

Tune in to your emotions: a robust personalized affective music player

Joris H. Janssen · Egon L. van den Broek ·
Joyce H. D. M. Westerink

Received: 5 August 2010 / Accepted in revised form: 18 August 2011 /

Published online: 7 October 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract The emotional power of music is exploited in a personalized affective music player (AMP) that selects music for mood enhancement. A biosignal approach is used to measure listeners' personal emotional reactions to their own music as input for affective user models. Regression and kernel density estimation are applied to model the physiological changes the music elicits. Using these models, personalized music selections based on an affective goal state can be made. The AMP was validated in real-world trials over the course of several weeks. Results show that our models can cope with noisy situations and handle large inter-individual differences in the music domain. The AMP augments music listening where its techniques enable automated affect guidance. Our approach provides valuable insights for affective computing and user modeling, for which the AMP is a suitable carrier application.

Keywords Mood · Music · Psychophysiology · User modeling ·
Kernel density estimation · Validation · Affective computing

J. H. Janssen (✉)

Human Technology Interaction, Department of Industrial Engineering and Innovation Sciences,
Eindhoven University of Technology, Den Dolech 2, 5600 MB Eindhoven, The Netherlands
e-mail: joris.h.janssen@philips.com

J. H. Janssen · J. H. D. M. Westerink

Department of Brain, Body, and Behavior, Philips Research, High Tech Campus 34,
5656 AE Eindhoven, The Netherlands

E. L. van den Broek

Human Media Interaction, Faculty of Electrical Engineering, Mathematics, and Computer Science,
University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

E. L. van den Broek

Karakter U.C., Radboud University Medical Center Nijmegen, P.O. Box 9101, 6500 HB Nijmegen,
The Netherlands

1 Introduction

Music is intrinsically intertwined with our everyday lives. People use music to get them going in the morning, to relax after work and to make it through a workout. College students use music while studying and brain surgeons perform their most intensive procedures with music in the background. In sum, many people use music to enhance their moods (North and David 2000; North et al. 2004).

Over the last 20 years, new technologies have digitalized our music listening experience. In today's world, we have access to large amounts of music everywhere and all of the time. We listen to music on the road, while working, cleaning the house, or while enjoying a book. Moreover, we prefer different types of music for these different occasions. While cleaning you want to be energized, whereas for enjoying a book, relaxing background music is probably preferable. Having access to large amounts of digital music opens up opportunities for automated intelligent music selection techniques. Some of these techniques have been explored in the recent past (Mandel et al. 2006; Sotiropoulos et al. 2008). For instance, Pandora, Last.fm, or Apple's Genius make it possible to select music automatically, based on content similarity.

Another way to augment the music selection process is by tapping into the emotional power of music. If a music player had emotional intelligence, it could automatically generate playlists that energize you, relax you, or make you happier. In this way, such technology could focus on the affective qualities specific activities require. Moreover, this technology can tune into your mood, fitting the selected music to your current affective state. Aside from hedonic motives, the ability to impact mood is also relevant to one's cognitive performance (Gendolla 2000), and health and well-being (Vaillant 2003). Positive moods increase creativity, improve decision making processes, and enhance social relationships (Baas et al. 2008; Rusting 1998; Clore and Palmer 2009). Moreover, positive moods relieve us from stress that can otherwise have devastating influences on our health and well-being (Gendolla and Brinkman 2005).

The idea of an affective music player (AMP) is not new. Picard (1997) was probably the first to describe it in her seminal work on *affective computing*. The field of affective computing arose from Picard's book and deals with the modeling, recognition, influence, and design of affective experiences (Carberry and de Rosis 2008). Interestingly, many of the challenges encountered in contemporary affective computing come together in the AMP (Picard 2003; Van den Broek et al. 2009). First, a method for assessing a user's affective state is necessary. This method has to be able to unobtrusively measure affective signals while being robust to all the noise encountered in real-world settings. Subsequently, these affective signals have to be modeled so that relevant information can be extracted. This modeling requires dealing with interpersonal differences, ground truth, and integration of different measurement modalities (Van den Broek et al. 2010). Finally, actuator selections need to be made based on these models. All these challenges are not only central to affective computing, they are also at the core of an AMP. So, not only is the AMP a promising application in itself, it is also a useful carrier application to investigate different strategies for affective computing. User modeling plays a central role in this process, allowing researchers to deal with the prominent individual differences in affective computing (de Rosis 2001; Carberry and de Rosis 2008).

In this paper, we take on the challenge of designing and validating a robust real-time AMP. The rest of the paper is organized as follows. In Sect. 2, we describe relevant related work on AMPs. There, we consider the strengths and weaknesses of previous AMPs and explain how we build on this research. In Sect. 3, we introduce six central pillars of our approach, taking into account theory and findings from a range of fields. Our AMP should be able to measure a user's affective responses to music and use this to construct affective user models on which music selection can be based. Models to achieve this are described in Sects. 4 and 5. The music player should work in the real world, and, as such, should be able to deal with the noise involved in real-world sensing and interpretation (Fairclough 2009). Moreover, sensing and recognition mechanisms should be able to distinguish between affective changes caused by the music and affective changes caused by other factors. Therefore, a validation in real-world trials is described in Sect. 6. Finally, we discuss our results, limitations, and their implications.

2 Related work

Throughout the last decade, some AMPs have been proposed. The Body Rest system by Liljedahl et al. (2005) is a biofeedback system that works by adapting the tempo of music to fit one's heart rate (HR). Using this system, users can gain insight into their own feeling state. This system uses a biofeedback approach that changes the music itself (i.e., the tempo of the music), instead of adapting the selection of the music. It is unclear how the system performs in practice and how user's respond to adaptations of the music. Beside the fact that the HR information is presented through the music, there will probably also be an effect of music on the affective state and HR of the user. This complex interaction might make it difficult to use in practice.

The MPTrain system by Oliver and Flores-Mangas (2006) determines HR responses to music while exercising. It uses this information to develop a model of the relationship between music characteristics, exercise level, and HR. Subsequently, it uses these models to select music to direct HR based on a workout program. Observations taken from one user show that the MPTrain system can improve the user's exercise performance. However, this system is specific to exercise, basing music selection on characteristics of the music that relate to exercise performance.

The Affective DJ by Healey et al. (1998) saves skin conductance changes a song elicits to a database. Subsequently, it uses this information to select music that directs to a relaxed or aroused state. Healey et al. acknowledge that "Testing such a system is tricky and time-consuming, as there are a huge number of variables to control for, and the system really needs to have a lot of music and to be worn for a lengthy time before it can provide the planned advantages". As they state themselves, their failure to show a successful performance of the system might have been because of a high repetition of only a few songs. This may have annoyed users. Nonetheless, the overall approach of repeatedly measuring physiological reactions to songs provides promising ground for further development.

The above three examples provide very useful first inquiries in the use of physiological signals for AMPs. We will build on their approach by using physiological

signals to measure the effects of songs. However, as will be elaborated in the following section, we extend this by incorporating more sophisticated physiological signal processing. Instead of using simple averages, we argue for the use of baselining, specific time windows, and the incorporation of the law of initial values (LIV). We expect that this will help us to deal with the noise of real-world measurements, providing better results.

As has become clear from the three examples above, there is a need for thorough empirical testing and validation of these systems. No matter how well-founded and integrated a system may be in theory, the complexities of human emotion and physiological responses make it necessary to rigorously test every system in the real-world (Chin 2001; Healey 2009; Van den Broek et al. 2009). We try to do this by taking a three step approach. First, we gather data from different users over different days during their regular working activities. Second, we use this data to train our user models. Third, and most importantly, we use these trained user models to make music selections and see whether or not these music selections elicit the desired affective state. By taking these steps in a real-world context, we hope to provide an adequate and convincing validation of our AMP.

3 Design considerations

Research on affective technology has been shaped by a range of disciplines. Based on the related work presented in the previous section and insights from psychology, we describe six considerations that are central to our AMP.

3.1 Music is personal

First, musical taste is very personal. For instance, a heavy metal fan will probably differentiate between relaxing and arousing heavy metal music. In contrast, someone who listens mostly to classical music will probably find all heavy metal pieces highly arousing. Many personal factors can be identified that influence the effect of music on affect; for instance, personality (Rentfrow and Gosling 2003), familiarity with the music (Ritossa and Rickard 2004), age (Pelletier 2004), gender (Webster and Weir 2005), and musical preference (Sloboda 2005). In line with this, it has been shown that personally selected music induces affect more strongly than music selected by an experimenter (Rickard 2004). Therefore, models for predicting the emotional effects of music should be personalized (Kim and André 2008).

To deal with the personal aspects of music listening, we take an approach similar to that of Oliver and Kregor-Stickles (2006) and Healey et al. (1998). Every time a user listens to a song, the affective changes the song elicits are calculated and stored. Based on this data, affective changes elicited by a specific song can be predicted. The advantage of such a personal approach is that all models are completely tailored to the user and his or her own personal music library. The user models can, therefore, capture all the personal preferences, associations, and memories a listener has for each specific song.

3.2 Moods are not emotions

So far, we have used the words affect, mood, and emotion interchangeably. There are, however, clear differences between these concepts in the psychological literature, and it is important to take these into account. Affect is often used as the common denominator of mood and emotion (Russell 2003). Emotions are often defined as relatively short-lasting hedonic reactions to specific stimuli in the environment. They result in physiological and behavioral changes to deal with the sudden change in the environment (Prinz 2004). In contrast to emotions, moods (1) are long lasting and change gradually, (2) are not object related, and (3) are often experienced without awareness of their origin (Frijda 1986). Moods can also be accompanied by physiological changes but do not result in direct action tendencies (Beedie et al. 2005). Instead, they influence behavior indirectly (Gendolla 2000). Moods are often operationalized in terms of valence and energy or arousal (Matthews et al. 1990; Wilhelm and Schoebi 2007; Thayer 1989). Valence refers to the pleasantness of the mood ranging from very sad to very happy. Energy (or arousal) ranges from very relaxed to very excited (Thayer 1989).

For our AMP, we are not specifically interested in influencing short-lived emotional changes, but in more gradual changes in the listener's mood. Moreover, it is important for the music not to be too prominent in its effects, as this could potentially decrease task performance (Ophira et al. 2009). Music is often used in the background, while one is doing other tasks. So, we try to avoid strong sudden reactions and instead focus on longer lasting changes that can be induced relatively unconsciously. Although media multitasking typically decreases task performance (Ophira et al. 2009), background music might be the one exception, as it can actually increase task performance (Lesiuk 2005). Given these considerations, we are not interested in emotional effects within a song, but rather in gradual changes over one or more songs. Studies on musical mood induction have shown a period of eight minutes (typically equivalent to two or three songs) to be sufficient to influence mood (Gendolla and Krüsken 2001). So, although music is very dynamic and time-varying in nature, we deal with this issue by averaging over longer time periods. This filters out the short dynamic changes within the music and affective state of the listener and, instead, provides an indication of more gradual changes in the listener's mood.

3.3 Physiological signals can measure affect

Several modalities can be used to measure a user's affective state (Zeng et al. 2009; Van den Broek et al. 2009). The three most common modalities are facial expressions as recorded by a video camera (Pantic and Patras 2006), speech recorded by a microphone (D'Mello et al. 2008), and physiological signals measured through body sensor networks (Hanson et al. 2009). Video capture and speech recording have some clear disadvantages for a music player. People are not always in front of a camera and they tend not to talk when listening to music. In contrast, physiological signals can be measured unobtrusively with wearable sensors incorporated in, for instance, a bracelet (Westerink et al. 2009). In this way, many different signals can be recorded from

the body; for instance, HR, skin conductance level, respiration, and skin temperature (ST). All these physiological signals have been used before to measure affective states (Yannakakis et al. 2008).

For our AMP, we focus on skin temperature as the measure of choice for different reasons. First, several studies have shown that skin temperature is related to valence (Baumgartner et al. 2006; McFarland and Kennison 1989; Rimm-Kaufman and Kagan 1996). Therefore, skin temperature is likely to provide a reliable indication of valence. Second, skin temperature changes gradually. Although it is, therefore, not very suitable for measuring emotions, it makes skin temperature ideal for mood measurement. Gradual changes make skin temperature more robust against sensor displacement and movement artifacts than other emotional indices such as skin conductance or HR. We have plotted two example measurements in Fig. 1. Finally, the sensor is cheap and small; see Fig. 2. Hence, it can easily be incorporated into wearable items such as rings or bracelets.

As said, several studies have shown a relationship between skin temperature and valence (Baumgartner et al. 2006; McFarland and Kennison 1989; Rimm-Kaufman and Kagan 1996). Nonetheless, there have also been studies that have not found such an effect (Rickard 2004). Although not finding a significant effect does not mean

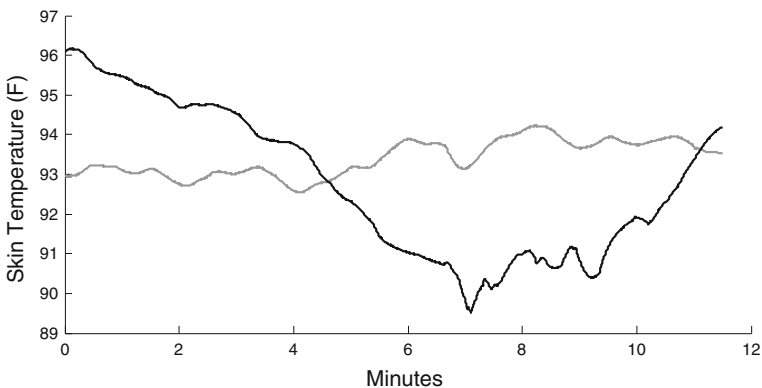
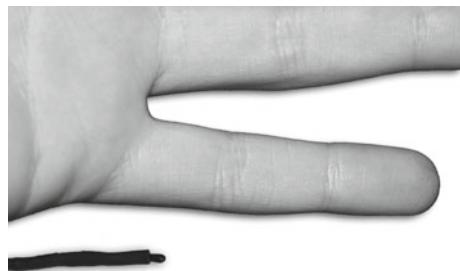


Fig. 1 Examples of two skin temperature traces of 11 min from the same person. The *black trace* decreases at first, and later increases. This suggests a gradual increase in valence with a decrease back to baseline level later on. The *gray trace* fluctuates around a baseline level, which suggests that there were no strong changes in valence during that time

Fig. 2 The skin temperature sensor we used together with a little finger. The tip of the black cable is the actual sensor. In our experiments, the sensor was attached to the proximal phalanx of the non-dominant hand with medical adhesive tape. The small size of the sensor makes it ideal for incorporation into all kinds of wearable devices; e.g., a ring



such an effect does not exist, we wanted to address this apparent discrepancy by running a study ourselves assessing skin temperature responses to very positive and very negative music. This is important for the validation of the music player. In the next paragraphs, we present a study that constituted the first part of a larger experiment. For the sake of clarity and brevity we only report this part, as the rest of that experiment is not relevant for the current work.

Twenty-four participants came into the lab and we attached different physiological sensors to them, including a skin temperature sensor which was taped to the little finger of the non-dominant hand. Participants were randomly assigned to one of two mood induction conditions, in which either very positive music or very negative music was administered. The positive music consisted of three songs that the participants chose from their own music selection, which they believed put them in a very positive mood every time they listened to this music. These songs were shortened to two minutes and 40 seconds from the beginning of the song, using a two-second fade out at the start and end to mask the shortening. They were combined into one audio file of 8 minutes in length. Because participants were unlikely to have very negative music in their own database, we selected a validated negative sad cello piece composed by Hans Zimmer from the movie the House of Spirits for the negative music condition (Gendolla and Krüsken 2001).

The experiment started with a neutral baseline session of eight minutes in which all participants listened to neutral meditation music (i.e., The Temple by Ray Linch). This way we made sure all participants were in the same state at the start of the mood induction. After the baseline session, participants either listened to eight minutes of the very negative music or eight minutes of the very positive music. The UWIST Mood Adjective Checklist (Matthews et al. 1990) was administered directly after the baseline mood induction and after the music mood induction.

Self-reported mood was assessed by averaging the score on the positive adjectives (happy, joyful, contented) with the scores on the reverse coded negative adjectives (sad, frustrated, depressed), yielding a Cronbach's α of 0.80. An independent samples t-test on these scores confirmed that the mood induction was successful ($t(22) = 9.01$; $p < 0.05$; Cohen's $d = 3.81$). Mood was more positive after listening to positive music ($M = 3.46$) than after listening to negative music ($M = 2.19$). Furthermore, individual differences in skin temperature were standardized over the entire experiment by subtracting the mean and dividing by the standard deviation for each participant (Boucsein 1992). Subsequently, an independent samples t-test was done on the means of the skin temperature over the eight minutes of mood induction. The results showed a significant difference ($t(22) = 2.73$; $p < 0.05$; Cohen's $d = 1.16$), with the negative mood induction ($M = 0.92$) resulting in a higher skin temperature than the positive mood induction ($M = 0.21$). To conclude, the two different moods resulted in a significant difference in skin temperature.

3.4 Affective loops

Systems that try to measure and, subsequently, influence affective states have been called affective loops (Fairclough 2009). Most often, an affective loop is defined by

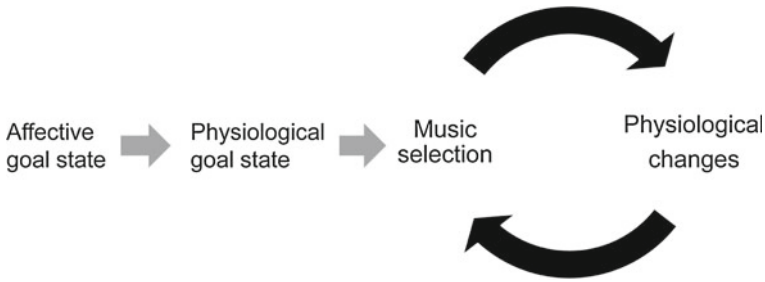


Fig. 3 The AMP's closed loop system. Music is selected based on the physiological changes it elicits and a physiological goal state. The physiological goal state depends on an affective goal state

three steps: (a) infer a user's current affective state from physiology, (b) select an actuator setting, and (c) measure the affective physiological changes the actuator elicits (Fairclough 2009). Subsequently, the loop starts again with step (a) to infer affective changes from the physiological changes. The most significant problem with this approach is the exact relation between affect and physiology (i.e., the first step (a), (Peter and Herbon 2006)), as machine learning studies trying to recognize affect from physiology report disappointing performances (Van den Broek et al. 2009). In fact, it is a common critique on affective computing that automated recognition of affect is difficult, if not impossible (Boehner et al. 2007).

For these reasons, we adopt an alternative affective loop after the ideas of Höök (2009) and Tractinsky (2004). They suggest that there are applications that do not need the inference step from physiology to affect. In our case, we can base music selection on a physiological goal state and model the physiological effects of a song. The physiological goal state can be inferred from psychological studies like the one we described. We know that decreases in skin temperature are related to increases in valence and that increases in skin temperature are related to decreases in valence. Hence, to select music that increases valence, we select music that decreases skin temperature. This approach is depicted in Fig. 3. Our user models should thus contain physiological changes instead of affective changes. This overcomes the problematic inference step from physiology to affect but still allows us to regulate mood in an affective physiological loop.

3.5 Probabilistic models

Physiology is responsive to many psychological and physical influences beside affect (Cacioppo and Tassinary 1990). Because of that, many factors contaminate the affective information in physiological signals. For instance, physical activity, cognitive workload, or simply a cup of coffee all influence our physiological signatures (Bak and Grobbee 1990; Wilson and Russell 2003). Moreover, other emotional influences not originating from music can influence the affective state of the user; for example, an upset colleague or a happy e-mail. As we want to investigate a system that works in the real world, we have no control over all these factors. Hence, the AMP should be able to deal with these different factors.

To deal with these effects we use probabilistic models that naturally deal with noise in the data (Bishop 2006). Assuming that the environmental noise is independent from the music, we will have noise centered around the main effect of the music. Therefore, as the number of data points per song increases, the effect of the song itself will become clearer. This way, we also have an indication of the song's strength compared to the influence of other environmental factors. Songs that do not produce a strong effect are most likely not very strong mood inducers for that user.

3.6 The law of initial values

The effect of a stimulus on physiological change depends on the physiological level before stimulus onset. This effect is called the LIV (Geenen and van de Vijver 1993; Wilder 1967). For instance, when skin temperature is high, it tends to decrease, whereas when it is low it tends to increase. In turn, a stimulus that normally increases skin temperature, might, when skin temperature is already high, only keep it at the same high level. This forms a challenge for our AMP, as we will not be able to control the prestimulus level in real-world settings.

To accommodate this, we model the effect of the LIV as a linear regression line in our AMP. This linear model describes the relationship between the initial value before the stimulus and the change the stimulus elicits. Such a model can be used to correct any measured value based on the natural change expected by the prestimulus level. Resulting corrected values give an indication of the effect of the stimulus (i.e., a song), independent of the prestimulus physiological state.

The six design considerations above form the basis for the empirical development of the music player. From that, we constructed the global architecture of our system. This architecture is depicted in Fig. 4 and will be explained in detail in Sects. 4 and 5. The empirical development of this system was conducted in three steps. First, we gathered data to be used to construct the user models, as is described in the next section. Second, the gathered data was used to train the user models. Third, we use the trained models to select different types of music and test whether or not these music selections were successful in inducing a desired mood.

4 Step 1: data gathering

4.1 Participants and materials

Three male volunteers (aged 22, 26, and 27) participated. All participants signed an informed consent form. Although three participants may seem like a small number, we were interested in developing a different user model for each user individually. We did not want to construct a generic model that would be the same for all users. Therefore, we were not very interested in the between-person variance that is often studied over large groups to estimate population averages. We wanted to obtain reliable estimates for each individual. So, instead of gathering a few data points for many individuals, we put our effort into gathering many data points over many sessions for a smaller

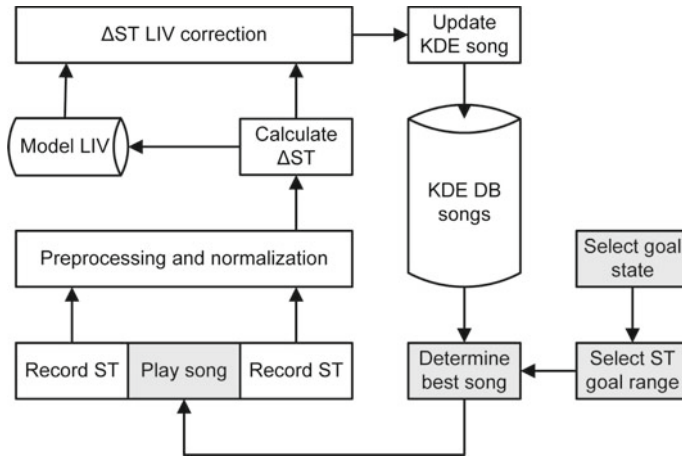


Fig. 4 Architecture of our system, with each *box* depicting a different process. The *white boxes* form the personalization part of the system. The *gray boxes* form the music selection part of the system. For each song that is played, ST is measured, preprocessed, and normalized. Subsequently, the change in skin temperature Δ ST the song elicits is calculated and corrected for the LIV. In addition, the change the song elicits is added to the LIV model that contains all the skin temperature changes measured over all the songs so far. The corrected skin temperature change is used to update a KDE of that song, as stored in the KDE database (DB) of the songs. Music can be selected by calculating the probability of each song that it directs to a certain skin temperature goal state. This skin temperature goal state is based on an affective goal state

number of individuals. This way, we got a good sense of the within-person variance for each of the three participants and were able to construct person-specific user models.

The participants rated 400 randomly selected songs from their own music library on a 5-point valence scale expressing how they expected the songs made them feel, ranging from very negative (-2) to very positive ($+2$). From the ratings, nine positive songs ($+2$), nine negative songs (-2), and nine neutral songs (0) were randomly selected. This resulted in 27 songs per participant for which the physiological mood data was gathered.

Each song was listened to nine times, over nine different sessions. In each session, every song was played exactly once. For each session, we used a different trigram balanced order (similar to Wagenaar 1969) based on the feeling ratings. A trigram balanced order is, first of all, completely counterbalanced. Furthermore, the type of the two items preceding the third item is also counterbalanced. Hence, for each feeling (positive / neutral / negative), all nine combinations of two feelings preceded every feeling once in an order. This gives a list of 27 (i.e., 3^3) songs, which required two additional songs to complete the beginning of the balancing scheme. Finally, the nine different songs per feeling were balanced over the nine different slots per feeling for nine different sessions; consequently, every song was positioned in every slot once.

The data gathering sessions were conducted during regular working activities, including writing and data analysis. All three participants conducted the experiment at their own desk at their work. They all shared a room with at least four other researchers. Hence, the data gathered contained various types of noise; for instance, noise generated by physical movements, joking colleagues, or disrupting emails. This is

important, as we want our AMP to be able to deal with these various sources of noise. The experimentation software ran on an independent computer, so that it did not confound participants' working activities. The songs were presented through a Philips SBC HP400 headphone.

The NeXus-10 apparatus of Mind Media b.v. was used for the measurement of skin temperature. The skin temperature sensor was connected to a portable device, using coated cables that provide an active shield precluding movement artifacts and noise in the signal. The skin temperature sensor was attached with medical adhesive tape to the proximal phalanx of the little finger of the non-dominant hand and could measure changes up to 0.001°C in a range of $10\text{--}40^{\circ}\text{C}$. The signal was sampled at 128 Hz. The data was saved onto a computer, using a wireless bluetooth connection. This made it possible for the participants to walk around while wearing the sensor equipment.

4.2 Procedure

For every participant, the data gathering consisted of nine sessions, on nine different days. The participants decided themselves when to run a session. They set the volume of the audio before the start of the first session and kept it the same throughout the rest of the experiment.

At the start of a session, the participants connected themselves to the skin temperature sensor, checked the signal, and started the experiment. Sometimes the participants had to break the experiment; for instance, because they were interrupted by a colleague or had to go to the restroom. Then, the song currently playing was positioned at the end of the playlist and, whenever they continued, the last minute of this song was played before continuing to the next song, so as to provide normalization data needed for data preprocessing. The complete data gathering took about one month per participant.

4.3 Data preprocessing

To be able to calculate the skin temperature changes, a number of preprocessing steps were performed. For every song k , in every session n , the mean skin temperature of the last minute of the song, denoted by x_{kn} , was extracted. This mean was standardized over the session using

$$z_{kn} = \frac{x_{kn} - \mu_n}{\sigma_n}, \quad (1)$$

where μ_n is the mean and σ_n the standard deviation over all songs of session n . This has proved to be a successful method for standardizing physiological signals (Boucsein 1992). Next, delta scores Δz_{kn} were computed that indicated the effect of song k in session n on the physiology, by using

$$\Delta z_{kn} = z_{kn} - z_{(k-1)n}, \quad (2)$$

with $k \geq 2$. In other words, the average skin temperature over one minute before the song was subtracted from the average skin temperature over the last minute of the song. These delta scores, describing the skin temperature change a song elicited, were used as input for our user models.

5 Step 2: personalization

The AMP was personalized, using the preprocessed gathered data. A different user model was learned for each participant. The personalized parts of the AMP are constituted by a personal model of the LIV and personal probability distributions for the effects of every song. These will be described in the next sections.

5.1 Law of initial values

The delta score Δz_{kn} of a stimulus depends on the prestimulus level $z_{(k-1)n}$ (Wilder 1967; Geenen and van de Vijver 1993). When the prestimulus level is high, the delta score tends to decrease, whereas, when the prestimulus level is low the delta score tends to increase. In turn, the further away the prestimulus level is from the neutral level, the stronger the effect of the song should be to counter the effect of the prestimulus-to-delta relationship. To extract the prestimulus effect from the delta score, a regression line was used to model this relation:

$$y(z) = w_1 z + w_0, \quad (3)$$

where w_0 and w_1 are the parameters of the regression line and $y(z)$ is the predicted change value based on the prestimulus value z . The parameters w_0 and w_1 were estimated based on all data for each participant. Relatively strong correlations r between $z_{(k-1)n}$ and Δz_{kn} were indeed found for each participant: 0.40, 0.59, and 0.35. Although w_1 was always < 0 , in line with the LIV, it differed per person: -0.30 , -0.68 , and -0.24 . We, thus, estimated the regression line over the different sessions of each person individually.

After the regression was established, corrected delta scores $\Delta' z_{kn}$ were computed as follows (see also Fig. 5):

$$\Delta' z_{kn} = \Delta z_{kn} - y(z_{(k-1)n}). \quad (4)$$

These delta scores, describing the skin temperature change a song elicited, formed the input for the personalized probability distributions.

5.2 Personal probability distributions

After extracting the effect of the prestimulus level from each delta score, we wanted to construct a probabilistic model of the corrected delta scores for each song. For this, we treated the delta score each song elicited as a random variable. A random variable

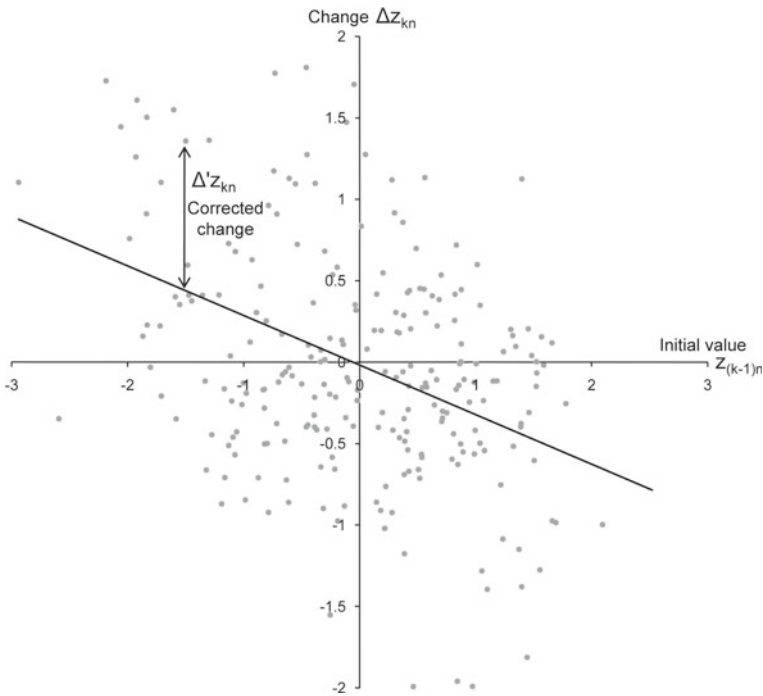


Fig. 5 The effect of a stimulus on physiological change depends on the physiological level before stimulus onset. This LIV is modeled by a *linear regression line*. The *gray dots* represent the data points measured and the *black line* depicts the *regression line*. The variables are defined in Sect. 5

can be described completely by its probability density function (pdf). This positive real-valued function integrates to 1 and can be used to exploit important characteristics of the variable; for instance, mean, quantiles, or power spectral density (Heinz and Seeger 2008). Moreover, the pdf naturally deals with uncertainty in the data. Hence, it makes sense to describe the physiological change a song elicits by a pdf over $\Delta'z$. This is the dimension on which corrected delta scores $\Delta'z_{kn}$ lie. However, the pdfs are unknown and we only have a limited number of observations of $\Delta'z$, so we have to estimate the pdf.

A well-established approach for estimating a pdf over observations is kernel density estimation (Silverman 1986). Kernel density estimates (KDEs) are unsupervised and nonparametric; in other words, they make no a priori assumptions of the underlying distribution and can approximate any distribution. In addition, KDEs are asymptotically unbiased: they are unbiased when the number of sample points tends to infinity. Hence, the more sample points, the better the estimation of the pdf will be.

For every song k , the KDE contains a radial kernel function $K(\Delta'z | \Delta'z_{kn}, h_k)$, with precision h_k , around all N_k measured points $\Delta'z_{kn}$ of this song. The KDE averages over all these N_k kernels:

$$p_k(\Delta'z) = \frac{1}{N_k} \sum_{n=1}^{N_k} K(\Delta'z | \Delta'z_{kn}, h_k). \tag{5}$$

This yields the pdf $p_k(\Delta'z)$ over $\Delta'z$ for song k .

The selection of the precision h is important: when h is chosen too large, the probability distribution is over smoothed; yet, when h is chosen too small, the probability distribution is under smoothed. Various methods have been proposed to calculate an accurate h (see (Turlach 1993), for a review). To calculate h , we adopt:

$$h_k = 1.06 \cdot \min \left(\sigma_k, \frac{R_k}{1.34} \right) \cdot N_k^{-\frac{1}{5}}, \quad (6)$$

where R_k is the interquartile range and σ_k is the standard deviation of the corrected delta scores $\Delta'z_{kn}$ of song k . This method was introduced by Härdle (1991), and combines a computationally efficient approach with robustness against outliers.

A radial Gaussian kernel was used, where the mean is $\Delta'z_{kn}$ and the standard deviation is h_k . The Gaussian is often employed for its analytical properties. Moreover, it provides a smooth distribution (Heinz and Seeger 2008):

$$K(\Delta'z \mid \Delta'z_{kn}, h_k) = \frac{1}{h_k \sqrt{2\pi}} \exp \left(-\frac{(\Delta'z - \Delta'z_{kn})^2}{2h_k^2} \right) \quad (7)$$

5.3 Music selection

Music can be selected on the basis of the probabilities calculated from the KDEs. First of all, the physiological signal can be directed to one of the extreme values. In that case, music selection entails calculating the probability of each song in the increasing $[0, \infty)$ or decreasing $(-\infty, 0]$ range and selecting the song with the highest probability. As an extension of this, a neutral range can also be defined (see Fig. 6). As listeners habituate to stimuli, in practice, constraints such as the variety of songs also play a role that needs to be taken into account.

Sometimes, the optimal affective state is not reached at either one of the extremes of a physiological measure. For instance, when trying to focus on a task, music that is too relaxing will lower the working spirit, whereas music that is too arousing might distract one from the task (Csíkszentmihályi 1990; Kowal and Fortier 1999). In those cases, the physiological models should also allow direction towards a specific point. To do this, first, the necessary change Δz has to be calculated, taking into account the prestimulus level and the LIV model; see Fig. 5. When this Δz is known, a level of precision p has to be defined that together with Δz creates the probability interval of the song: $[\Delta z - p, \Delta z + p]$. Then, music selection entails selecting the song with the highest probability in this interval. This requires specific music for each interval. If this music is available, the KDEs do allow selection of the song with the highest probability in a specific interval.

These different approaches show the versatility with which KDEs can be used. Comparing this to the approaches described in related work shows two fundamental differences. First of all, other approaches typically use means to estimate the effect of a song (e.g., Healey et al. 1998). Although means can be informative, they are very sensitive to outliers that are likely to be encountered in real-world data. Furthermore, they only provide a very rough estimate of the song. By employing probability

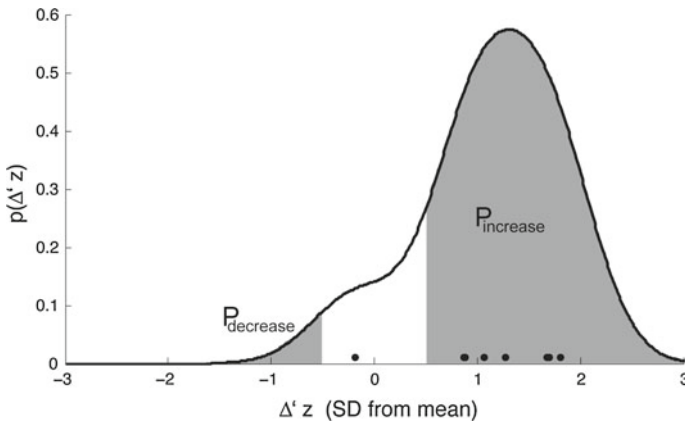


Fig. 6 An example of a KDE of one song. The *black dots* depict measured skin ST changes for this song. The *gray areas* under the *curve* indicate the probability that the song decreases or increases skin temperature. The *white area* under the *curve* indicates the probability that the song does not strongly influence skin temperature. The KDE of this song shows that this song has a very high probability of increasing skin temperature

distributions, the predictions made by the model also contain levels of uncertainty. For instance, a song that decreased skin temperature in half of the cases and increased skin temperature in the other half of the cases would probably have a mean of zero. In this case, the mean would thus provide a very uncertain measure of the effect of the song. Our model would capture the uncertainty of predictions for this song and not select the song. Second, in contrast to the MPTrain system (Oliver and Flores-Mangas 2006), we do not use any specific properties of the music. Although music characteristics can provide useful information (Husain et al. 2002; Webster and Weir 2005), we base our system completely on the listener's personal reactions to the song. We believe these measurements provide the most direct information about a song's effect on a given listener.

For the music selection in the rest of the paper, we divided the KDEs in three ranges: a negative range $(-\infty, -0.5]$, a neutral range $(-0.5, 0.5)$, and a positive range $[0.5, \infty)$. The KDEs resulting from the data gathered in Step 1 show that for each participant 10 songs have a $p > 0.40$ of increasing skin temperature and five songs have a $p > 0.50$ of increasing skin temperature. Furthermore, nine songs have a $p > 0.40$ of decreasing skin temperature and four songs have a $p > 0.50$ of decreasing skin temperature. This suggests that a selection of songs can be used to direct the skin temperature of the listener. Figure 7 depicts two typical examples of resulting KDEs. These KDEs show the difference between a song with a rather clear effect and a song that probably does not have a strong effect on the user's mood.

6 Step 3: validation

With the personalized AMP developed, its empirical validation was the essential subsequent step. To test the performance of our models, we submitted them to a two-week real-world trial with the same users as we built the models for. We validated these

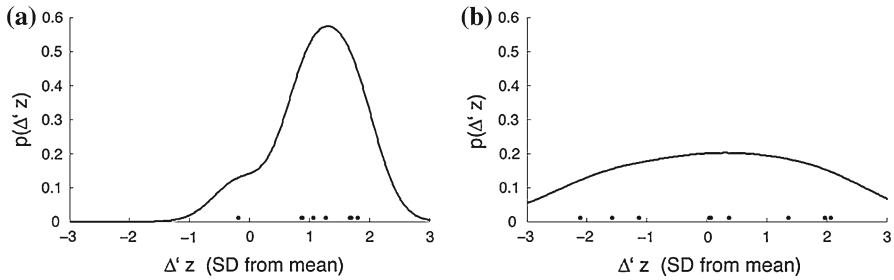


Fig. 7 Typical examples of participants' KDEs. The *dots* depict measured values, and the *line* represents the KDE over ST. The first song (**a**) is an example of a song that increases skin temperature. The second song (**b**) has no systematic effect on the participant's skin temperature and can, therefore, not be used to direct skin temperature. **a** Increase of ST (participant 2, song 21). **b** Undirected ST (participant 1, song 17)

Table 1 Examples of a positive and negative playlist generated in the validation session

Positive playlist	Negative playlist
Racoon—Feel like flying	Ilse de Lange—Old tears
Dave Brubeck Quartet—Blue rondo a la turk	Marco Borsato—De waarheid
Queen—Don't stop me now	Bob Dylan—Blowin in the wind
Live—Simple creed	Chris de Burgh—Carry me
Bon Jovi—It's my life	Counting crows—Colorblind

models against the valence ratings obtained before we started gathering the data (see also the data gathering section).

6.1 Experiment setup

The same setup and materials were used as in Step 1. The same three participants participated in four sessions of music listening; each session comprised of 18 songs. To be able to assess μ_n and σ_n for session standardization (see also Eq. 1), every session started with eight neutral songs defined as the songs with the highest probabilities in $(-0.5, 0.5)$. Then, two conditions followed: one that tried to increase skin temperature level and one that tried to decrease skin temperature level. For both of these conditions, we selected a block of five songs from the trained user models described in the previous section. For each participant, a block of five songs was selected with the highest increasing probability and a block of five songs was selected with the highest decreasing probability. This was done by integrating the pdf (Eq. 5) over $\Delta'z \in [0.5, \infty)$ and $\Delta'z \in (-\infty, -0.5]$, respectively, and selecting the songs with the highest probability values in each interval. An example of two playlists that were generated based on this method can be found in Table 1. The order of the two conditions was counterbalanced over the four sessions. Each session was conducted on a different day. Again, the sessions were conducted during regular working activities. The validation phase took about two weeks per participant.

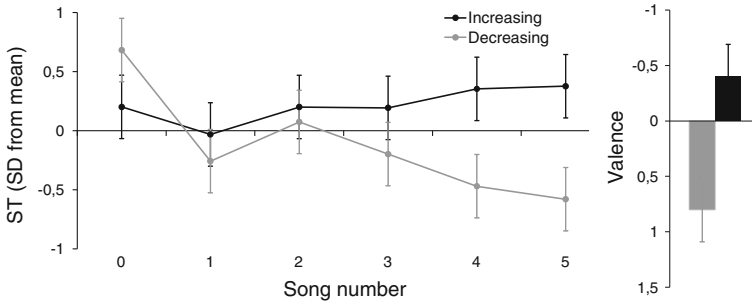


Fig. 8 The left graph presents the mean skin ST over the five songs selected to direct it towards a positive or negative level, as averaged over the participants and the sessions. The graph on the right presents the corresponding valence ratings of the music played in decreasing (gray) and increasing (black) blocks of music. Note that the valence scale is inverted to facilitate comparison with the left graph. The error bars depict ± 1 SE

6.2 Results

6.2.1 Skin temperature

The mean skin temperature of the last minute of each song was extracted and standardized over all songs in the session, using Eq. 1. One session of one participant contained measurement errors due to a loose sensor and was, therefore, excluded from further analyses. Means and standard errors (SEs) over all participants and sessions are depicted in Fig. 8.

To see whether or not the direction of skin temperature was successful, we ran a repeated measures ANOVA on skin temperature including data from all sessions of all participants. Song number (0/1/2/3/4/5; where 0 is the song directly preceding the directing block) was a within-block factor and direction (increasing / decreasing) was a between-block factor. Here, block is the set of five songs selected to either increase or decrease skin temperature. An interaction effect of song number \times direction was found ($F(5, 90) = 2.43, p < 0.050, \eta^2 = 0.11$). Pairwise comparisons showed that, for the decreasing direction, skin temperature of song number 0 ($m = 0.68$) was higher than skin temperature of song number 4 ($m = -0.47; t(20) = 3.06, p < 0.006$, Cohen's $d = 1.30$) and 5 ($m = -0.58; t(20) = 3.49, p < 0.002$, Cohen's $d = 1.49$). No effects were found for the increasing direction.

Subsequent ANOVAs on skin temperature over selected parts of the block with direction (increasing / decreasing) as between-block factor showed that the effect of direction becomes stronger over time (see Table 2). Moreover, pairwise comparisons showed that, for song number 4 and 5, skin temperature is lower in the decreasing direction than in the increasing direction, respectively $m_{4neg} = -0.47$ and $m_{4pos} = 0.35$ ($t(20) = -2.25, p < 0.036, d = 0.96$) and $m_{5neg} = -0.58$ and $m_{5pos} = 0.38$ ($t(20) = -2.42, p < 0.025$, Cohen's $d = 1.03$).

Participants' individual skin temperature changes were also examined by running repeated measures ANOVAs on skin temperature for each individual participant, with song number as within-block factor and direction as between-block factor. For the

Table 2 Results of ANOVAs on skin temperature, with direction (increasing/decreasing) as between-block factor

Song numbers	$F(1, 20)$	p	η^2
0, 1, 2, 3, 4, 5	2.24	0.150	0.11
1, 2, 3, 4, 5	3.79	0.066	0.16
2, 3, 4, 5	4.11	0.056	0.17
3, 4, 5	5.34	0.032	0.21

first participant, a main effect of direction was found ($F(1, 4) = 13.7$, $p < 0.021$, $\eta^2 = 0.77$), where skin temperature was higher in the increasing direction ($m = 0.45$) than in the decreasing direction ($m = -0.31$). For the second participant, an interaction effect of song number \times direction was found ($F(5, 30) = 4.74$, $p < 0.003$, $\eta^2 = 0.44$). A t test showed that skin temperature declined in the decreasing direction ($t(6) = 3.64$, $p < 0.011$, $d = 2.97$) but did not change in the increasing direction. For the third participant, no effects were found. Inspection of the individual sessions for this participant revealed an inverted pattern for one of the four sessions. This was confirmed by t -tests: in the inverted session, skin temperature was higher in the decreasing ($m = 0.61$) than in the increasing direction ($m = -0.61$; $t(8) = 4.67$, $p < 0.002$, $d = 3.30$). In contrast, in the other three sessions skin temperature was higher in the increasing direction ($m = 0.31$) than in the decreasing direction ($m = -0.29$; $t(28) = 2.10$, $p < 0.045$, $d = 0.79$).

6.2.2 Subjective measures

Before the data gathering (Step 1), the participants had expressed how they expected the songs would make them feel on a valence scale. We used these as a self-report validation metric. The ratings of the songs selected for the increasing skin temperature direction were compared to the ratings of the songs selected for the decreasing skin temperature direction. This was done through an independent samples t -test on the feeling ratings, with direction as independent variable. As expected, the songs of the decreasing skin temperature direction ($m = 0.8$) were more positively rated than the songs in the increasing skin temperature direction ($m = -0.4$; $t(28) = 2.05$, $p < 0.05$, Cohen's $d = 0.77$). This means that in the block that successfully decreased skin temperature, participants judged the selected songs as more positive compared to the music that was selected to increase skin temperature; see also Fig. 8.

7 Discussion

Founded on six considerations, we developed and validated an AMP. The AMP uses personalized probabilistic models that deal with individual differences and environmental noise. Moreover, we have employed the LIV in a physiological closed-loop system. Finally, instead of measuring short-term emotions, we have focused on longer-term effects of moods. The development and validation of the AMP comprised three

steps: (1) physiological responses during real-world music listening were gathered; (2) user-specific parts of the AMPs were personalized; and (3) the resulting system was validated in the real world, using skin temperature and subjective measures.

The validation of the AMP was successful. The songs selected to reduce skin temperature did indeed decrease skin temperature. It took three songs before these effects reached significance (Fig. 8). This period is in line with controlled laboratory experiments that employ musical mood induction periods of eight minutes (Gendolla and Krüsken 2001). These results were not only found over all data, but also for each individual participant. This is important, as we are most interested in developing user-specific models, instead of an overall population model. That is also the reason why we put our effort into gathering a large amount of real-world data for a few individuals, instead of a few data points for many people.

We validated our physiological results against self-report measurements. Based on our pilot study, we expected lower skin temperature to relate to more positive valence. This was confirmed by the fact that the music that decreased skin temperature was positively rated, whereas the music that increased skin temperature was negatively rated. These ratings also indicated that lowering skin temperature (and increasing valence) was more successful than increasing skin temperature (and decreasing valence). This might be because the participants were more likely to have music in their personal collection that would make them feel better as opposed to music that would make them feel worse, as is confirmed by the more neutral ratings of music selected for increasing skin temperature (and decreasing valence).

7.1 Limitations and further research

A first limitation of our work is the fact that we used only one physiological measure (i.e., skin temperature) to direct mood. Most traditional machine learning approaches to affective computing employ many different modalities and features to predict mood or emotions (Calvo and D'Mello 2010). However, because of the novelty of our approach, we decided to use only one physiological signal to make the results easier to understand and explain. With this one signal our system was already able to direct mood to two extremes. Nonetheless, the use of more signals might significantly improve the performance of the system and the resolution of mood states the system can model. In that light, it is good to note that our model can be extended to deal with multiple signals in two ways. First of all, if we assume the different physiological changes to be independent of each other, selections could be made, based on multiple pdfs to direct several physiological signals simultaneously. Second, as the physiological changes will probably be correlated, multi-dimensional KDEs (Scott and Sain 2005) can be employed to model the effects of multiple physiological variables in one pdf.

A second limitation of our system is the fact that new users have to listen to each song a few times before accurate predictions can be made (i.e., the cold start problem). There are several approaches that can be investigated to overcome this problem. First, aggregates can be made of other user's physiological responses to songs that can serve as a population model. Such a population model can be employed as an a priori distribution that is updated with KDEs each time the new user listens to the song.

Second, once a user model has been constructed for one song, this user model can be used as an a priori model for similar songs. There are many ways of calculating song similarity already available (Mandel et al. 2006; Sotiropoulos et al. 2008). Finally, a priori models can also be based on personality and music preference as affective responses to music depend on these variables (Rentfrow and Gosling 2003).

In the current work, we have focused on two types of moods: very positive and very negative moods. We chose these moods as they form two extremes that are easy to interpret. Nonetheless, this is not the whole spectrum of possible moods. Moods can differ on tension, energy, and valence (Matthews et al. 1990; Thayer 1989). Although our results are validated on valence, we did not control for tension and energy. The question remains whether music can direct mood on each of these dimensions independent of the other two dimensions. If this is possible, future research can aim to create more complex models to differentiate between directing tension, energy, and valence. This will probably require the use of multiple physiological signals.

For one session of one participant, the effect of the music was inverted. The participant characterized his mood that day as very frustrated, sad, and tense. As he was feeling very negative, happy positive music might actually have annoyed the participant, whereas gradually less negative music might have been a better way to regulate his mood to a more positive state (Saarikallio and Erkkilä 2007). If this turns out to be the case in future research, the AMP could adapt to this by selecting music that is neutral, so as not to invoke any annoyance. Once the user has then reached a more neutral level, the music player can shift to select more positive music. Another solution is to have the user control the type of music selected. That way, the user can change the setting when annoyingly happy music is selected. Furthermore, besides basing the goal state on the user's current emotional state, it might also be useful to base the goal state on the context the user is in. For instance, when the user is exercising, energizing music is more relevant than relaxing music. In sum, more research on suitable goal states is likely to improve the AMPs performance.

7.2 Implications and applications

Although there are still issues to be resolved with future research, we did successfully develop and validate an AMP. As we made a large effort to use real-world contexts, we were able to achieve a high ecological validity. This is important, as real-world settings could render the use of physiological signals problematic. Nevertheless, our models have shown to be robust against the real-world noise.

As we stated in the introduction, the AMP is also a good carrier application for research on affective computing. First of all, when developing affective technology, it is important to consider the different affective dimensions (e.g., moods, emotions, and attitudes) and understand the implications of each of them. We focused on moods, as they change gradually and changes in mood are not very distracting. As a result, we investigated effects over multiple songs, instead of focusing on effects within a song. Second, when using physiological signals, it is important to consider their properties and their relation to psychological constructs. We used skin temperature as our and others' research indicated its relation to valence (Baumgartner et al. 2006; McFarland

and Kennison 1989; Rimm-Kaufman and Kagan 1996). In addition, we used the LIV (Wilder 1967) to correct our measured values for prestimulus levels. If we had not done this, our models would have contained more noise, and might not have performed as well as they did now. Third, it might not be necessary to continually infer affective states from physiological signals, which is often the approach taken in affective computing applications (e.g., Calvo and D’Mello 2010). Continuously inferring affective states from physiological signals is a very challenging process (Peter and Herbon 2006; Fairclough 2009; Van den Broek et al. 2009) and adopting a *physiological* closed-loop (Tractinsky 2004; Höök 2009) helped to overcome this problem. Finally, affective responses are often very personal, so user-specific models should be employed.

A range of applications other than a music player can be identified that could benefit from the different elements of our approach. First, consider an application that tries to enhance human emotion communication. It could try to continuously infer the communicators’ affective states from physiology and share those inferred affective states with other people. However, the physiological signals could also be communicated directly instead of having a computer interpret them first (Janssen et al. 2010). This creates a physiological closed-loop like the one we have used for the AMP. Furthermore, to keep such physiological messages understandable, the LIV can be used to filter out noise that might otherwise be difficult to deal with for the receiver. A second example is a haptic movie enhancement system, which tries to stimulate emotional reactions to a movie through haptic stimulation (Lemmens et al. 2009). Physiological responses could be used to measure emotional responses to haptic patterns, which could teach the system which haptic patterns work well for that particular user. In this way, the system could be personalized. Here, a system similar to our music player could be implemented, building KDEs for different haptic patterns instead of different songs. As a third example, consider an in-car system that detects a driver’s stress and responds appropriately (Healey and Picard 2005). Such a system could use physiological signals to continually measure stress levels. Hence, it could benefit from employing the LIV to correct the measured physiology. Furthermore, the physiological closed-loop could be used to model how users respond to different adaptations the system makes to reduce driver stress. For instance, for some people it might be good to reduce the volume of the music, while for other’s colored lighting in the car might work better (Caberletti et al. 2009). Finally, our approach to the AMP could also be interesting for other applications outside of affective computing. For instance, a domain in which our approach might be useful is persuasive technology (Fogg 2003) that deals with technology intended to change our behavior (e.g., to get us to exercise more). It is becoming increasingly clear that different users respond differently to different persuasive messages (Kaptein and Eckles 2010). For instance, while some people might be particularly susceptible to consensus arguments (e.g., “everybody does this”), others are more susceptible to authority (e.g., “a doctor recommends this”). Our approach to modeling physiological signals could be employed to model users’ emotional reactions to different arguments or system behaviors. This way, the system can learn what the most effective messages are for that particular user. Again, it would be important to carefully consider whether it is attitudes, moods, or emotions that are of interest in such a system. Moreover, the LIV and probabilistic modeling of the physiological signals can also improve the quality of the system.

To conclude, the lessons presented in this paper are not only relevant for affective computing, but also for other fields. Although real-world evaluations come with many drawbacks and challenges, they are necessary to test how systems perform in practice. Through real-world evaluations of considerations such as the ones we posed, we gain a good overview of the approaches that work well and the ones that do not. In turn, such research can help science to find its way to technology, hopefully resulting in many meaningful innovations.

Acknowledgments We gratefully acknowledge Marjolein van der Zwaag, Tim Tijs, Kathryn Segovia, and Maurits Kaptein for their helpful comments and vivid discussions on an earlier draft of this paper. We also thank three anonymous reviewers and the editor who all provided us detailed feedback on two earlier versions of this paper. Thanks to their comments and suggestions we have been able to revise this article substantially. Finally, we gratefully acknowledge Lynn Packwood for her careful proof reading.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Baas, M., De Dreu, C.K.W., Nijstad, B.A.: A meta-analysis of 25 years of moodcreativity research: hedonic tone, activation, or regulatory focus? *Psychol. Bull.* **134**, 779–806 (2008)
- Bak, A.A., Grobbee, D.E.: A randomized study on coffee and blood pressure. *J. Hum. Hypertens.* **4**, 259–264 (1990)
- Baumgartner, T., Esslen, M., Jäncke, L.: From emotion perception to emotion experience: emotions evoked by pictures and classical music. *Int. J. Psychophysiol.* **60**(1), 34–43 (2006)
- Beedie, C.J., Terry, P.C., Lane, A.M.: Distinctions between emotion and mood. *Cogn. Emot.* **19**, 847–878 (2005)
- Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer, New York (2006)
- Boehner, K., DePaula, R., Dourish, P., Sengers, P.: How emotion is made and measured. *Int. J. Hum.-Comput. Stud.* **65**, 275–291 (2007)
- Boucsein, W.: *Electrodermal Activity*. Plenum Press, New York (1992)
- Caberletti, L., Elfmann, K., Kümmel, M., Schierz, C.: Influence of ambient lighting in vehicle interior on the driver's perception. In: de Kort, Y., IJsselstein, W., Vogels, I., Aarts, M., Tenner, A., Smolders, K. (eds.) *Proceedings of Experiencing Light 2009 International Conference on the Effects of Light on Wellbeing*, pp. 5–13, Eindhoven, The Netherlands (2009)
- Cacioppo, J., Tassinary, L.: Inferring psychological significance from physiological signals. *Am. Psychol.* **45**, 16–28 (1990)
- Calvo, R.A., D'Mello, S.: Affect detection: an interdisciplinary review of models, methods, and their applications. *IEEE Trans. Affect. Comput.* **1**, 18–37 (2010)
- Carberry, S., de Rosis, F.: Introduction to special issue on affective modeling and adaptation. *User Model. User-Adapt. Interact.* **18**, 1–9 (2008)
- Chin, N.D.: Empirical evaluation of user models and user adaptive systems. *User Model. User-Adapt. Interact.* **11**, 181–194 (2001)
- Clore, G.L., Palmer, J.: Affective guidance of intelligent agents: how emotion controls cognition. *Cogn. Syst. Res.* **10**(1), 21–30 (2009)
- Csikszentmihályi, M.: *Flow: The Psychology of Optimal Experience*. Harper Collins, Sussex, UK (1990)
- de Rosis, F.: Preface: towards adaptation of interaction to affective factors. *User Model. User-Adapt. Interact.* **11**, 267–278 (2001)
- D'Mello, S., Craig, S.D., Witherspoon, A., McDaniel, B., Graesser, A.: Automatic detection of learners affect from conversational cues. *User Model. User-Adapt. Interact.* **18**, 45–80 (2008)
- Fairclough, S.H.: Fundamentals of physiological computing. *Interact. Comput.* **21**, 133–145 (2009)
- Fogg, B.J.: *Persuas. Technology*. Morgan Kaufmann Publishers, New York (2003)
- Frijda, N.H.: *The Emotions*. Cambridge University Press, New York (1986)

- Geenen, R., van de Vijver, F.J.R.: A simple test of the law of initial values. *Psychophysiology* **30**(5), 525–530 (1993)
- Gendolla, G.H.E.: On the impact of mood on behavior: an integrative theory and a review. *Rev. Gen. Psychol.* **4**, 378–408 (2000)
- Gendolla, G.H.E., Brinkman, K.: The role of mood states in self-regulation: effects on action preferences and resource mobilization. *Eur. Psychol.* **10**, 187–198 (2005)
- Gendolla, G.H.E., Krüsen, J.: Mood state and cardiovascular response in active coping with an affect-regulative challenge. *Int. J. Psychophysiol.* **41**, 169–180 (2001)
- Hanson, M.A., Powell, H.C. Jr., Barth, A.T., Ringgenberg, K., Calhoun, B.H., Aylor, J.H. et al.: Body area sensor networks: challenges and opportunities. *IEEE Comput.* **42**, 58–65 (2009)
- Härdle, W.: *Smoothing Techniques, with Implementations* in S. Springer, New York (1991)
- Healey, J.A.: Affect detection in the real world: recording and processing physiological signals. In: *Proceedings of the IEEE 3rd International Conference on Affective Computing and Intelligent Interaction, ACII, Vol. 1*, pp. 729–734. IEEE Press, Amsterdam (2009)
- Healey, J.A., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transp. Syst.* **6**, 156–166 (2005)
- Healey, J.A., Picard, R.W., Dabek, F.: A new affect-perceiving interface and its application to personalized music selection. In: Turk, M. (ed.) *Proceedings of the 1998 Workshop on Perceptual User Interfaces (PUI)*, San Francisco, CA, USA (1998)
- Heinz, C., Seeger, B.: Cluster kernels: resource-aware kernel density estimators over streaming data. *IEEE Trans. Knowl. Data Eng.* **20**, 880–893 (2008)
- Höök, K.: Affective loop experiences: designing for interactional embodiment. *Philos. Trans. R. Soc. B* **364**, 3585–3595 (2009)
- Husain, G., Thompson, W.F., Schellenberg, E.G.: Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Percept.* **20**, 151–171 (2002)
- Janssen, J.H., Bailenson, J.N., IJsselstein, W.A., Westerink, J.H.D.M.: Intimate heartbeats: opportunities for affective communication technology. *IEEE Trans. Affect. Comput.* **1**(2), 72–80 (2010)
- Kaptein, M., Eckles, D.: Selecting effective means to any end: futures and ethics of per30 sasion profiling. In: Ploug, T., Hasle, P., Oinas-Kukkonen, H. (eds.) *Persuasive Technology*, pp. 82–93. Springer, Berlin (2010)
- Kim, J., André, E.: Emotion recognition based on physiological changes in music listening. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**, 2067–2083 (2008)
- Kowal, J., Fortier, M.: Motivational determinants of flow: contributions from selfdetermination theory. *J. Soc. Psychol.* **139**, 355–368 (1999)
- Lemmens, P., Crompvoets, F., Brokken, D., van den Eerenbeemd, J., De Vries, G.-J.: A body-conforming tactile jacket to enrich movie viewing. In *World Haptics Conference*, pp. 7–12. IEEE, Los Alamitos (2009)
- Lesiuk, T.: The effect of music listening on work performance. *Psychol. Music* **33**, 173–191 (2005)
- Liljedahl, M., Sjömark, C., Lefford, N.: Using music to promote physical well-being via computer-mediated interaction. In *MusicNetwork Open Workshop*, 5 (2005)
- Mandel, M., Poliner, G., Ellis, D.: Support vector machine active learning for music retrieval. *ACM Multimed. Syst. J.* **12**, 3–13 (2006)
- Matthews, G., Jones, D.M., Chamberlain, A.G.: Refining the measurement of mood: the UWIST mood adjective checklist. *Br. J. Psychol.* **81**, 17–42 (1990)
- McFarland, R.A., Kennison, R.: Asymmetry in the relationship between finger temperature changes and emotional state in males. *Appl. Psychophysiol. Biofeedback* **14**(4), 281–290 (1989)
- North, A.C.H., David, J.: Musical preferences during and after relaxation and exercise. *Am. J. Psychol.* **113**, 43–67 (2000)
- North, A.C.H., Hargreaves, D.J., Hargreaves, J.J.: Uses of music in everyday life. *Music Percept.* **22**, 41–77 (2004)
- Oliver, N., Flores-Mangas, F.: MPTrain: a mobile, music and physiology-based personal trainer. In: *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 21–28. ACM, New York (2006)
- Oliver, N., Kregor-Stickles, L.: PAPA: physiology and purpose-aware automatic playlist generation. In: Lemström, K., Tindale, A., Dannenberg, R. (eds.) *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR)*, pp. 250–253, Victoria, Canada, 8–12 October 2006
- Ophira, E., Nass, C., Wagner, A.D.: Cognitive control in media multitaskers. *Proc. Natl Acad. Sci.* **106**, 15583–15587 (2009)

- Pantic, M., Patras, I.: Dynamics of facial expressions: recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Trans. Man Syst. Cybernet. B* **36**, 433–449 (2006)
- Pelletier, C.L.: The effect of music on decreasing arousal due to stress: a meta-analysis. *J. Music Ther.* **41**, 192–214 (2004)
- Peter, C., Herbon, A.: Emotion representation and physiology assignments in digital systems. *Interact. Comput.* **18**, 139–170 (2006)
- Picard, R.W.: *Affective Computing*. MIT Press, Cambridge (1997)
- Picard, R.W.: Affective computing: challenges. *Int. J. Hum.-Comput. Stud.* **59**, 55–64 (2003)
- Prinz, J.J.: *Gut Reactions: A Perceptual Theory of Emotion*. Oxford University Press, New York (2004)
- Rentfrow, P.J., Gosling, S.D.: The do re mi's of everyday life: the structure and personality correlates of music preference. *J. Pers. Soc. Psychol.* **84**, 1236–1256 (2003)
- Rickard, N.S.: Intense emotional responses to music: A test of the physiological arousal hypothesis. *Psychol. Music* **32**, 371–388 (2004)
- Rimm-Kaufman, S.E., Kagan, J.: The psychological significance of changes in skin temperature. *Motiv. Emot.* **20**(1), 64–78 (1996)
- Ritossa, D.A., Rickard, N.S.: The relative utility of 'pleasantness' and 'liking' dimensions in predicting the emotions expressed by music. *Psychol. Music* **32**, 5–22 (2004)
- Russell, J.A.: Core affect and the psychological construction of emotion. *Psychol. Rev.* **110**, 145–172 (2003)
- Rusting, C.L.: Personality, mood, and cognitive processing of emotional information: three conceptual frameworks. *Psychol. Bull.* **124**, 165–196 (1998)
- Saarikallio, S., Erkkilä, J.: Role of music in adolescents' mood regulation. *Psychol. Music* **35**, 88–109 (2007)
- Scott, D., Sain, S.: Multidimensional density estimation. In: Rao, C.R., Wegman, E.J., Solka, J.L. (eds.) *Handbook of Statistics*, Vol. 24, pp. 229–261. Elsevier, North Holland (2005)
- Silverman, B.W.: *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London (1986)
- Sloboda, J.A.: *Exploring the Musical Mind: Cognition, Emotion, Ability, Function*. Oxford University Press, New York (2005)
- Sotiropoulos, D.N., Lampropoulos, A.S., Tsihrintzis, G.A.: MUSIPER: a system for modeling music similarity perception based on objective feature subset selection. *User Model. User-Adapt. Interact.* **18**, 315–348 (2008)
- Thayer, R.E.: *The Biopsychology of Mood and Activation*. Oxford University Press, New York (1989)
- Tractinsky, N.: Tools over solutions? comments on interacting with computers special issue on affective computing. *Interact. Comput.* **16**, 751–757 (2004)
- Turlach, B.A.: Bandwidth selection in kernel density estimation: a review. Discussion Paper 9317, Institut de Statistique, Voie du Roman Pays 34, B-1348 Louvain-la-Neuve (1993)
- Vaillant, G.: *Aging Well: Surprising Guideposts to a Happier Life from the Landmark Harvard Study of Adult Development*. Little, Brown and Company, Boston (2003)
- Van den Broek, E.L., Janssen, J.H., Westerink, J.H.D.M.: Guidelines for Affective Signal Processing (ASP): From lab to life. In *Proceedings of the IEEE 3rd international conference on affective computing and intelligent interaction, ACII*, Vol. 1, pp. 704–709. IEEE Press, Amsterdam, The Netherlands (2009a)
- Van den Broek, E.L., Janssen, J.H., Westerink, J.H.D.M., Healey, J.A.: Prerequisites for Affective Signal Processing (ASP). In: Encarnaça, P., Veloso, A. (eds.) *Biosignals 2009: Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, pp. 426–433, Porto, Portugal (2009b)
- Van den Broek, E.L., Lisý, V., Janssen, J.H., Westerink, J.H.D.M., Schut, M.H., Tuinenbreijer, K.: Affective man-machine interface: unveiling human emotions through biosignals. In: Fred, A., Filipe, J., Gamboa, H. (eds.) *Biomedical Engineering Systems and Technologies: BIOSTEC2009 Selected Revised Papers*. Vol. 52., pp. 21–47. Springer, Berlin (2010)
- Wagenaar, W.A.: Note on the construction of digram-balanced latin squares. *Psychol. Bull.* **72**, 384–386 (1969)
- Webster, G.D., Weir, C.G.: Emotional responses to music: interactive effects of mode, texture, and tempo. *Motiv. Emot.* **29**, 19–39 (2005)
- Westerink, J.H.D.M., De Vries, G., Waele, S., Eerenbeemd, J., Boven, M., Ouwkerk, M.: Emotion measurement platform for daily life situations. In: Nijholt, A., Cohn, J., Pantic, M. (eds.) *Proceedings of ACII'09: Affective Computing and Intelligent Interaction*, pp. 217–223. IEEE, Los Alamitos (2009)
- Wilder, J.: *Stimulus and Response: The Law of Initial Values*. Wright, Bristol (1967)

- Wilhelm, P., Schoebi, D.: Assessing mood in daily life: structural validity, sensitivity to change, and reliability of a short-scale to measure three basic dimensions of mood. *Eur. J. Psychol. Assess.* **23**, 258–267 (2007)
- Wilson, G.F., Russell, C.A.: Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Hum. Factors* **45**, 635–644 (2003)
- Yannakakis, G.N., Hallam, J., Lund, H.H.: Entertainment capture through heart rate activity in physical interactive playgrounds. *User Model. User-Adapt. Interact.* **18**, 207–243 (2008)
- Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**, 39–58 (2009)

Author Biographies

Joris H. Janssen is a Ph.D. Candidate at Eindhoven University of Technology, researcher at Philips Research, and a visiting scholar at Stanford University. He received his B.Sc. (2007) and M.Sc. (2008; cum laude) degrees in Artificial Intelligence from the Radboud University Nijmegen (The Netherlands). His research interests are centered around affective computing, emotion, psychophysiology, and social connectedness. His work appeared in about 25 peer-reviewed conference and journal articles as well as book chapters.

Egon L. van den Broek is Consultant and Assistant Professor (University of Twente and Radboud University Medical Center Nijmegen, NL). He received his M.Sc. (2001) in Artificial Intelligence (AI) and Ph.D. (2005) in image retrieval both from the Radboud University Nijmegen and a Ph.D. (2011) in affective computing from the University of Twente, all in NL. He conducts research on intelligent and adaptive systems, usability, cognitive computer vision, biosignals, affective computing, and computational geometry. He supervised more than 50 students, published more than 150 papers, and holds five patent applications and various awards.

Joyce H. D. M. Westerink studied physics at Utrecht University (M.Sc., 1985). She then joined Philips Research in Eindhoven, the Netherlands, and specialized in human perception and interaction with technology. She received her Ph.D. from Eindhoven University of Technology (1991) as part of her work at Philips Research, and has remained active in that field ever since. She published about 50 book chapters and journal articles, and holds 20 patents and patent applications.