AMAI: Adaptive music for affect improvement

AMAI: Adaptive Music for Affect Improvement

David Su
MIT Media Lab
davidsu@media.mit.edu

Rosalind Picard
MIT Media Lab
picard@media.mit.edu

Yan Liu
Harvard Graduate School of Design
yliu1@gsd.harvard.edu

ABSTRACT

This paper introduces Adaptive Music for Affect Improvement (AMAI), a music generation and playback system whose goal is to steer the listener towards a state of more positive affect. AMAI utilizes techniques from game music in order to adjust elements of the music being heard; such adjustments are made adaptively in response to the valence levels of the listener as measured via facial expression and emotion detection. A user study involving AMAI was conducted, with N=19 participants across three groups, one for each strategy of Discharge, Diversion, and Discharge → Diversion. Significant differences in valence levels between music-related stages of the study were found between the three groups, with Discharge → Diversion exhibiting the greatest increase in valence, followed by Diversion and finally Discharge. Significant differences in positive affect between groups were also found in one before-music and after-music pair of self-reported affect surveys, with Discharge → Diversion exhibiting the greatest decrease in positive affect, followed by Diversion and finally Discharge; the resulting differences in facial expression valence and self-reported affect offer contrasting conclusions.

1. INTRODUCTION

Music and emotion are inextricably linked, and the relationship between the two has fascinated humans for centuries. Plato wrote in *Laws* that "rhythm and music generally are a reproduction expressing the moods of better and worse men" [1], and Minsky in his essay *Music, Mind, and Meaning* grapples with the mystery of why we enjoy listening to music and the ways in which "it touches our emotions" (as well as how to reason about music) [2].

While musicians do not always create music with the express purpose of improving listeners' affects, there is great potential for music to steer listeners towards a more positive affective state, given the power that music has over our emotions. Indeed, Thayer et al. [3], Gallup & Castelli [4], and Parker & Brown [5] found music to be the 2nd, 6th, and 10th most effective method for mood regulation in comparison to other methods such as exercise and food consumption.

This paper explains the design, background, implementation, and validation of Adaptive Music for Affect Improvement (AMAI), a system that uses game music techniques to generate music which varies in accordance to a listener's affective state, with the ultimate goal of improving the listener's affect.

2. RELATED WORK

2.1 Music, Mood, and Emotion

Saarikallio formulated a scale for the "use of music for mood regulation" [6], proposing that such regulation occurs primarily through "seven distinct regulatory strategies: *Entertainment*, *Revival*, *Strong Sensation*, *Diversion*, *Discharge*, *Mental Work*, and *Solace*". These strategies form the basis of the different ways in which AMAI adapts to listeners' affects. Similarly, Zentner, Meylan, and Scherer developed a model for describing and measuring emotions evoked by music [7]. In the process, they describe the distinction between perceived emotion and felt (or induced) emotion in music, and report a previous finding that positively valenced emotions tend to be perceived and felt at roughly the same ratio, while negatively valenced emotions tend to be perceived more than felt. Their model and insights were used to inform the creation of the adaptive music in AMAI.

Kawamaki et al. demonstrate that while sad music leads to perception of more tragic emotions, the emotions that are actually induced fall more into categories of romance and blitheness [8]. This could be seen as an application of Saarikallio's discharge strategy as a means of "releasing anger or sadness through music that expresses these emotions" [6], with the listener's perceived emotions corresponding to the music's expressed emotions and the listener's felt emotions corresponding to the listener's release of those perceived emotions. As discharge is one of the mood regulation strategies that the AMAI study aims to examine, Kawamaki et al.'s findings were helpful in analyzing participants' affect, in particular looking at whether the facial expression data and self-reported data differ in a manner similar to that of felt and perceived emotion in the case of Kawamaki et al.'s study.

2.2 Affective Properties of Musical Elements

The musical underpinnings of the AMAI system draw from research on the affective properties of music. In comparing different musical characteristics' effects on affect, Hevner found that "the use of the major or minor mode is of the most clear-cut significance" [9], noting that "the major mode is very strongly associated with happiness, gayety, playfulness and sprightliness, and the minor with sadness, and with sentimental yearning, tender effects". As such, AMAI gives major tonalities greater prominence when attempting to steer individuals towards happiness,

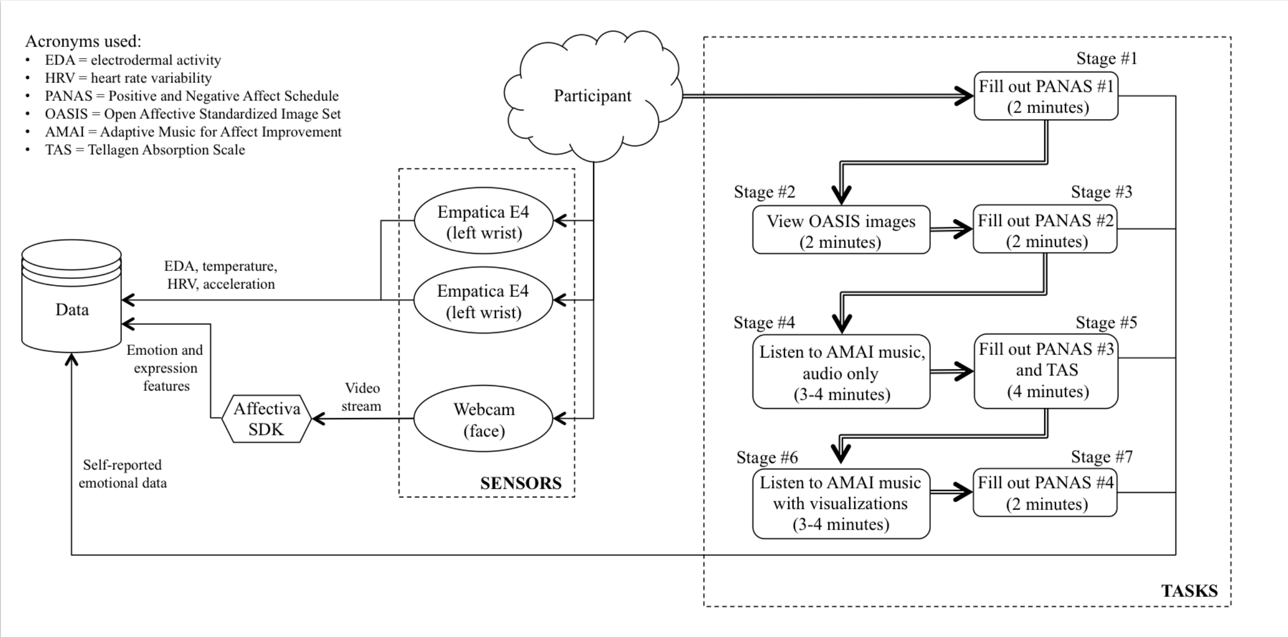

Figure 1. AMAI experiment design. Data from sensors are collected during all tasks.

while minor tonalities are used when attempting to reflect individuals' sadness.

Balkwill and Thompson found that even listeners who are unfamiliar with a tonal system are sensitive to the emotion being expressed by a piece of music in that tonal system [10]. It should be noted that Balkwill and Thompson focused on Western listeners listening to Hindustani ragas, while AMAI operates in the context of a Western tonal system. However, their findings suggest that harmonic expression of affect is to some extent cross-cultural.

Hunter et al. examined how tempo and harmony in music affect listeners' self-reported emotional responses [11]. They found that faster tempos and major modes elicited more happiness, while slower tempos and minor modes elicited more sadness. (Interestingly, both happiness and sadness ratings were higher when faster tempos were combined with major modes, or slower tempos with minor modes.) These findings are consistent with Hevner's conclusions on harmony, and support the use of tempo and tonality modulation within the AMAI system. In addition, the study found that ratings were higher for perceived emotion than felt emotion, especially for sad music; this is consistent with Zentner, Grandjean, and Scherer's results.

Lundqvist et al. demonstrate that listeners' responses to happy musical stimuli and sad musical stimuli (both of which were composed specifically for the study) differ in autonomic activity, facial expression, and self-reported emotion [12]. The happy stimuli "generated more zygomatic facial muscle activity, greater skin conductance, lower finger temperature, more happiness, and less sadness" than the sad stimuli. These results can be used as a basis for evaluating and comparing the happiness and sadness evoked by the AMAI system for different groups.

3. METHODOLOGY

3.1 Experiment Design

N=19 participants took part in the study, and were randomly separated into three groups corresponding to three different MMR strategies: discharge (7), diversion (6), and a combination of the two (6). In the case of the combination, the AMAI system would initially operate in the mode of discharge and then transition to applying diversion strategies. The groups are thus referred to as *Discharge*, *Diversion*, and *Discharge* → *Diversion*.

Each session lasted 20-30 minutes, and participants were told that the study's purpose was to compare affects of different visual and musical stimuli, so as not to bias their affective perceptions of the different tasks. A post-study debriefing was given in order to reveal the true nature of the study to participants. The study protocol was reviewed and approved by the MIT Committee on the Use of Humans as Experimental Subjects. As compensation for their involvement in the study, participants were given $25 Amazon gift cards after the end of the study.

Figure 1 shows the AMAI experiment design. Participants were asked to engage in three tasks: view a set of 20 images from the OASIS dataset [13] (specifically chosen to elicit negative affect), listen to 3 minutes of music generated by the AMAI system, and then listen again to 3 minutes of AMAI-generated music with visual accompaniment. In between each task as well as at the beginning and end of the study, participants were asked to fill out a Positive and Negative Affect Schedule (PANAS) survey [14]; before the final task, participants were also asked to fill out the Tellegen Absorption Scale (TAS) [15]. The tasks in combination with the surveys are collectively referred to as *stages*; as shown in Figure 1, a study session comprises seven stages in total. The three tasks in the experiment were implemented using the Unity engine (version 2017.1), with the five surveys (four PANAS + one TAS) being filled out using Google Forms.

Throughout the study, participants had their face analyzed for emotion and expression data using the Affectiva Emotion SDK (version 2.3), in addition to having their physiological data recorded using two Empatica E4 sensors, one on each wrist. Facial data was saved as a series of timestamped emotion and expression values.

3.2 Music Implementation

Of Saarikallio's seven MMR strategies [6], we chose to focus on *discharge* ("releasing anger or sadness through music that expresses these emotions"), *diversion* ("forgetting unwanted thoughts and feelings with the help of pleasant music"), and the progression from discharge to diversion. Given Lundqvist et al.'s results, our hypothesis was that the combination of the two approaches would result in the greatest increase in positive affect, followed by diversion on its own and finally discharge on its own.

The music generated by AMAI makes use of the game music principles of horizontal re-sequencing, which refers to the idea that "the sequence of a musical composition can be rearranged", and vertical layering, which refers to the "independent manipulation of the layers, enabling the overall track to change in accordance with the fluctuating state of the game" [16].

Original music was composed for six distinct sections; Table 1 describes the musical properties that characterize each of those sections. Horizontal re-sequencing was then applied to the sections, yielding the overall forms outlined in Table 2. In addition, transitions between sections were triggered by peaks in valence, detected from the participant's face. If the peak threshold was not reached within a given time period, the system would advance to the next section regardless.

Section	Key	Tempo	Instrumentation
A	G minor	72 BPM	piano, guitar
B	G minor	72 BPM	piano, guitar, cello, vibraphone
C	G minor / Bb major	74 BPM	piano, guitar, cello, vibraphone
D	Bb major	76 BPM	piano, guitar, cello, kick drum
E	Bb major	76 BPM	piano, guitar, cello, kick drum, vibraphone

Table 1. Musical characteristics of each section of AMAI music.

Group/Strategy	Form	Final Chord
Discharge	A-B-C-A-B	Gmin
Diversion	D-E-C-D-E	Bbmaj
Discharge $\rightarrow$ *Diversion*	A-B-C-D-E	Bbmaj

Table 2. Overall form, or progression between sections via horizontal resequencing, for each group/strategy.

Vertical layering was applied to individual instruments in all sections; in particular, the volume of the guitar, cello, and vibraphone, in addition to entrances and exits of those instruments, were tied to the valence deteced from the listener's face. In addition, the specific notes of the melody were stochastically determined so as not to tire the listener with exact repetition and looping.

Figure 2. Screenshot of the AMAI backend in FMOD.

The AMAI system was implemented using Firelight Technologies' audio middleware engine FMOD (version 1.09.06), which integrated into the experiment's Unity pipeline. Figure 2 shows a screenshot of the AMAI system being built in FMOD. Audio examples of AMAI output for each strategy can be found at `http://usdivad.com/amai/examples.html`.

4. RESULTS AND DISCUSSION

4.1 Facial data analysis

A one way ANOVA comparing the change in average valence, ranging from -1 to 1 as defined by the Affectiva Emotion SDK, between each stage yielded significant differences in the three groups' valence levels for all stages except for between stages 1 and 2 (that is, between filling out PANAS #1 and viewing OASIS images).

Figure 3 shows the averages in change in valence levels between all stages, among all participants. Of particular interest are the changes between average valence between stages 3 and 4 (filling out PANAS #2 and listening to AMAI music for the first time, audio only) $[F(1,17) = 7.016, p = 0.017, \eta^2 = 0.292]$, between stages 4 and 5 (listening to AMAI music for the first time and filling out PANAS #3), between stages 5 and 6 (filling out PANAS #3 and listening to AMAI music for the second time, with visualizations) $[F(1,17) = 14.933, p = 0.001, \eta^2 = 0.483]$, and between stages 6 and 7 (listening to AMAI music with visualizations and filling out PANAS #4) $[F(1,17) = 11.784, p = 0.003, \eta^2 = 0.424]$. These correspond to the stages before, during, and after experiencing the music generated by AMAI.

4.2 Self-reported data analysis

Figure 4 shows self-reported positive affect, calculated from participants' responses to the PANAS questionnaire, for each time they filled out the questionnaire.

A one way ANOVA comparing the change in affect between each questionnaire yielded significant differences between groups' affect levels reported in PANAS #2 (stage 3, between viewing OASIS images and listening to AMAI

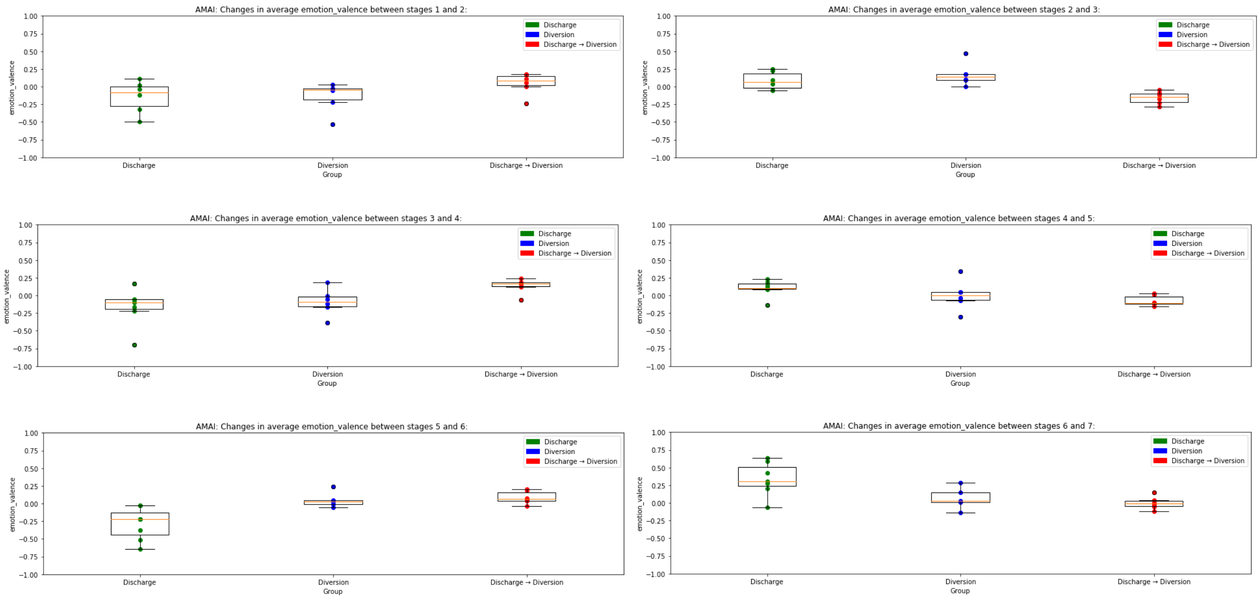

Figure 3. Changes in average valence values between stages 1 to 2, 2 to 3, 3 to 4, 4 to 5, 5 to 6, and 6 to 7.

music for the first time) and PANAS #3 (stage 5, after listening to AMAI music for the first time) $[F(1, 17) = 7.701, p = 0.013, \eta^2 = 0.312]$. In particular, the *Discharge* group's positive affect increased by an average of 9.43%, while the *Diversion* group's positive affect decreased by an average of 0.67% and the *Discharge* $\rightarrow$ *Diversion* group's positive affect decreased by an average of 3.67%.

4.3 Discussion

In terms of changes in valence as determined by facial data, for both cases where the change is measured between a before-music stage and a during-music stage (stages 3 to 4 and stages 5 to 6), the *Discharge* $\rightarrow$ *Diversion* group had the highest mean as well as the lowest standard deviation ($\mu \pm \sigma = 0.136 \pm 0.095$ and 0.088 ± 0.086 respectively), followed by the *Diversion* group ($\mu \pm \sigma = -0.093 \pm 0.173$ and 0.051 ± 0.102 respectively), with the *Discharge* group having the lowest mean and highest standard deviation ($\mu \pm \sigma = -0.162 \pm 0.245$ and -0.290 ± 0.218 respectively). This finding confirms our hypothesis, suggesting that the use of discharge followed by diversion may have more of a consistent effect across listeners than the employment of discharge or diversion strategies alone. If this is true, then a case can be made for adaptive music's unique potential to steer listeners' affects with more effectiveness than static music such as an MP3; in the case of AMAI, the transition from discharge to diversion was controlled by the user's affective state, with the system having control of the precise timing of when to switch sections and strategies. However, such a case must be supplemented by further examination of static music experiences that also employ discharge followed by diversion.

The findings from the self-reported data stand in contrast to the facial data results from Section 4.1. In addition, these results appear to contradict Lundqvist's findings that sad stimuli result in lower levels of self-reported happiness. However, some participants, in making subjective comments after the experiment, did note that the music intended to feel sad often felt more melancholy or soothing. Perhaps sections A and B of the music did not follow strongly enough Hevner [9] and Hunter et al.'s [11] findings regarding affective properties of musical characteristics. Another possibility is that despite the use of minor modes and slower tempos, *other* characteristics of the music did not adequately convey sadness to the participants. With regards to potential temporal issues, perhaps the transitions between *Discharge* to *Diversion* happened too suddenly; similarly, the sections themselves may have been too short. Yet another possibility is that participants simply do not accurately report the emotions that they feel; while Zentner found that negatively valenced emotions tended to be perceived at a greater magnitude than they were felt [7], under this interpretation the results from the AMAI study appear to indicate the opposite.

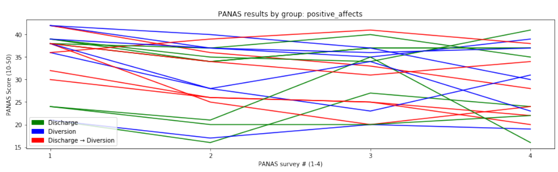

Figure 4. Changes in positive affect ratings, as derived from PANAS survey responses, for all three groups.

5. FUTURE WORK

Electrodermal activity, heart rate, and temperature data were also gathered, but have not yet been analyzed at this point in time. It may prove fruitful to examine how the physiological data compare to Lundqvist et al.'s findings [12] mentioned in Section 2.2. Further analysis can also be done comparing how self-reported affect differs from that inferred by facial expression and physiological signals.

The AMAI music generation process itself can also be improved. As mentioned in Section 4.3, the musical sections intended to elicit sadness may not actually do so succesfully; adjustments to characteristics such as transition time and section length may prove useful, as could more research into musical elements beyond harmony and tempo. In particular, Webster & Weir note that "non-harmonized melodies were rated happier than harmonized melodies for both major and minor modes" [17]; a more nuanced consideration of musical texture may improve the accuracy of AMAI's affective intentions.

As mentioned in Section 4.3, further investigation can be done into the necessity of an *adaptive* music system, and how the effectiveness of a static music experience, employing the same strategy as an adaptive one, might compare.

For future participants, we plan to add a TAS-20 questionnaire [18] in order to distinguish between people who are alexithymic and those who are not.

Finally, we hope to visualize the music in augmented reality (AR), using the Microsoft HoloLens headset and drawing inspiration from work on colored-hearing synesthesia [19]. Mapping sounds to colors, providing immersion through AR, and giving listeners more active agency (rather than having the system control all aspects of the music) could further improve AMAI's effectiveness.

6. CONCLUSION

In this paper we described the design, implementation, and validation of Adaptive Music for Affect Improvement (AMAI), a system for generation and playback of music that adapts to listeners' affective states with the goal of increasing positive affect. In particular, we explored the effectiveness of employing MMR strategies of discharge, diversion, and a combination of the two (with discharge followed by diversion).

We conducted a user study to evaluate the effectiveness of these three strategies. The results yielded from analysis of facial emotion and expression data as well as self-reported affect data suggest that employing discharge followed by diversion may steer listeners towards a state of more positive affect than either diversion or discharge alone. Self-reported data did not corroborate the facial expression and emotion data; potential reasons behind this range include inadequate elicitation of affect via music, inattention to musical details aside from harmony and tempo, and human behavior in the area of self-reporting. Finally, we outlined several recommendations for future work.

It is our hope that AMAI demonstrates the potential for an adaptive music system to improve affect and wellbeing, and that our results will be of benefit to the research community and society at large.

Acknowledgments

AMAI was made possible through funding from the Advancing Wellbeing initiative at the MIT Media Lab. Many thanks to Fengjiao Peng for her generous assistance throughout the study.

7. REFERENCES

[1] Plato and A. E. Taylor, *The Laws of Plato*. JM Dent & Sons, 1934.

[2] M. Minsky, "Music, mind, and meaning," in *Music, mind, and brain*. Springer, 1982, pp. 1–19.

[3] R. E. Thayer, J. R. Newman, and T. M. McClain, "Self-regulation of mood: Strategies for changing a bad mood, raising energy, and reducing tension." *Journal of personality and social psychology*, vol. 67, no. 5, p. 910, 1994.

[4] G. Gallup and J. Castelli, *The people's religion: American faith in the 90's*. Macmillan Publishing Company, 1989.

[5] G. B. Parker and L. B. Brown, "Coping behaviors that mediate between life events and depression," *Archives of General Psychiatry*, vol. 39, no. 12, pp. 1386–1391, 1982.

[6] S. H. Saarikallio, "Music in mood regulation: Initial scale development," *Musicae Scientiae*, vol. 12, no. 2, pp. 291–309, 2008.

[7] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions evoked by the sound of music: characterization, classification, and measurement." *Emotion*, vol. 8, no. 4, p. 494, 2008.

[8] A. Kawakami, K. Furukawa, K. Katahira, and K. Okanoya, "Sad music induces pleasant emotion," *Frontiers in psychology*, vol. 4, 2013.

[9] K. Hevner, "Experimental studies of the elements of expression in music," *The American Journal of Psychology*, vol. 48, no. 2, pp. 246–268, 1936.

[10] L.-L. Balkwill and W. F. Thompson, "A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues," *Music perception: an interdisciplinary journal*, vol. 17, no. 1, pp. 43–64, 1999.

[11] P. G. Hunter, E. G. Schellenberg, and U. Schimmack, "Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions." *Psychology of Aesthetics, Creativity, and the Arts*, vol. 4, no. 1, p. 47, 2010.

[12] L.-O. Lundqvist, F. Carlsson, P. Hilmersson, and P. N. Juslin, "Emotional responses to music: experience, expression, and physiology," *Psychology of music*, vol. 37, no. 1, pp. 61–90, 2009.

[13] B. Kurdi, S. Lozano, and M. R. Banaji, "Introducing the Open Affective Standardized Image Set (OASIS)," *Behavior research methods*, vol. 49, no. 2, pp. 457–470, 2017.

[14] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the PANAS scales." *Journal of personality and social psychology*, vol. 54, no. 6, p. 1063, 1988.

[15] A. Tellegen and G. Atkinson, "Openness to absorbing and self-altering experiences (" absorption"), a trait related to hypnotic susceptibility." *Journal of abnormal psychology*, vol. 83, no. 3, p. 268, 1974.

[16] W. Phillips, *A Composers Guide to Game Music*, ser. MIT Press. The MIT Press, 2014.

[17] G. D. Webster and C. G. Weir, "Emotional responses to music: Interactive effects of mode, texture, and tempo," *Motivation and Emotion*, vol. 29, no. 1, pp. 19–39, 2005.

[18] R. M. Bagby, J. D. Parker, and G. J. Taylor, "The twenty-item Toronto Alexithymia ScaleI. Item selection and cross-validation of the factor structure," *Journal of psychosomatic research*, vol. 38, no. 1, pp. 23–32, 1994.

[19] L. E. Marks, "On colored-hearing synesthesia: cross-modal translations of sensory dimensions." *Psychological bulletin*, vol. 82, no. 3, p. 303, 1975.