1-1-2011

An ambient multimedia user experience feedback framework based on user tagging and EEG biosignals

Eva Cheng
RMIT University, ecc04@uow.edu.au

Stephen J. Davis
University of Wollongong, stdavis@uow.edu.au

Ian Burnett
RMIT University, ianb@uow.edu.au

Christian H. Ritz
University of Wollongong, critz@uow.edu.au

An ambient multimedia user experience feedback framework based on user tagging and EEG biosignals

Abstract

Multimedia is increasingly accessed online and within social networks; however, users are typically limited to visual/auditory stimulus through media presented onscreen with accompanying audio over speakers. Whilst recent research studying additional ambient sensory multimedia effects recorded numerical scores of perceptual quality, the users' time-varying emotional response to the ambient sensory feedback is not considered. This paper thus introduces a framework to evaluate user ambient quality of multimedia experience and discover users' time-varying emotional responses through explicit user tagging and implicit EEG biosignal analysis. In the proposed framework, users interact with the media via discrete tagging activities whilst their EEG biosignal emotional feedback is continuously monitored in-between user tagging events with emotional states correlated with media content and tags.

Keywords

ambient, multimedia, user, experience, feedback, framework, based, user, tagging, EEG, biosignals

Disciplines

Physical Sciences and Mathematics

Publication Details

Cheng, E., Davis, S. J., Burnett, I. & Ritz, C. H. (2011). An ambient multimedia user experience feedback framework based on user tagging and EEG biosignals. 4th International Workshop on Semantic Ambient Media Experience (SAME 2011) (pp. 61-66). Tampere:

An Ambient Multimedia User Experience Feedback Framework Based on User Tagging and EEG Biosignals

Eva Cheng[1], Stephen Davis[2], Ian Burnett[1], Christian Ritz[2]

[1]School of Electrical and Computer Engineering
RMIT University
Melbourne, VIC, Australia
{eva.cheng, ian.burnett}@rmit.edu.au

[2]School of Electrical, Computer and
Telecommunications Engineering
University of Wollongong
Wollongong, NSW, Australia
{stdavis, critz}@uow.edu.au

ABSTRACT
Multimedia is increasingly accessed online and within social networks; however, users are typically limited to visual/auditory stimulus through media presented onscreen with accompanying audio over speakers. Whilst recent research studying additional ambient sensory multimedia effects recorded numerical scores of perceptual quality, the users' time-varying emotional response to the ambient sensory feedback is not considered. This paper thus introduces a framework to evaluate user ambient quality of multimedia experience and discover users' time-varying emotional responses through explicit user tagging and implicit EEG biosignal analysis. In the proposed framework, users interact with the media via discrete tagging activities whilst their EEG biosignal emotional feedback is continuously monitored in-between user tagging events with emotional states correlated with media content and tags.

Keywords
Ambient sensory effects, quality of multimedia experience, tagging, EEG biosignals, social networks.

INTRODUCTION
Readily available Internet-enabled multimedia mobile/camera devices and broadband Internet has resulted in a plethora of multimedia being made available online. Currently, user personal consumption of multimedia is largely through visual media displayed on a screen with accompanying audio (on a mobile phone, tablet, laptop, PC), primarily only stimulating the human audio/visual senses. Recent research has thus considered enhancing a user's Quality Of Multimedia Experience (QoMEX) through ambient sensory effects such as light, wind, and vibration to engage other human senses in addition to vision and audition [10] [11][12].

In addition, recent standardization efforts from the ISO/IEC MPEG community include the Sensory Effect Description Language (SEDL), which aims at compact representation of ambient sensory effect metadata (SEM) for multimedia. Utilising SEDL, Waltl et al. [10][11][12] proposed a test-bed for ambient sensory multimedia experience in the form of a video sensory annotation tool, sensory simulation tool, Sensory Effect Media Player (SEMP) and real test environment using the amBX (amBient eXperience) hardware platform. Subsequent user subjective tests concluded that videos accompanied with amBX sensory effects rated higher Mean Opinion Scores (MOS) than videos without effects over varying bitrates, especially for documentary content (rather than fast action sequences) [11].

Whilst Waltl et al. [10][11][12] utilized a modified MOS metric to measure user QoMEX; even in continuous use MOS provides a numerical score of perceptual quality that does not convey the user's time-varying emotional response to sensory feedback. Thus, this paper aims to extend upon the work of Waltl et al. [10][11][12] to propose a multimedia QoMEX evaluation framework that seeks to collect time-varying user emotional feedback. The proposed framework draws together explicit user tagging activities with implicit user biosignal feedback, such as electroencephalography (EEG) signals that can be directly mapped to human emotional states [1][2]. In addition to evaluating user emotional response to sensory effects, user sensory preferences can also be recorded to personalise the effects depending on the user's content preferences, mood etc. Current methods to derive sensory effects for media content typically present the same effects for each user [10][11][12]; however, in reality, different users are likely to respond differently to varying sensory effects.

The proposed framework builds upon the work of Davis et al. [4], which introduced a social multimedia adaptation framework that utilised semantics derived from user media preferences. In [4], user preferences were inferred from users' interactions with media e.g., tagging media content within social groups/networks. Whilst [4] presented a multimedia adaptation framework, as part of the adaptation process, user QoMEX data and feedback are collected and thus the framework can also be utilized for QoMEX evaluation. This paper hence extends upon the framework of [4] for the evaluation of ambient multimedia experiences and collection of user sensory preferences: users explicitly indicate a response through tagging activities whilst their EEG biosignal emotional feedback is continuously

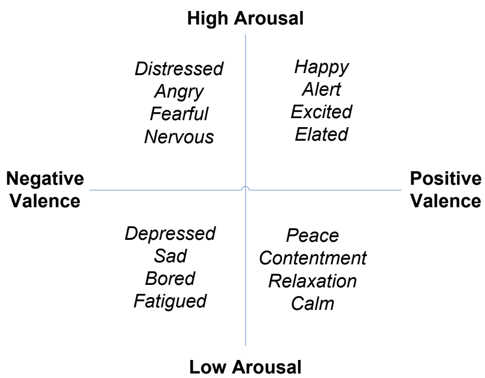

Fig. 1. Valence/arousal emotional space

monitored in-between user tagging events with emotional states correlated with media content and tags.

BACKGROUND

Emotional feedback for multimedia content is often obtained in the form of user-entered descriptions and tags. For example, on YouTube users can add titles, descriptions and tags/keywords to their videos whilst viewers can comment and tag like/dislike (also available on Facebook). One disadvantage of user-entered tags, however, is tag ambiguities between users [5]. For example, what one user tags as 'amusing' can be tagged as 'happy', 'hilarious' or 'funny' by other users; thus, similar emotional responses often elicit varying emotional tags from users. Further, whilst multimedia tags allow for user-specific emotional feedback, the act of tagging is a deliberate and discrete user-driven event. In contrast, multimedia content and thus corresponding ambient sensory effects can significantly change within a video (and even within a scene). Davis et al. [3][4] found user emotions and responses vary greatly in-between explicit tagging events and thus temporal user emotional feedback is required for ambient experience evaluation, rather than a single score for the entire media clip.

To overcome these constraints of explicit user tagging for capturing user emotional feedback, recent research has shown that user biosignals (e.g., facial expressions, galvanic skin response (GSR), electrocardiogram (ECG), electromyogram (EMG) and electroencephalography (EEG)) can provide valuable multimedia feedback [7][9][13]. EEGs are of particular interest as they are less susceptible to voluntary manipulation (that facial expressions are prone to) and ongoing research suggests that EEGs can directly map to human emotional states [1][2]. In addition, affordable consumer EEG headsets such as the Emotiv EPOC[1] and Neurosky Mindset[2] have recently become available on the market, thus enabling user EEG signal collection in personal computing environments to be of practical possibility.

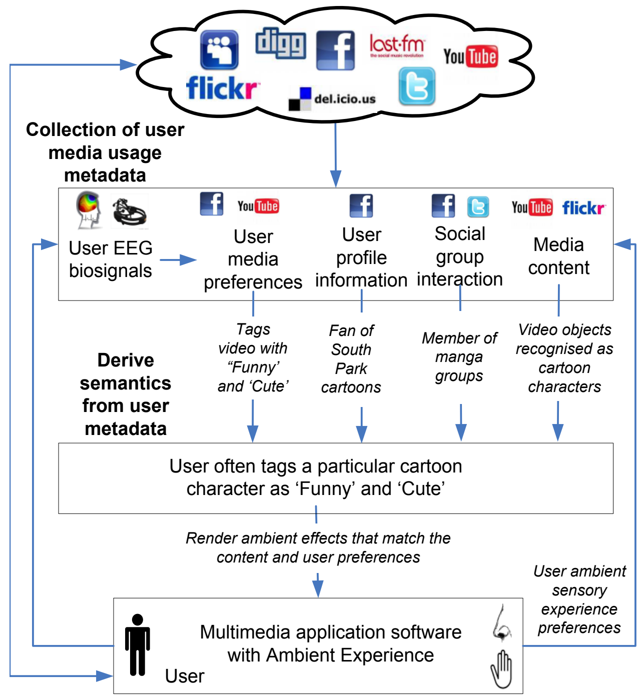

Fig. 2. Proposed ambient sensory feedback framework

Yazdani et al. [13] proposed the use of a medical-grade EEG cap for implicit emotional tagging of multimedia, where EEG signals were clustered to form user emotional responses based on corresponding tags. However, the EEG biosignals replaced an explicit emotional tagging interface rather than inferring implicit tags or emotions from the user's real-time physiological response. Various methods to directly map EEG signals to emotional states have been studied: Choppin [2] applied neural network classifiers for three emotional states based on the valence/arousal model from cognitive theory (see Fig. 1) [8]. As shown in Fig. 1, an emotion exists as a point in 2D continuous valence/arousal space, where valence indicates the extent of positive/negative emotion whilst arousal represents the degree of excitement. Also adopting this valence/arousal paradigm, Chanel et al. [1] found EEGs to highly correlate (compared to other biosignals) to emotional states. Thus, as a key indicator to revealing a user's emotional state mapped into continuous valence/arousal space, continuous EEG biosignal feedback is a valid measure to explore for evaluating users' ambient multimedia experiences.

PROPOSED FRAMEWORK

The proposed framework aims to combine and build on the bodies of work discussed in the previous Section to present an ambient QoMEX evaluation methodology based on user tagging and EEG emotional feedback. Fig. 2 illustrates the proposed framework extended from Davis et al. [4] to utilise tagging activities combined with EEG biosignal feedback for evaluating ambient multimedia experiences. The framework in Fig. 2 collects user experience-based Quality of Perception (QoP) metadata: analysing the tagging activity whilst consuming media, continuously

[1] http://www.emotiv.com/apps/epoc/299/
[2] http://www.neurosky.com/

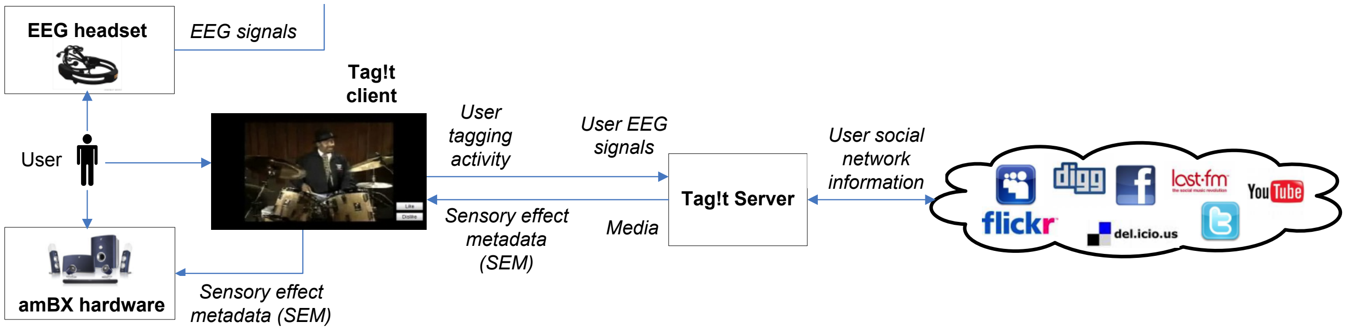

Fig. 3. System flow diagram

collecting EEG biosignal responses, and performing content-based media analyses. Due to the prevalent sharing and consumption of media via social networks, additional metadata can also be collected from the user's volunteered profile information and interaction on social networks to further build up knowledge about the user's preferences. Fig. 2 loosely groups the different metadata types into four categories [4]:

- User media preferences e.g., video interaction (start/stop/pause etc.) and tagging (like/dislike/funny etc.), augmented with EEG biosignal feedback as proposed in this paper;
- User profile information e.g., from social networks;
- Social group interaction e.g., collaborative tagging activity, sharing of multimedia content; and
- Media content-based metadata e.g., signal processing analysis.

Whilst tagging within this framework was explored in [4], this paper introduces continuous EEG user emotional feedback (e.g., using consumer EEG headsets) to map user emotional responses to tagging activity and media content in response to ambient experiences.

SYSTEM IMPLEMENTATION

The authors are implementing the ambient QoMEX evaluation framework of Fig. 2, and the system flow diagram is shown in Fig. 3. Utilising presently available technologies, the implementation involves consumer EEG headsets with accompanying SDKs (Emotiv EPOC and Neurosky Mindset), the amBient eXperience (amBX) computer peripheral hardware and SDK (as used by Waltl et al. [10][11][12])[3], and a modified version of the Tag!t tagging client/server technology tool [3].

As shown in Fig. 3, the user views videos using the Tag!t client interface, which communicates with the Tag!t server to send user tagging activity and EEG signals (interfaced to the EEG headset). The Tag!t client receives ambient sensory effect metadata (SEM) from the Tag!t server, and renders the effects on the amBX hardware through the amBX API.

EEG Signal Collection

Utilising the headset SDKs, the user's EEG signals (or emotional states predefined for the headsets) are periodically polled each second and collected for analysis. The single-sensor Neurosky Mindset, in addition to access to raw EEGs as delta, theta, alpha, beta and gamma waves, defines two emotional states using propriety algorithms: attention (i.e., focus) and meditation (i.e., calmness). Similarly, the 14-sensor Emotiv EPOC provides access to the raw electrode signals, facial expression sensing (via facial muscle EEGs), cognitive thought sensing, dual-axis gyroscope (for head movements) and five emotional states defined using propriety algorithms: engagement/boredom, meditation, frustration, and short and long-term excitement. Whilst initial studies can utilise the emotional states predefined in the headset SDKs, the framework aims to support directly mapping EEG signals to varying emotional states building on learning classification techniques [1][2].

Ambient Multimedia Effects Rendering

The amBX hardware is composed of one rumble wrist pad with two variable speed motors, two fans of variable speed control up to 5000 rpm, two 40W speakers with LED array satellite lights, 80W subwoofer, and high power LED wall-washer lights of 3 LED arrays capable of over 16 million RGB colours. The Windows SDK for the amBX hardware is freely available, and allows for direct control of the amBX hardware via the API by parsing the sensory effect metadata (SEM) sent from the Tag!t server.

The amBX effects can be either predefined and stored on the Tag!t server or media-specific effects derived in real-time though content analysis algorithms running on the Tag!t server (see Fig. 2). For the preliminary evaluations of the proposed framework, amBX effects are predefined and generated offline. Different colour effects corresponding to video content are derived based on average frame colour and dominant frame colour, as obtained through colour histograms [10]. Whilst light effects are continuous for all video content, wind (fan) and force feedback (rumble) are manually derived for relevant video content.

Sensory Effect Metadata

Sensory Effect Metadata (SEM) is represented using the Sensory Effect Description Language (SEDL), currently undergoing standardisation in MPEG-V Part 3 [6]. For flexibility, sensory effects, which can be application-

[3] http://www.ambx.com/

```
<SEM>

  <GroupOfEffects si:pts="57" duration="200" fade = "30" position = "urn:mpeg:mpeg-
  v:01-SI-PositionCS-NS:center:*:front">
      <Effect xsi:type="sev:LightType" intensity-value="0.5478" intensity-
      range="0.0000 1.000" color="#808080 />
      <Effect xsi:type="sev:WindType" intensity-value="0.2061" intensity-
      range="0.0000 1.000" />
      <Effect xsi:type="sev:VibrationType" intensity-value="0.1467" intensity-
      range="0.0000 1.000" />
   </GroupOfEffects>

  <GroupOfEffects ...>
      <Effect ... />
  </GroupOfEffects>

</SEM>
```

Fig. 4. Example SEM

Fig. 5. Screenshot of Tag!t client interface

dependent, are not defined in SEDL but within a Sensory Effect Vocabulary (SEV).

SEDL is an XML schema-based language, an example SEM for the amBX hardware is shown in Fig. 4. For brevity, a series of effects for a given timestamp (`pts`) can be defined in a `GroupOfEffects` element. All attributes are inherited by the child `Effect` elements e.g., `duration` and `fade` times, and `position` of the effects. For the amBX API, device intensities are restricted to a floating point range of 0 to 1, hence the `intensity-range` and `intensity-value` attributes shown in Fig. 4.

Tag!t User Client Tagging Interface

The collection of user media tagging metadata requires a tagging interface tool, developed as 'Tag!t' by Davis et al. [3]. Tag!t has been modified by the authors to additionally collect and send EEG signals from consumer headsets, with an example interface screenshot shown in Fig. 5. Tag!t is a client/server application that allows users to tag multimedia directly from their client, be it a web browser or standalone Tag!t application, where the user is presented with a video and a separate tagging panel (see Fig. 5). Whilst the example interface in Fig. 5 supports like/dislike tagging akin to the like/dislike video user interaction on YouTube and Facebook, Davis et al. [3] have studied the use of emotion tags ('emotitags'), prompted tagging, different tagsets and user-defined tags.

Tag!t Server-Side Processing

The Tag!t server is configured in a Linux, Apache, MySQL and PHP (commonly known as LAMP) environment and implemented using the Recess RESTful PHP framework. All user tagging activity and EEG data from the Tag!t client are sent to the server and saved into a database. As the media file is streamed via the Tag!t server (see Fig. 3), the amBX effects metadata can either be retrieved from predefined SEM files or derived in real-time based on media content analysis (see Fig. 2) and delivered to the amBX hardware via the Tag!t client.

DISCUSSION AND EVALUATION

To evaluate the proposed framework, the key research questions aim to discover users' time-varying emotional responses to ambient multimedia experiences. Whilst Waltl et al. [10][11][12] showed that users do enjoy amBX effects, how does the quality of experience vary within a video, would users prefer parts of videos containing certain content to render more or less effects, at what rate of change do effects cease to improve ambient QoMEX and become annoying/unpleasant, and how to effects preferences vary with corresponding modality (i.e., video/audio stimulus)? Further, building on the social multimedia findings of Davis et al. [3][4], the proposed framework (see Fig. 2) can also investigate whether users exhibit shared sensory preferences within social groups e.g., are users who are fans of horror movies likely to enjoy added thrilling sensory effects, whereas other users may not desire the additional suspense?

Typically, for QoMEX evaluation experiments to investigate the research questions above, users will be instructed to watch a series of videos with and without ambient sensory effects, and asked to tag at any point during the videos whilst wearing an EEG headset. EEG signals are continuously recorded with all tagging activity and EEG data sent to the Tag!t server for processing and

Fig. 6. Example amBX evaluation

evaluation. Fig. 6 illustrates an example user evaluation setup using the Emotiv headset, where the room lighting has been accentuated for visibility; hence, the amBX lighting effects appear muted. Whilst the user is seated quite close to the amBX and video stimulus in Fig. 6, different audience distances will be investigated to discover 'optimal' distances for experiencing near-field ambient effects such as the wind/fan, and far-field effects such as the wall-washer lights.

By collecting and evaluating users' ambient QoMEX, user preferences can also be gauged from their emotional response to effects and thus QoMEX improved by personalising effects and adapting multimedia according to media content preferences, the user's emotional state as detected from their EEG signals, or according to their social group preferences. Further, shared ambient sensory preferences among members of social groups can help to build a QoMEX preference profile for new group members.

CONCLUSIONS AND FUTURE WORK

This paper introduced a framework to evaluate time-varying user emotional responses to ambient quality of multimedia experience. In the proposed framework, users explicitly indicate a response through tagging activities whilst their EEG biosignal emotional feedback is continuously monitored in-between user tagging events with emotional states correlated with media content and tags. The proposed framework also supports the collection of user sensory preferences for ambient effects personalisation, in addition to exploring sensory preferences shared among social groups. The authors are currently implementing the framework with readily available consumer EEG headsets, amBX sensory effects hardware and client/server user tagging software tools; a number of user evaluation experiments will shortly follow.

REFERENCES

1. Chanel, G., Kronegg, J., Grandjean, D., Pun, T. Emotion assessment: Arousal evaluation using EEGs and peripheral physiological signals. in *Proc. Int. W'shop on Multimedia Content Representation, Classification and Security* (Istanbul, Turkey, 11-13 Sept. 2006), 530–537.
2. Choppin, A. EEG-based human interface for disabled individuals: Emotion expression with neural networks, *Master's thesis*, Tokyo Institute of Technology, 2000.
3. Davis, S., Burnett, I., Ritz, C. Using Social Networking and Collections to Enable Video Semantics Acquisition, *IEEE Multimedia*, 16, 4 (Oct.-Dec. 2009) 52-61.
4. Davis, S., Cheng, E., Burnett, I., Ritz, C. Multimedia Adaptation Based on Semantics from Social Network Users Interacting with Media, in *Proc. 2nd W'shop on QoMEX* (Trondheim, Norway, 21-23 June 2010), IEEE.
5. Golder, S.A., Huberman, B.A. Usage patterns of collaborative tagging systems, *Journal of Information Science*, 32 (Apr. 2006), 198-208.
6. ISO/IEC 23005-3 FCD: Information technology – Media context and control – Sensory information. ISO/IEC JTC 1/SC 29/WG 11/N10987, Xi'an, China, Oct. 2009.
7. Kierkels, J. J. M., Soleymani, M., Pun, T. Queries and tags in affect-based multimedia retrieval, in *Proc. IEEE Int. Conf. on Multimedia and Expo* (New York, USA, 28 June-3 July 2009), IEEE, 1436-1439.
8. Lang, P. J. The Emotion Probe: Studies of Motivation and Attention. *American Psychologist*, 50, 5 (1995), 372–385.
9. Soleymani, M., Chanel, G., Kierkels, J. J. M., Pun, T. Affective Characterization of Movie Scenes Based on Multimedia Content Analysis and User's Physiological Emotional Responses, in *Proc. 10th IEEE Int. Symp. on Multimedia* (Berkeley, California, 15-17 Dec. 2008), IEEE, 228-235.
10. Waltl, M., Timmerer, C., Hellwagner, H., "A Test-Bed for Quality of Multimedia Experience Evaluation of Sensory Effects", in *Proc. 1st International Workshop on Quality of Multimedia Experience (QoMEX 2009)*, San Diego, USA, July 29-31, 2009.
11. Waltl, M., Timmerer, C., Hellwagner. Improving the Quality of Multimedia Experience Through Sensory Effects, in *Proc. 2nd W'shop on QoMEX 2010* (Trondheim, Norway, 21-23 June 2010), IEEE, 124-129.
12. Waltl, M., Timmerer, C., Hellwagner, H. Increasing the User Experience of Multimedia Presentations with Sensory Effects, in *Proc. WIAMIS 2010* (Desenzano del Garda, Italy, 12-14 April 2010), IEEE, 1-4.
13. Yazdani, A., Lee, J.-S., Ebrahimi, T. Implicit emotional tagging of multimedia using EEG signals and brain computer interface, in *Proc. ACM Multimedia W'shop on Social Media* (Beijing, China, 14-19 Oct. 2009).