Real Time Labeling of Affect in Music Using the AffectButton

Joost Broekens
MMI, TU Delft
Mekeweg 4, Delft
Netherlands
+31 (0)15 - 2786333
joost.broekens@gmail.com

Anne Pronker
Stedelijk Gymnasium Leiden
Fruinlaan 15, Leiden
Netherlands
+31(0)71 5121365
anne_pollewop@hotmail.com

Marian Neuteboom
Stedelijk Gymnasium Leiden
Fruinlaan 15, Leiden
Netherlands
+31(0)71 5121365
marian.neuteboom@tiscali.nl

ABSTRACT
Valid, reliable and quick measurement of emotion and affect is an important challenge for the use of emotion and affect in human-technology interaction. Emotion and affect can be measured in two different ways: explicit, the user is asked for feedback, and implicit, signals from the users are automatically translated to affective and emotional meaning (affect recognition). Here we focus on explicit affective feedback. More specifically, we focus on the evaluation of an affect measurement tool called the AffectButton. Previous evaluation studies [2] showed that the AffectButton enables users to give affective feedback in a low-effort, reliable and valid way. In this paper we report a study involving real-time affective labeling of movie music by primarily high school students, i.e., a realistic domain with mainstream users. Our results show that (a) users (n=21) are able to use the AffectButton in real time while listening to the music; (b) the labeling very-well follows the changes in the music and gives insight into the different affective dimensions of the music, and; (c) objective music properties correlate to these affective dimensions replicating findings of others. This provides evidence that the AffectButton is a viable affect measurement tool usable by non-expert users in real-time realistic domains.

Categories and Subject Descriptors
H5.2. [Information interfaces and presentation]: User Interfaces; interaction styles.

General Terms
Measurement, Human Factors.

Keywords
Affective feedback, music, user study, AffectButton.

1. INTRODUCTION
Emotions play an important role in our lives. Emotion and affect influence how we think, what we like as well as how we react to and reason about events. As such, emotions also play a role in how we interact with technology. In a variety of domains, measurement of emotion and affect has become a key issue. We briefly discuss several of these domains to underscore the importance of this issue.

First, in the development of so-called Living Labs, the aim is to better design products for people by capturing users’ experience with products during actual usage. This is a markedly different approach from the traditional controlled laboratory experiment. As emotion is a key aspect of experience [17], it is important to capture a user’s emotion in a valid, reliable and quick way.

Second, social software explicitly mediates or uses human-to-human communication. Network sites such as Facebook facilitate human contact. Blogs facilitate topic-based communication and discussion, and wiki’s facilitate collaborative creation. As emotion is important in human-to-human communication, measuring and presenting emotion is an important field of study in social software [6].

Third, certain applications can make explicit use of human affect. For example, content sharing systems such as Tribler [24] facilitate human-to-human content recommendations in a social way. Recommendations are made based on collaborative filtering, a user-to-user content matching mechanism based on overlapping interests in content items. Such mechanisms in essence try to capture the attitude a user has about a content item, and as such affective feedback could be used to enhance the feedback given to the system by the user. Other systems that could benefit from a valid, easy-to-use and reliable emotion measurement tool include e-learning and pervasive computing systems [10] as well as training simulators and tutoring systems [9].

Affective user feedback can be divided in two categories: explicit and implicit. Implicit methods sense the behavior of the user (face, body, heart rate, mouse-movement, etc.) and deduce an emotion. Explicit methods ask the user to input affective feedback directly. There is a long history of emotion measurement in the fields of psychology and affective computing. Typical explicit instruments used are self-report questionnaires, adjectives [27], and images [1, 13]. In addition, a set of implicit mechanisms exist that deduce affect and emotion from different physiological modalities such as heart rate, skin conductance, facial expression and voice (see [21, 22, 30] for overviews). As classical psychological feedback tools usually consist of paper or digitized
questions, these are difficult to embed in an interface or product for intuitive and quick affect feedback. On the other hand such mechanisms are often well-validated, and therefore offer meaningful feedback. Automatic affect recognition typically works well in laboratory environments, but currently fails when used as a generic mechanism to capture affect due to the large number of confounding factors in real world usage [30]. In essence this means that the state of the art is that classical measures need to be updated to digital, easier-to-use forms, until affect recognition has advanced to broad usage with cheap equipment. It is with this in mind that we propose and evaluate our solution to affect measurement. Others have also proposed explicit tools to capture user affect. We now briefly review these.

Several approaches exist towards explicit emotion feedback by means of digital systems [7, 11, 28]. Typically, these approaches have a fundamental tradeoff between precision and measurement speed/ease of use [11]. This means that the more detailed the feedback, the more effort involved for the user and therefore the less likely users will adopt the method as a common way of entering affective feedback. In our approach we specifically aim for both precision and speed/ease of use.

Further, many methods ask a user to input categorical (discrete) emotions, such as happy, sad, angry, jealous, with or without intensity. The benefit is that discrete emotions are easy to interpret by the user giving the feedback and the person or system interpreting the feedback. The drawback is that mixed emotions are difficult to express as there is no logical “emotional continuum” between categories, as well as cost more effort from the user to enter as feedback (feedback on two emotion categories requires more interaction than feedback on one emotion category).

Our approach focuses on continuous (vs. discrete) feedback on core affective dimensions (see next section). Affective dimensions define an affect space. The affective feedback consists of a point in this space. The benefit of a dimensional approach is that it enables graded (by which we mean of different intensities) and mixed affective feedback that is easy to interpret numerically by a computing device. Factor-based feedback enables numerical operations such as averaging of affective feedback, distribution densities, and the like. This is important for computing systems, for example in recommender systems because this allows the computation of average affective profiles for pieces of content or affective distance between user profiles.

Other approaches are able to, in principle, extract detailed affective information in a non-invasive manner but involve the use of human observers to evaluate the feedback and are more focused on measuring human emotion during the process of product development [5, 14]. Such approaches give detailed affective feedback but are not suitable for online affect measurement: i.e., affect measurement aimed at getting real-time affective feedback in a format that is usable by a computing device. Our goal is the latter. We aim for a simple, easy to use tool that can be used for affective feedback without the need for human observer intervention.

Several methods exist that are based on the Self-Assessment Manikins (SAM) [2]. Key in these methods is that they measure emotion factors (pleasure, arousal and dominance) directly and separately. For each factor the user selects a picture from a set of pictures showing emotional faces that express different intensities for that factor. Although the SAM method is by now well-validated, a potentially unresolved limitation is that users must understand the three emotion dimensions before they can use the method. A second drawback is that the method takes up a considerable amount of screen space, and is thus difficult to embed in an interface. This is particularly relevant for mobile devices, or embedded technology with small screens. The AffectButton aims to address both issues. In essence we focus on a usable digital form of the SAM tool, but one that requires limited user interaction and no user interpretation of the three dimensions. As such our work is related to FEELRACE [4], with two key differences. First, the AffectButton enables affective feedback on three dimensions analogous to SAM, while FEELRACE uses two (activation/arousal and evaluation/valence). Second, the AffectButton shows the user a facial expression that represents the values on three dimensions, eliminating the need for users to understand the semantics of, and relation between affective dimensions.

To summarize our motivation for developing the AffectButton; we aim for a measurement method for affect that is thoroughly evaluated with respect to validity, reliability, and ease of use, that is easy to embed in an interface, of which the data is machine interpretable, and that can be used to measure continuous values of affect. A related advantage we see of having such a tool is that it can be used for psychological studies where affect needs to be measured. Its usage would be the same as paper or digital forms of SAM, but quicker, simpler to embed in an interface and easier to understand. In an earlier study the validity (i.e., do I measure what I want) and reliability (i.e., is my measure precise) of the AffectButton has been shown [2].

In this paper we address a different question. We investigate the usability of the AffectButton. In particular, we investigate if the AffectButton can be used in a realistic domain in real time with mainstream users. We report our findings from a study involving real-time affective labeling of movie music by primarily high school students (n=21), indeed a realistic domain with mainstream users.

2. AFFECTBUTTON: PLEASURE-AROUSAL-DOMINANCE FEEDBACK

Emotion is a complex topic, and agreement on one solid definition does not really exist. We do not detail the topic of emotion, as many excellent works have been published from different perspectives (see, e.g., [8, 15, 18-21, 25, 26, 29]). We explain how to interpret what the AffectButton measures in relation to emotion, and above mentioned references can be seen as “collective source”.

Typically, affect refers to the underlying core of emotion, mood, and affective attitude towards persons and things. Emotion, mood and affective attitude are different but strongly related and influence each other. In general, emotion is related to facial expression, feeling, cognitive processing, physiological change and action readiness. Furthermore, emotion refers to a short but intense episode that, in addition to the previously mentioned aspects such as facial expressions, is characterized by “attributed affect to a causal factor”. An emotion is a noticeable and often powerful experience. For example, I feel (notice I am) happy about seeing an old friend. In contrast, mood refers to a silent presence of moderate levels of affect. I can feel frustrated for half
a day without knowing why. Mood is rarely (consciously) attributed to a causal factor. Affective attitude refers to how one generally feels about something or someone, not specifically because of that thing or person. For example, I like popular science books. To complicate matters a little, affect is also used as commonplace term for everything that has to do with the above.

There are several theoretical views on how to think about emotion. The following categorization that uses two axes is particularly useful. The first axis defines the level of abstraction at which emotion is studied: social, psychological, biological, and physiological. The second axis defines the three main ways emotion is represented: categories of emotion, components that form an emotion, and principle factors. For example, Ekman’s well-known six basic emotions are categorical (fear, anger, happiness, etc.). Cognitive appraisal theories are componential, as these describe emotion as a combination of the activation of different sub processes (e.g., evaluation of an event in terms of novelty and goal conduciveness). Finally, emotion and affect can be described using continuous factors (e.g., valence, arousal and control).

Disregarding these different views, many emotion researchers agree upon two core affective factors that are useful to describe a mood, emotion or attitude: valence and arousal. The difference in opinion is not so much about these factors but about how to interpret what they are. Are these factors the emotion, do they represent something real in the brain and if so which brain areas are involved, are they independent (orthogonal), are they artifacts of statistical analysis of many factors, etc.

One of the goals of our research is to have a measurement device to input mixed and graded intensity affective feedback, regardless of what that feedback relates to (attitude, emotion, and mood). We have chosen for a dimensional approach that relates to core affect as explained above. Because a substantial number of emotions cannot be represented clearly as points in this 2D affective space, we have used a related theory as basis for the AffectButton. This theory uses three factors, pleasure (i.e., valence), arousal and dominance (i.e. control) [16, 19] (PAD), and is more expressive a model for affect but less generic as it is unclear if the dominance dimension is a fundamental element of core affect. Valence relates to the positiveness vs. negativeness of affect; arousal to the activation, and dominance to whether the environment is imposing influence over us or the inverse. The PAD factor-based theory states that every object/ emotion/mood/etc has a mapping to a point in PAD space. The reverse mapping is not the case, i.e., not every point has a unique emotion attached to it. The mapping is many (object/mood/emotion) to one (PAD triplet).

The AffectButton was developed as a simple button (implemented in Java) and is available from http://www.joostbroekens.com. It is scalable in size and is a square (Fig. 1). It does not unfold or pop-out, and can therefore be considered a static element in an interface. The button itself renders a face that changes directly according to the mouse position in the button. The mouse x and y within the button define the values on the dimensions Pleasure, Dominance and Arousal respectively. Resulting values are numbers between -1 and 1. The user can therefore select a large range of affective values from the whole PAD spectrum by moving the mouse within the button and clicking. Affective values are represented by the rendered facial expressions, so the user selects an emotional expression by clicking the button. Based on the PAD coordinates, the face displayed is interpolated between nine prototypical expressions. Therefore, a user can enter mixed emotions (e.g., confused) as well as low and high intensity prototypical ones (e.g., little happy, elated). The 3D PAD space has been mapped to a 2D interaction, resulting in a slight reduction in resolution, particularly on the arousal dimension. For more detail on the construction, workings and design choices of the AffectButton, see [2]. A limitation of the design of the AffectButton is that emotions that have a detailed (cognitive) meaning, such as jealous, envy, hope, and caution, cannot be uniquely selected. However, this is a limitation of the underlying PAD space, and the same holds for all approaches using PAD (e.g., SAM). As such we consider the AffectButton a tool to enter affect (as is SAM), not emotion per se.

![Figure 1. The AffectButton and its extreme affective states: elated (PAD=1,1,1), afraid (-1,1,-1), surprised (1,1,-1), sad (-1,-1,-1), angry (-1,1,1), relaxed (1,-1,1), content (1,-1,1), frustrated (-1,-1,1). Labels are exemplary. Note that the AffectButton allows for continuous input in the PAD space.](image)

### 3. EXPERIMENT SETUP

To investigate usability of the AffectButton in a realistic domain we set up an experiment in which users had to listen to 3 different pieces of music. The music consisted of three songs from the movie Pirates of the Caribbean:

- Walk the plank  
  [http://www.youtube.com/watch?v=ICvzkd_Nsg0](http://www.youtube.com/watch?v=ICvzkd_Nsg0)
- Underwater march  
  [http://www.youtube.com/watch?v=bh_h-6cnD1M](http://www.youtube.com/watch?v=bh_h-6cnD1M)
- Blood rituals  
  [http://www.youtube.com/watch?v=hpBwvXMpA4](http://www.youtube.com/watch?v=hpBwvXMpA4)

The Youtube links for these songs are given, so that readers can listen to the songs and easily appreciate the qualitative analysis presented in the next section. Users were explained how to use the AffectButton (but not how it works nor the meaning of affective dimensions), and could familiarize themselves with it for several minutes. No problems were reported with understanding its use. Users were explained that they were about to listen to 3 songs, and that they could use the AffectButton whenever they wanted to indicate the affective content of (explained as emotion of) the music they heard, but in any case whenever they thought it changed significantly. Subsequently, users filled in a short questionnaire (including demographics). Then users listened to each of the songs, and rated the music in real time. Each song was listened to in full. The application used by the subjects was a standalone implementation of the AffectButton, downloadable from [http://www.joostbroekens.com](http://www.joostbroekens.com) (Figure 2). In total 21 subjects participated (18 high school students, 3 parents, 13 female, average age high school students =16.7, stdev 1.5, average age parents=54.6, stdev=3.8). The feedback was time stamped and logged to a central server. This resulted in an affect trace for each participant and each song for the duration of a song for the three affective dimensions pleasure, arousal and dominance.
To analyze the affect trace, we employed two strategies. First, for each song separately, we binned all measurements for all participants together. We sorted the measurements based on time stamp, resulting in three “overall” affect traces, one for each song. Then, for each affective dimension, song and second in that song, we created average scores based on a window of the past 20 measurements relative to the second in the song. Figures 3a-c present the resulting plots for each song and each dimension. These graphs are used to qualitatively analyze the affect traces. To appreciate this analysis, it is best to listen to the songs and read the affect graphs plus textual explanation of what happens.

Second, to quantitatively analyze the affect traces, one of us rated all songs manually on a three point scale, each 5 seconds based on 5 objective music properties. As we had only one rater, we lack inter-rater consistency data. These ratings were created before the experiment and are coupled to all affective feedback based on the timestamps of the feedback after the experiment. We used:

- **Rhythm presence** (1=no clear presence, 2=normal, 3=a lot)
- **Orchestration** (1 little instruments, 2=normal, 3=a lot)
- **Tonality** (1=moves up in key, 2=stable, 3=moves down)
- **Volume** (1=soft, 2=normal, 3=loud)
- **Tempo** (1=long notes, 2=normal, 3=short notes).

Figure 2. AffectButton application used to score music

4. RESULTS AND DISCUSSION

As mentioned, we analyze our results in two steps. First we present a qualitative analysis, then a quantitative one. In this paper we present only the complete qualitative analysis of the first and second song (walk the plank and underwater march). The last song is left for the reader as a “challenge”, as the affect traces are quite clean and easy to read if one understands the three affect dimensions (Figure 3). For the analysis, please remind that valence relates to the positiveness vs negativeness of affect; arousal to the activation, and dominance to whether the environment is imposing influence over us or the inverse. Please also keep in mind that users scored as a reaction to the music, so aspects of the music will be seen in the graph with a delay of about 7 seconds (see [2] for the average time needed to use the AffectButton for scoring). We now provide the affective time line (in seconds in the song, not the graph) for underwater march:

- **0-17:** dark, gloomy, sudden climax at 17, characterized by low pleasure, neutral dominance, and arousal peak at 17 (a bit later in graph).
- **17-75:** sadness, minor key, death, slow, characterized by low pleasure, arousal and dominance. Please note the crescent of pleasure and dominance before second 50, where music seems to go towards a more light tone, but doesn’t, as reflected by the subsequent dip in pleasure. Also note that the dominance values reflect the fact that the music is building up to something.
- **75-93:** change of key, brighter music, still calm, end with silence, tension, characterized by a small peak in pleasure before a dip, arousal is low, and dominance is slightly below neutral.
- **93-110:** last “wave” working out of the minor key, ending with clear transition, characterized by low pleasure, and a high arousal for anticipation.
- **110-130:** relief, a bit peaceful, characterized by high dominance, high pleasure, low arousal.
- **130-160:** melancholic, pausing, characterized by decreasing pleasure and dominance, low arousal.
- **160-190:** working towards a peaceful ending, closure, relief, still calm, characterized by high dominance, high pleasure, low arousal, but then....
- **190-253:** the march, first gloomy (decreasing pleasure), then rhythmic marching, heavy, dark, then very active, characterized by low pleasure, high dominance and increased arousal.

Our qualitative analysis clearly shows an affective fingerprint of several key aspects of the music, including minor-major changes in key and working towards a climax. Interesting to see is the independent behavior of dominance and pleasure (at the end of both songs), showing that it is essential to use dominance as an affective feedback dimension. Also, subjects labeled affect in the music “as experienced”, indicated, e.g., by high dominance for angry, heavy, dark parts (i.e., the music feels dominant, instead of is dominant). These findings indicate that the AffectButton can be used to measure user affect in a realistic setting in real-time and results in well-interpretable data.

To provide a more objective, quantitative analysis, we have performed a MANOVA for each objective property with the values of that property as independent variables, and pleasure, arousal and dominance as dependent variables. Statistical unit was a PAD tuple and corresponding objective property values in the song at the time the PAD feedback was given by the user. We tested for each objective music property (mentioned above) if it influences affective dimensions (a well-known fact in consumer psychology, see, e.g. [12]). We did not test for interaction effects between properties. All of the effects were significant with p<0.001 (Rhythm: F(6, 1700)=47, Orchestration: F(6, 1700)=27, Tonality: F(6, 1700)=4.8, Volume: F(6,1700)=18, Tempo: F(6,
Effects of music properties on individual affective dimensions (Graphs in Figure 4) was strongly significant in all cases with $p<0.001$ except for orchestration and pleasure, and tonality and arousal (not significant), and tonality and dominance, and volume and pleasure ($p<0.02$). Most of the effect sizes are small, except those on dominance (table and bar charts in Figure 4). Our results partially replicate results from [12] who found significant relations between texture, tempo and tonality on the one hand and pleasure and arousal on the other. Like we, they also failed to find a significant relation between tonality and arousal. We did not find a relation between orchestration and pleasure, while [12] did. This difference can be explained by our stimuli. Both the positive and negative parts of our songs have more orchestration.

Figure 3a-c. Affect trace of walk the plank, underwater march and blood rituals respectively (seconds=x, PAD value=y).

Figure 4. Graphs show means and standard errors for each affective dimension per objective property value. Table shows effect sizes of properties on affective dimensions (Pearson r).

In this paper, we focus on real-time use of the AffectButton in a relevant domain, not on the relation between music properties and affect per se. Therefore, the comparison between our findings and those of others is limited. The main aim of the qualitative and quantitative analysis is to show that the AffectButton is usable in a realistic setting and produces valid data. Further, the fact that we have no inter-rater consistency data on the objective music properties is a limitation for our conclusions with regards to the effect sizes. We do not know the amount of “noise” in the objective property ratings. However, given the fact that the affect traces are very clear, we feel confident that the AffectButton itself produces valid data in this particular study. If we relate this to findings in earlier studies with the AffectButton that all reported strong correlations, where expected, between the affective feedback given by users using the AffectButton and other measures [2, 3, 23] (e.g., Likert scales on liking), we conclude that the AffectButton is a viable affect measurement tool.

5. CONCLUSIONS

We have presented experimental results that provide evidence that the AffectButton is a viable affect measurement tool usable by non-expert users in real-time realistic domains. We believe this is an important finding as in many domains there is a need for easy-to-use, valid, and reliable affect measurement tools that can be embedded in human computer interfaces. Future work is planned to objectively compare ease of use of different factor-based affect measurement mechanisms, including the Self-Assessment-
Manikins. Finally, we would like to stress that the AffectButton is downloadable and freely available for research and other non-commercial activities.

6. ACKNOWLEDGEMENTS
This research is supported by the Dutch Technology Foundation STW, division of NWO and the Technology Program of the Ministry of Economic Affairs. It is part of the Pocket Negotiator project VICI-project nr 08075.

7. REFERENCES