Deep Learning for Classification of Peak Emotions within Virtual Reality Systems

Denise Quesnel
School of Interactive Arts & Technology
Simon Fraser University
Canada
dquesnel@sfu.ca

Steve DiPaola
School of Interactive Arts & Technology
Simon Fraser University
Canada
sdipaola@sfu.ca

Bernhard E. Riecke
School of Interactive Arts & Technology
Simon Fraser University
Canada
ber1@sfu.ca

ABSTRACT

Research has demonstrated well-being benefits from positive, 'peak' emotions such as awe and wonder, prompting the HCI community to utilize affective computing and AI modelling for elicitation and measurement of those target emotional states. The immersive nature of virtual reality (VR) content and systems can lead to feelings of awe and wonder, especially with a responsive, personalized environment based on biosignals. However, an accurate model is required to differentiate between emotional states that have similar biosignal input, such as awe and fear. Deep learning may provide a solution since the subtleties of these emotional states and affect may be recognized, with biosignal data viewed in a time series so that researchers and designers can understand which features of the system may have influenced target emotions. The proposed deep learning fusion system in this paper will use data collected from a corpus, created through collection of physiological biosignals and ranked qualitative data, and will classify these multimodal signals into target outputs of affect. This model will be real-time for the evaluation of VR system features which influence awe/wonder, using a bio-responsive environment. Since biosignal data will be collected through wireless, wearable sensor technology, and modelled through the same computer powering the VR system, it can be used in field research and studios.

CCS CONCEPTS

• Human-centered computing ~ Virtual reality • Computing methodologies ~ Artificial intelligence

ADDITIONAL KEYWORDS AND PHRASES

Virtual reality; Deep Learning; Affect; Emotion; Biosignals

1 INTRODUCTION

Awe experiences can lead to shifts in perspective, changes to how people see their relationship with the world. Furthermore, awe tends to provide therapeutic and educational benefits [1-5].

Additionally, awe is correlated with increased willingness to volunteer, and increased life satisfaction [6]. Being awestruck is also good for physical health: of six positive emotions, awe was found to be the strongest predictor of reduction in inflammatory cytokines [7], which are responsible for the initiation and persistence of pain. There are more examples of how being awestruck is good for the environment, social connectivity, and health, and many individuals have their own stories of how natural wonders, spaceflight, transformative life events, and artistic artifacts have elicited awe.

Awe-inspiring events in our natural world are quite rare but when they occur, it is with an immediate and intense manner that can be felt as a sensation of 'chills' or with the elicitation of goose bumps on the skin [8], see **Fig. 1**. This can be measured via physiological biosignals via skin surface video recording, which delivers an objective, unobtrusive means of recording and analyzing triggers of these states. In the study of profound transformative experiences, it is important to be able to objectively monitor the phenomenon for validity, but also to remove the burden of self-report from the participant. Reliance on self-reports of feelings can have major drawbacks, including the need for users to continuously self-monitor which can have the unwanted effect of removing them from an emotional state, and this data typically is retrospective. To address this, Benedek & Kaernbach [9] suggest physiological monitoring of these peak emotional events with many real-time physiological devices now that can collect data. While researchers now understand many of the psychophysiological mechanics that correlate with powerful affective states like awe and fear, there is currently few ways to measure the magnitude of this correlation, and therefore classify the target states accurately.

Figure 1: *Appearance of goose bumps on skin.*

1.1 Motivation

The consequence of successfully prompting an awe-inspiring experience with immersive virtual environments could be profound. Strong experiences of affect like awe are rare phenomena, and suffer from lower intensities in a lab environment due to possible issues with ecological validity [10]. For example, it is difficult to simulate an experience of great natural beauty and vastness to the same level of aesthetic detail and embodiment as in real life, thus missing a critical wellness experience for the participant, especially those with mobility or travel logistic issues. To address this, immersive, virtual environments can be personalized, with stimuli presented in a controlled, private,

In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

physically accommodating manner superior to similar immersive delivery methods (large format screens like IMAX theatres, motion-tracked CAVE automatic virtual environments). Chirico [11] posits that these very personalized features may assist in eliciting higher-intensity awe in a controlled setting. Except for our pilot study of 16 participants in VR [12], this is largely unexplored especially with interactivity features. The reasons for this are largely two-fold: 1) awe-inspiring events are rare in controlled-laboratory environments; and 2) there is a lack of common verbal expression for profound, emotional experiences. Phenomenological methodology, self-reports, and interview data collection methods provide some insight into the experience of awe [8-10, 13-15], but the aforementioned drawbacks exist, with data being retrospective, subjective, with possible misinterpretations between participant and facilitator; and requiring a laboratory environment and extensive time to collect data. Using these methods alone, it is challenging to collect data, or suggest that findings in the lab may transfer into the field.

At the same time, researchers need an affordable tool for measuring and delivering experiences of awe for further understanding of the phenomenon and the events that trigger it. With our proposed study, we aim to utilize a deep learning multimodal fusion model that can extract salient features from multiple biosensors and annotated qualitative events in performing classification of affective states. This system will be part of a wearable tool comprising of a wireless webcam for video of goose bumps and skin conductance sensors (Fig. 2) that can record the user's continuous biosignal data without the need to be in a lab environment. Our aim is that by learning salient features of peak positive emotional experiences like awe, we can identify features in the VR system that prompt these feelings, which in turn will allow us to create more effective VR experiences for profound emotion elicitation.

1.2 Data corpus for emotion and affect in VR

It has been reliably demonstrated that changes with the body's physiological signals represents a shift in emotional state. These changes reflect the way the autonomic nervous system (ANS) and central nervous system (CNS) process internal events within the body and mind. Many physiological sensors can validate emotional states and affective moments, with electrocardiograms (ECG), heart rate (HR), galvanized skin response (GSR), skin conductance (SC), electrodermal activity (EDA), electroencephalogram (EEG), and electromyogram (EMG) all reliably tested [16-17].

More recently, profound shifts in emotional states of awe and wonder have been detected through visible goose bumps on the skin. Kelter & Haidt [1] demonstrated that goose bumps are a distinct central autonomic nervous system marker of awe associated with sympathetic activity. The correlation of chills and goose bumps to awe inspiring, affective experiences are well established [8, 13, 18]. Sumpf, Jentschke, & Koelsch [19] proposed heart rate variability (HRV), and Galvanic Skin Response (GSR) as additional measures to goose bumps [20].

For human computer interaction (HCI) and UX, physiological biosensing are becoming popular for the exploration of user affective experience. Findings have revealed significant correlations between psychophysiological arousal via HR and EDA with self-reported emotion in gameplay [21], with biosensing technology now readily available with the surge of wearable biosensing technology like FitBit for health monitoring, and biosensing interfaces for entertainment (gaming) to replace traditional game controllers.

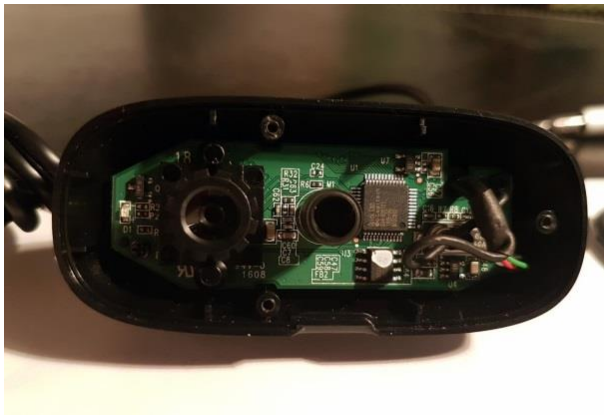

Figure 2: *Interior view of the wireless wearable research tool, consisting of a HD video camera for macro 1 x 1 inch coverage of participant's skin surface for goose bump recording.*

2 DESIGNING AFFECTIVE VR THROUGH AI

2.1 Use of the time series

A time series (TS) is a collection of values obtained from sequentially ordered measurements or values, often visually represented as a database or graph. Measurements are uniformly spaced (time instants) at a given sampling rate. Represented as *T*, a time series is as follows:

$$T = (t_1,\ldots,t_n),\ t_i \in \mathbf{R}. \quad (1)$$

While clusters are frequently used in data mining where labels are often unknown in the dataset, classification differs in that classes are known in advance, and the algorithm is trained on this example dataset [22]. First, we must understand what the features are, as they will belong to different classes. This way, when an unlabeled dataset is fed into the system, it can assign the appropriate class. For this study, TS classification will be the aim rather than clustering, as we can appropriately label intense feelings of positive emotion into categories. We need our system to identify the features as learned through the labeled training data set and classify new datasets. This can be done effectively with Convolutional Neural Networks (CNNs), which are suitable for TS classification through the use of window slicing for classification at the slice level [23], and are robust to scaling issues, i.e: different time series may contain differing time scales, common with biosignal data. As a deep learning method, the features extracted by CNNs are not handcrafted (manually added) as they are with non-deep learning models.

2.2 Pattern recognition and multimodal fusion

Pattern recognition consists of recognizing an object based on its unique attributes and features. While TS classification and pattern recognition is complex enough within a single biosignal TS, it is important to note that a comprehensive, accurate classification

In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

may not be possible without a multimodal approach. It should be noted that humans use multiple modalities to feel and express a state [24]; humans do not emote through one channel alone. An advantage to using deep learning (DL) in CNNs for multimodal biosensing recognition and classification is that the sensors can be collected from an unconstrained environment, meaning no baseline, and no significant data preprocessing or handcrafted feature selection is required [25]; learning can occur through ordered feature representation from raw data. Additionally, CNN based classifiers have been more accurate than k-Nearest Neighbor algorithm and Support Vector Machine models for event classification of multi-senor environments [26-27]. It is for these compelling reasons and because of the accuracy of DL CNN models of emotion and affect that we propose to use a similar method utilizing multi-sensor physiological data. Deep multimodal fusion models demonstrate filter-pooling methods provide the most effective fusion for accurate predictions [28]. User events such as annotated comments made by the participant (discrete signals) are recorded in parallel with physiological (continuous) biosignals, and used in a CNN with a Single Layer Perceptron (SLP). Through fusing aspects of the system such as user events with physiological data, we can better evaluate the VR system for target affect and emotion by understanding the elicitor of the target affect and emotional state, and then further facilitate these targets in VR.

3 METHODS

3.1 Emotional Inducement in VR

For this study, we are interested in feelings of peak positive affect while interacting with immersive VR. Our stimulus is a visualization of the Earth, in which many participants experience a sense of awe and wonder through visiting places on earth [12]. We call this stimulus the 'Earthgazing' content, and is presented in VR, as seen in Fig. 3. To understand how affect-inducing the Earthgazing content is, we need to compare it to content that was created for dissimilar purposes. We choose an educational game of the same length, designed to be informative in nature and not designed to induce positive affect. Both sets of content are created by the researchers in Unity3D, and use the same navigational interface (HTC Vive hand controllers), and head-mounted display (HTC Vive).

Figure 3: *experimental setup with participant 'earthgazing'.*

DEFINITIONS

Cued recall debrief: a video situated recall methodology applied in naturalistic research, with the aim of re-immersing the participant and gaining insight into their thoughts and feelings of the system being evaluated. (Bentley et al., 2005)

Affect: a short term, discrete, and conscious subjective feeling that may have an influence upon a person's overall emotion (Bentley et al., 2005, p. 3). Affective meaning is the degree at which a situation or object changes an individual's reaction to the environment (Duncan & Barrett, 2007).

3.2 Data collection and labelling

For the training data, it is important to accurately label the samples accordingly to their affective state. This will be conducted through a process of cued recall debriefing, which re-immerses the participant with the VR system content and gains insight into their thoughts and feelings [29]. This is a trusted method of data collection and analytics for UX evaluation, because it can isolate features of the system design that may be responsible an affective event. During cued recall, the facilitator can view the biosignals on TS and annotate a comment made by the participant about their affective state (see **Fig. 4**). Such an affective statement may be "Whoa!" indicating surprise or wonder, or "I felt I was part of a greater collective", indicating a potential awe-inspiring experience. Participants may also directly categorize their affective states, such as statements like "entering this scene made me feel afraid right there", or "I felt awe when I looked at the Earth". Such statements can be coded into their appropriate affective states, through thematic analysis. Regardless of how they are coded, the comments made by the participant will be annotated on the time series, but coded after the session with the participant is concluded.

These discrete affective annotations can be transformed into binary continuous signals and when a window is generated around them, features from biosignals may be seen within the window. As each annotated event will be within its own window, the window as a sample will be pairwise ranked for feelings of awe, excitement, motion sickness, and frustration based on the interpretation from the cued recall debriefing (coded thematic analysis).

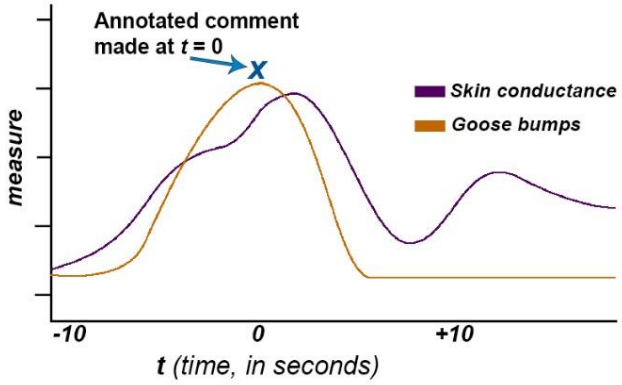

Figure 4: *affective comment annotated within a joint display of physiological data.*

A final step after the cued-recall debriefing is concluded, is that the participant will be asked to rank the entire experience of the Earthgazing content and educational content. The facilitator will

In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

ask the participant to rank the more awe-inspiring/exciting/motion sickness inducing/frustrating pair: A or B; both are equally so; neither are so. The purpose of ranking the entire experience and annotating the events themselves is so that we have rankings for the preference learning of the model across the experience, and within windows centered on events. With this method, we aim for experimental validity and to utilize the framework set by Yannakakis & Martínez [30] that demonstrates rank-based questions yield more reliable models on the constructs of emotion and preference than rating-based questions.

4 MODELING, TRAINING, & CLASSIFICATION

Each neuron defines one local feature which is extracted at every position of the input signal, with the output creating a new signal called a feature map (lower time resolution output). Reduced dimensionality of the convolutional layer is created through filtered feature maps consisting of the fusion of the biosignals and discrete data events (annotated comments from cued recall debriefing) that are transformed to continuous binary signals with a decay rate; see **Fig. 5**. Windows are centered on events, displaying the newly created signal of patterns.

The classification is the output layer of the neural network. In this case, the classification/output is a representation of the affective states. The output will be determined from the outcome of the initial corpus data collection, where researchers will be assigning annotated events from participants into categories of awe, excitement, motion sickness, and frustration. From these categories, classifications will be generated. Automation of feature extraction is possible through Preference Deep Learning (PDL), which was first applied to psychophysiology by Martínez et al., [31] with pairwise preference events of affect across two biosignals. Preference learning handles pairs of data samples (xP, xN), as a model that "outputs higher values for the samples preferred on each pair and lower for the non-preferred" sample [28, p. 4], xP is preferred over/greater than xN. Since we know some of the main features and labels that we'd like to use in our training data but still wish to discover new and unknown features, this is an appropriate technique. With PDL, the SLP is trained to predict the affective state of a user via their biosignals through backpropagation, like Rank Margin error function.

If accurate, this will allow for the VR system to become bio-reactive and adjust the participant's environment to personalize their experience, IE: if they are predicted to be experiencing fear, the system integrates calming audio-visual stimuli; if the participant is possibly experiencing awe, the system may respond with a crescendo of music and aesthetic beauty.

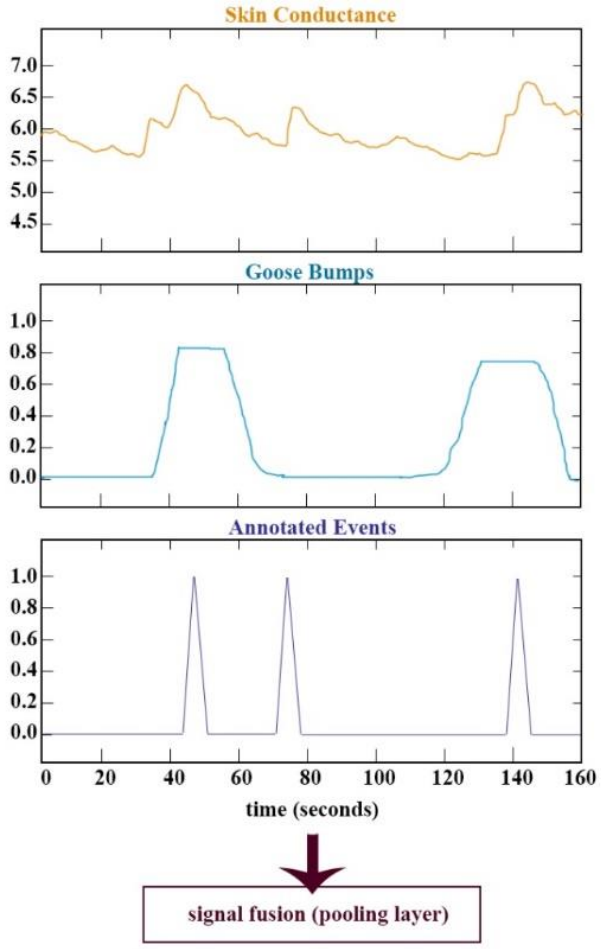

Figure 5: *Fusion of continuous biosignals and events (binary continuous).*

5 FUTURE WORK

Through our proposed study, we aim to augment existing rank-based PDL systems [28, 31] with a validation technique of qualitative cued recall data and categorization of participant affective states for the labelling of our corpus datasets for positive, profound emotion classification. In fusing multimodal physiological signals with reported events of affect like awe and wonder, our aim is to utilize a CNN for accurate classification of output states, demonstrating that complex time series data can be integrated in a model of affect. This work involves the collection of skin conductance and goose bump physiological data alongside the cued recall annotated events from participants for training data (the corpus), and once collected will be made open-sourced for public use. We will test the model and make iterative adjustments to the model parameters, including number of layers, the model will be evaluated for accuracy, and reported with descriptive and inferential statistics. If the system accurately classifies the target output affect states, we will observe whether new salient features have been discovered.

Such a tool as this can be regarded a type of "emotive media", specifically fitting into the characteristic of bio-feedback sensors creating a pathway for interactive applications to discover emotional states [32]. Because emotional VR has been seen to elicit

In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

specific verbal and physical responses in our studies, such as a sharp intake of breath, outbursts of 'whoa!', or physically relaxing in posture [33], it may be valuable to collect basic social signals. Such signals are used in human-robot interaction, with speech, posture, and body orientation data collected and modelled into behavior generation modules for state-sensitive behaviours [34]. Such data could be collected during the creation of the data corpus, and coded based on the cues noted in video recordings. This data could be particularly useful in helping to differentiate between profound affective states that may have similar biosignal features in at least one sensor modality, for example between a fear-inducing moment, or a profound awe-inducing moment, both of which may induce physiological goosebumps- yet verbal or posture data may clearly indicate which state is more probable.

In utilizing a deep learning approach to a wearable biosensing research tool for study of positive peak emotion, we hope the work may be utilized by researchers, designers, and artists for evaluation and creation of bio-responsive, personalized VR environments.

ACKNOWLEDGMENTS

This work is partially supported through a Simon Fraser University graduate student merit-based scholarship award, KEY Big Data Initiative. The authors thank Simon Fraser University for its support.

REFERENCES

[1] Keltner, D. and Haidt, J. 2003. Approaching awe, a moral, spiritual, and aesthetic emotion. *Cognition and Emotion*. 17, 2 (Jan. 2003), 297–314. DOI:https://doi.org/10.1080/02699930302297.

[2] Piff, P.K. et al. 2015. Awe, the small self, and prosocial behavior. *Journal of Personality and Social Psychology*. 108, 6 (2015), 883–899. DOI:https://doi.org/10.1037/pspi0000018.

[3] Riva, G. et al. 2016. Transforming Experience: The Potential of Augmented Reality and Virtual Reality for Enhancing Personal and Clinical Change. *Frontiers in Psychiatry*. 7, (2016), 164. DOI:https://doi.org/10.3389/fpsyt.2016.00164.

[4] Shiota, M.N. et al. 2011. Feeling good: autonomic nervous system responding in five positive emotions. *Emotion (Washington, D.C.)*. 11, 6 (Dec. 2011), 1368–1378. DOI:https://doi.org/10.1037/a0024278.

[5] Shiota, M.N. et al. 2007. The nature of awe: Elicitors, appraisals, and effects on self-concept. *Cognition and Emotion*. 21, 5 (Aug. 2007), 944–963. DOI:https://doi.org/10.1080/02699930600923668.

[6] Rudd, M. et al. 2011. Awe Expands People's Perception of Time, Alters Decision Making, and Enhances Well-Being. *NA - Advances in Consumer Research Volume 39*. (2011). DOI:https://doi.org/10.1177/0956797612438731.

[7] Stellar, J.E. et al. 2015. Positive affect and markers of inflammation: discrete positive emotions predict lower levels of inflammatory cytokines. *Emotion (Washington, D.C.)*. 15, 2 (Apr. 2015), 129–133. DOI:https://doi.org/10.1037/emo0000033.

[8] Grewe, O. et al. 2009. The Chill Parameter: Goose Bumps and Shivers as Promising Measures in Emotion Research. *Music Perception: An Interdisciplinary Journal*. 27, 1 (Sep. 2009), 61–74. DOI:https://doi.org/10.1525/mp.2009.27.1.61.

[9] Benedek, M. and Kaernbach, C. 2011. Physiological correlates and emotional specificity of human piloerection. *Biological Psychology*. 86, 3 (Mar. 2011), 320–329. DOI:https://doi.org/10.1016/j.biopsycho.2010.12.012.

[10] Silvia, P.J. et al. 2015. Openness to experience and awe in response to nature and music: Personality and profound aesthetic experiences. *Psychology of Aesthetics, Creativity, and the Arts*. 9, 4 (2015), 376–384. DOI:https://doi.org/10.1037/aca0000028.

[11] Chirico, A. et al. 2016. The Potential of Virtual Reality for the Investigation of Awe. *Human-Media Interaction*. (2016), 1766. DOI:https://doi.org/10.3389/fpsyg.2016.01766.

[12] Quesnel, D., & Riecke, B. E. (2017). Awestruck: Natural interaction with virtual reality on eliciting awe (pp. 205–206). IEEE. https://doi.org/10.1109/3DUI.2017.7893343

[13] Benedek, M. et al. 2010. Objective and continuous measurement of piloerection. *Psychophysiology*. 47, 5 (Sep. 2010), 989–993. DOI:https://doi.org/10.1111/j.1469-8986.2010.01003.x.

[14] Gallagher, S. et al. 2015. *A Neurophenomenology of Awe and Wonder*. Palgrave Macmillan UK.

[15] Grewe, O. et al. 2011. Chills in different sensory domains: Frisson elicited by acoustical, visual, tactile and gustatory stimuli. *Psychology of Music*. 39, 2 (Apr. 2011), 220–239. DOI:https://doi.org/10.1177/0305735610362950.

[16] Li, L. and Chen, J. h 2006. Emotion Recognition Using Physiological Signals from Multiple Subjects. *2006 International Conference on Intelligent Information Hiding and Multimedia* (Dec. 2006), 355–358.

[17] Szwoch, W. 2015. Emotion Recognition Using Physiological Signals. *Proceedings of the Mulitimedia, Interaction, Design and Innnovation* (New York, NY, USA, 2015), 15:1–15:8.

[18] Schurtz, D.R. et al. 2011. Exploring the social aspects of goose bumps and their role in awe and envy. *Motivation and Emotion*. 36, 2 (Sep. 2011), 205–217. DOI:https://doi.org/10.1007/s11031-011-9243-8.

[19] Sumpf, M. et al. 2015. Effects of Aesthetic Chills on a Cardiac Signature of Emotionality. *PLOS ONE*. 10, 6 (Jun. 2015), e0130117. DOI:https://doi.org/10.1371/journal.pone.0130117.

[20] Colver, M.C. and El-Alayli, A. 2016. Getting aesthetic chills from music: The connection between openness to experience and frisson. *Psychology of Music*. 44, 3 (May 2016), 413–427. DOI:https://doi.org/10.1177/0305735615572358.

[21] Drachen, A. et al. 2010. Correlation Between Heart Rate, Electrodermal Activity and Player Experience in First-person Shooter Games. *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games* (New York, NY, USA, 2010), 49–54.

[22] Esling, P. and Agon, C. 2012. Time-series Data Mining. *ACM Comput. Surv.* 45, 1 (Dec. 2012), 12:1–12:34. DOI:https://doi.org/10.1145/2379776.2379788.

[23] Guennec, A.L. et al. 2016. Data Augmentation for Time Series Classification using Convolutional Neural Networks. (Sep. 2016).

In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

[24] Chen, L.S. et al. 1998. Multimodal human emotion/expression recognition. (1998), 366–371.

[25] Sarkar, S. et al. 2016. Wearable EEG-based Activity Recognition in PHM-related Service Environment via Deep Learning | PHM Society. *INTERNATIONAL JOURNAL OF PROGNOSTICS AND HEALTH MANAGEMENT*. 7, Special Issue Big Data and Analytics (Sep. 2016), 10.

[26] Tong, C. et al. 2017. A convolutional neural network based method for event classification in event-driven multi-sensor network. *Computers & Electrical Engineering*. (Jan. 2017). DOI:https://doi.org/10.1016/j.compeleceng.2017.01.005.

[27] Zhang, Y. et al. 2013. Multi-metric Learning for Multi-sensor Fusion Based Classification. *Inf. Fusion*. 14, 4 (Oct. 2013), 431–440. DOI:https://doi.org/10.1016/j.inffus.2012.05.002.

[28] Martínez, H.P. and Yannakakis, G.N. 2014. Deep Multimodal Fusion: Combining Discrete Events and Continuous Signals. *Proceedings of the 16th International Conference on Multimodal Interaction* (New York, NY, USA, 2014), 34–41.

[29] Bentley, T. et al. 2005. Evaluation Using Cued-recall Debrief to Elicit Information About a User's Affective Experiences. Proceedings of the 17th Australia Conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future (Narrabundah, Australia, Australia, 2005), 1–10.

[30] Yannakakis, G.N. and Martínez, H.P. 2015. Ratings are Overrated! *Frontiers in ICT*. 2, (2015). DOI:https://doi.org/10.3389/fict.2015.00013.

[31] Martinez, H.P. et al. 2013. Learning deep physiological models of affect. *IEEE Computational Intelligence Magazine*. 8, 2 (May 2013), 20–33. DOI:https://doi.org/10.1109/MCI.2013.2247823.

[32] Lugmayr, A. 2016. Emotive media: a review of emotional interfaces and media in human-computer-interaction. Proceedings of the 28th Australian Conference on Computer-Human Interaction (OzCHI '16) (2016), 338–342. DOI: 10.1145/3010915.3010982

[33] Quesnel, D. and Riecke, B. 2017. Connected Through Awe: Can Interactive Virtual Reality Elicit Awe for Improved Well-Being? Poster Presented at the 3rd Annual Innovations in Psychiatry and Behavioral Health: Virtual Reality and Behavior Change (Stanford University, CA, Oct. 2017). DOI: 10.13140/RG.2.2.22177.10088

[34] Sun, M., Zhao, Z. and Ma, X. 2017. Sensing and Handling Engagement Dynamics in Human-Robot Interaction Involving Peripheral Computing Devices. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17) (2017), 556–567. DOI: 10.1145/3025453.3025469

In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal