, in order to avoid cultural stereotypes. The ideology of a non-narrative, non-discursive mode of expression dates back to first color organ creations, where subjective interactions of all sensory perceptions, which are directly related to participant's cognitive experience, are eliminated [14]. For most of people, a child's smile refers to hope or a red rose refers to love. AVIEM aimed to convey emotions without repeating such "clichés" to seize accurate emotional responses. In the end, the first challenge was to elicit target emotions with the strategic use of very primary notions such as geometry, color, movement, timbre and pitch.

AVIEM profits from today's cutting-edge computer graphics tools, which allow creating virtual environments that can be perceived almost as real. In contemporary computer simulations, immersion is mostly depending on the accuracy of lighting and shading. With the exploration of global illumination engines in 3d software, artists and designers have became able to create the most physically accurate virtual environments up to date. In this case, immersion is how much people believe in what they see and hear. Thereby, successful immersion of sound and moving image enhances the impact of their audiovisual stimuli making the emotive response of the participants highly intense and lasting. Current creation tools provide users full control over the amodal qualities; various parameters, e.g. intensity, rhythm, and form, can be fine tuned and mapped on to each other in order to design impactful audiovisual mappings. Moreover, they provide the most immersive digital renders up-to-date. Consequently, digital creation tools provide control over the key features for successful immersion and strong emotional response. Creative narration is also one of these key features, since the viewers gain an emotional numbness after consuming similar audiovisual media over and over in daily life.

In this respect, AVIEM library offers a wide range of experimentation and aims to become a source of inspiration for designers and artists that are eager to explore new ways of conveying emotions. In order to preserve the impact of the first impression and give users enough time to live an emotive range of experience, the duration of the audio and visual components within AVIEM are no longer than 40 seconds. Unlike most of the VJ samples and audio loops, they are not designed for synching one-another; knowing that temporally synched audiovisual relations increase the impact of the emotive response. This sacrifice was crucial for acquiring the diversity and interactivity of the library and ensuring a fruitful experimentation for users through numerous cross-model matching combinations. In order to overcome lack of synchronization, each cross-modal component is designed with several dynamic and textural structures to encourage happy accidents of temporal synchronization. For example, when a visual sequence which have several jump cuts from different camera angles (referring to cinematic montage techniques), it is highly possible to witness numerous synched moments if it interacts with a sound of a similar nature that has ups and downs in timbre and pitch with an arrhythmic tempo.

4. THE INTERFACE DESIGN

The interface design of AVIEM emphasizes on easy usage and accessibility in order let user to explore and experiment freely. Since AVIEM database (tag system) grows with user contribution, it is crucial to secure the accurate evaluation by users. Thus, the interface aims to provide a comfortable navigation with its compact and lucid design elements. The interface consists of two main parts: "audiovisual library" and "what do you feel?" These two sections are the core of the interface, as user can input data and provide statistical information for the other two sections that are "AVIEM Cloud" and "Stats".

Audiovisual Library section consists of three parts. First, the tag cloud visualization, where users can browse separately both audio and video library content according to emotion tags (Fig. 3). Second, the video selection tab, in which users choose a visual component between the videos corresponding to the emotion tag selected from video library tag cloud. In the third part, the audio selection tab works in the same manner as the video selection tab, in order to define an audio component. The important point is that users are able to sort and choose different emotions for audio and video components, which makes the final audiovisual evaluation interesting. Users may choose a visual component that addresses to "happiness" along with an audio component that is pre-tagged as "fear". Hence, the opposing nature of selected audio-visual elements is subjected to user evaluation.

Figure 3. AVIEM Cloud allowing users to choose and watch audiovisual compositions by selecting any desired emotion.

The tag cloud visualization and the emotion database are primarily designed to provide an easy navigation, but also not to affect the individual's subjective evaluation. Moreover, users are able to add their preferable emotion tag in both video and audio selection tabs and contribute to the pre-tagged emotion data. The pre-tag data are formed during the testing process with the contribution of 30 individuals. With the integration of tag cloud visualization, the statistic data from each audio and video emotion databases provide visual information on the emotive trends in both audio and visual libraries. Users can observe the quantity of the tags according to their font size. Hence, the tag cloud visualization enables users to see most the dominant emotions in both libraries at a first look. The second section, "what do you feel?" is launched once users choose the audiovisual components and hit the "play together" button. Here, users experience and evaluate their audiovisual creation by adding several emotion tags with the "tag an emotion" button or describing their experience through the comment tab. Finally, they can hit the "I'm done" button to complete their evaluation and view a pop-up screen that summarizes all the information on audio, video and audiovisual tags.

AVIEM Cloud and Stats sections are proposing various approaches in order to visualize the statistic data and strengthen the visual identity of a "library". AVIEM Cloud section is the visualization of an immense library that consists of all the audiovisual combinations created by users. It provides an easy navigation, in order to observe previously created audiovisual combinations. Thus, users are able to view all the previous creations as a whole library, add tags and write comments. Consequently, AVIEM Cloud section provides a wide range of observation and evaluation feature, and encourages users to contribute more. After creating and tagging their creations, users are able view their contribution to the massive pile of creation. The colored particles that symbolize audiovisual compositions can be selected randomly or sorted according to emotion tags. Here, users are able to analyze and compare different approaches to create a target emotion.
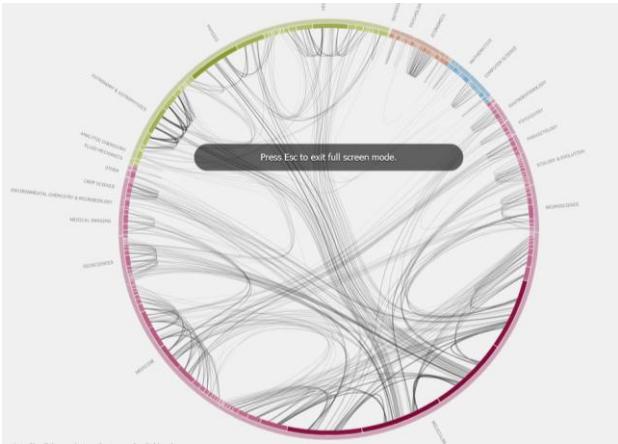

Figure 4. The dependency graph can be utilized to visualize dependencies among emotions within the AVIEM Library.

Lastly, the Stats section aims to visualize the emotional relations and trends in the AVIEM Cloud. The dependency graph can be utilized to visualize dependencies among classes within the AVIEM Library (Fig. 4). The outer and inner ring sections can symbolize the primary and subsequent emotions signifying the quality and the quantity of the class in the database. In this representation, Bezier curves highlight related emotion tags and creations upon selecting an emotion tag. Hence, users are able to see emotions that are often selected together. When AVIEM library gathers a considerable sample in its database with the contribution of users, in theory, the statistic interpretation of the dependency graph may provide key knowledge on emotional trends used in audiovisual media creation.

5. CONCLUSION

This paper documents the design of a tagged database and associated sound and video creation tool aimed at creating (and tagging) emotions elicited from viewers/listeners. The research is inspired by the emergence of "VJing", the practice of real-time creation of video in dance clubs, to complement the DJ's activity of soundscape creation. Moreover, the paper includes a discussion of the subjective and individual and temporal nature of emotional response.

All in all, the AVIEM project covers both the creation and evaluation processes and seeks the unrevealed possibilities of digital creation in audiovisual media design. It combines the very principals of cognitive emotion theory with the most up-to-date digital creation techniques in 3d animation and sound design. It provides experimental knowledge on audiovisual media design and aims to encourage exploration of new techniques emphasizing on the infinite possibilities of today's digital creation tools. In conclusion, AVIEM project run a comprehensive investigation (creation, experimentation and evaluation) on the relationship between audio, visual and emotion in digital creation. With further developing and user contribution, AVIEM aims to become an experimentation and inspiration source and a valuable data resource for cognitive emotion researches.

6. ACKNOWLEGEMENTS

7. REFERENCES

[1] Fahlenbrach, K., Feeling sounds. Emotional aspects of music videos. In Proceedings of IGEL, Hungary, 2002.

[2] Fahlenbrach, K., The emotional design of music videos. Approaches to audiovisual metaphors. Journal of Moving Image Studies, 3(1), 2005, 22-28.

[3] Faulkner, M., VJ: Audio-Visual Art and VJ Culture. London, UK: Laurence King Publishing, 2006.

[4] Grodal, T., Moving pictures. A new theory of film genres, feelings, and cognition: Oxford UP, 2002.

[5] Izard, Carroll E., Human Emotions. New York: Plenum Press, 1977, 100.

[6] Johnson, M., The body in the mind. The bodily basis of meaning, imagination, and reason. Chicago UP, 1987.

[7] Lazarus, R.S., Cognition and motivation in emotion. American Psychologist, 46(4), 1991, 352–367.

[8] Peacock, K., Famous early color organs. Experimental Musical Instruments, 7(2), 1991, 17-20.

[9] Rheingold, H., Virtual Reality. New York: Summit Books, 1991, 116.

[10] Stern, D., The Interpersonal World of the Infant. Stuttgart: Klett-Cotta, 1993.

[11] Thompson, J. G., The Psychobiology of Emotions. New York: Plenum Press, 1988.

[12] Watson, J.B. and Rayner, W., Conditioned Emotional Reactions. Journal of Experimental Psychology, 3(1), 1920, 1-14.

[13] Worth, S. E., Music, Emotion and Language: Using Music to Communicate. Twentieth World Congress of Philosophy, Boston, Massachusetts, 1998.

[14] Zilczer, J., Color Music: Synaesthesia and nineteenth-century sources for abstract art". Artibus et Historiae, 8(16), 1987, 101-126.