4ZL-03

Development of a Multimodal Mood Detection Mechanism for Bar Music Recommendation

Feike Xu† Shin'ichi Konomi‡
†Graduate School of ISEE, Kyushu University
‡Faculty of Arts and Science, Kyushu University

1. Introduction

When trying to recommend music for a public entertainment venue such as a bar, we often need to consider the emotion of the people in that venue to recommend music effectively. This significantly differs from traditional music recommendation systems, which typically focus on individual users. When recommending music in a bar, we need to consider the 'mood', which refers to the atmospheric characteristics or overall ambiance, of the bar.

In this work, we propose a multimodal mood detection mechanism that provides useful mood data for recommending music in bars. The proposed mechanism utilizes various sensor data representing alcohol concentration in the air, sound level, brightness level, crowd density to quantify the mood of the bar environment using a simple mixed weight-based calculation. Finally, we will elucidate how this mechanism operates in music recommendation.

2. Related Works

To detect mood, it is necessary to find an appropriate emotion space to evaluate the mood. Even though there is no standardized space, the two-dimensional emotion space model as shown in Figure 1 is well recognized, in which the fundamental aspects of music mood are represented in a two-dimensional space of valence and arousal [1]. In this model, valence axis describes the continuous scale from positive mood (happy, pleasant, etc.) to negative mood (unhappy, annoyed, etc.). The arousal axis represents the degree of calming or exciting. The mood is represented by a point based on the values of coordinates. Sometimes the mood categories can be designated in the two-dimensional emotion space as in Figure 1.

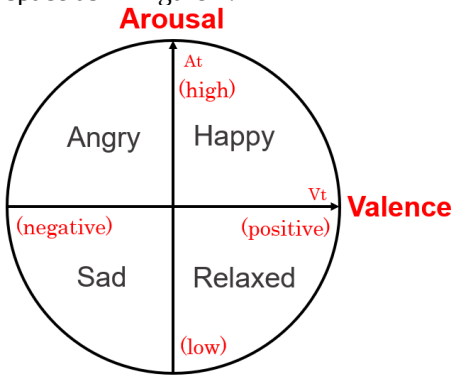

Figure 1. Arousal-valence space [2, 3].

Another aspect is the method to detect the mood of music. Bhattarai et al. [3] propose an automatic mood detection of music with a composition of transfer learning and multilayer, in which the five layered convolutional neural network pre-trained on Million Song dataset is used to extract the features from EmoMusic dataset. Considering the task of multimodal music mood prediction based on the audio signal and the lyrics of a track, Delbouys et al. [4] reproduce the implementation of traditional feature engineering-based approaches and propose a new model based on deep learning.

Extensive research has been conducted on the detection of music mood, contributing to the music recommendation aspect of our study.

3. Preliminary Field Study

To identify challenges in bar music recommendations, the first author conducted on-site surveys by physically visiting numerous bars, both in China (such as Doman Bar) and Japan (such as English Society Bar near Kyudai-Gakkentoshi Station, The En Bar near Chikuzen-Maebaru Station, etc.), and inquired with bar owners about their respective requirements. We received diverse responses, such as the need for music recommendations aligning with the bar's musical style (Doman Bar), recommendations tailored to the season and weather conditions (English Society Bar), music recommendations in the form of albums (The En Bar). After visiting the bars, we derived several valuable tips and advice for effective bar music recommendations. These include understanding the audience, music genres variety, mood-based playlists, and real-time feedback.

However, we also identified a crucial issue during the visits: for a relatively small-scale bar, the mood within the bar can be considered as a unified entity. Furthermore, the mood within the bar is highly likely to be influenced by the music being played. In other words, patrons within the bar exhibit a stronger preference for listening to music that matches the prevailing mood. Therefore, leveraging the mood of the bar for music recommendations appears to be a feasible approach. We aim to develop a mechanism capable of automatically recognizing the mood within small-scale bars.

An additional clarification is necessary: in this paper, the term 'mood' refers to the atmospheric characteristics of the bar, representing the overall ambiance, and is not

an aggregate of the individual emotions of all individuals within the venue. Therefore, when referring to 'mood detection' in this paper, it pertains to the detection of the ambiance within the venue.

4. A Mood-based Music Recommendation Mechanism

The general structure of the mood detection mechanism, as shown in Figure 2, involves sensors recognizing corresponding data within the bar. These data are then transformed into numerical values, which are input into the weight-based calculation. The output of the weight calculation comprises values for Arousal *At* and Valence *Vt*. Subsequently, these values are used to determine the mood by locating the corresponding coordinates in the two-dimensional emotion space (as Figure 1). Finally, the obtained mood is matched with the mood identified in music mood detection, thereby achieving the desired effect of bar music recommendation.

The following parts will describe several key components within the mechanism.

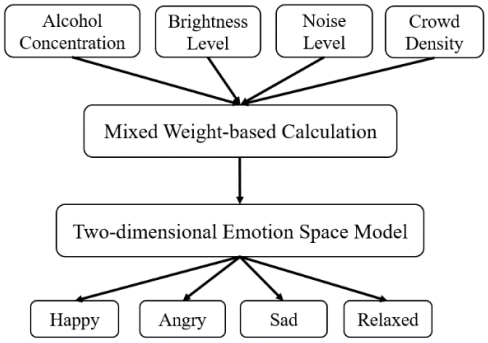

Figure 2. Architecture of the Mood Detection Mechanism

4.1 Sensing Mechanism

Table 1 shows the sensing method and sensing devices in this mechanism. The three sensors mentioned in Table 1 can all be connected to a computer through *Arduino*. For the detection of crowd density, our current plan is to have bar staff manually count the number of people and input this directly into the weight calculation. This could potentially be automated by using different sensor. Alternatively, automated counting of entries and exits can be achieved through devices such as cameras and infrared sensors.

Table 1. Sensing method/device

	Sensing method/device	Manufacturer
Alcohol Concentration	MQ-3 Gas Sensor	SparkFun
Brightness Level	Grove – Light Sensor	Seeed Studio
Sound Level	Grove – Sound Sensor	Seeed Studio
Crowd Density	Manual Counting	

4.2 Weight-based Calculation of Mood Detection

The inputs for the weight calculation comprise four parameters: alcohol concentration, brightness level, sound level, and crowd density, with values ranging from -50 to 50. Alcohol concentration and crowd density have a higher weight in Valence, while brightness level and sound level have a higher weight in Arousal. Additionally, weights can be adjusted according to specific user preferences and the characteristics of the bar. For instance, in a bar that places significant emphasis on the musical ambiance, the weight for noise level can be appropriately increased. An initial example of the weight calculation is shown in Table 2.

Table 2. An example of the weight calculation

	Data Value	Arousal Weighting Factor (w_a)	Valence Weighting Factor (w_v)	Arousal Value (At)	Valence Value (Vt)
Alcohol Concentration	40	0.1	0.3	4	12
Brightness Level	-10	0.4	0.2	-4	-2
Sound Level	-30	0.4	0.2	-12	-6
Crowd Density	10	0.1	0.3	1	3
Total		1	1	-11	7

4.3 Music Recommendation

The results of the weight calculation will be the values for both Arousal and Valence, which can be used to approximate the mood within the bar. Moreover, as introduced in the second section, researches in the field of music mood detection commonly employ a four-category classification method (i.e., Happy, Sad, Angry, Relaxed). Therefore, the core of music recommendation is to recommend music that matches the detected mood.

5. Conclusion

In this paper, we discussed music recommendation in bar environment and proposed a mechanism for mood detection within bars and elucidate how to leverage this mechanism for more effective bar music recommendations. We intend to conduct further explorations in our subsequent studies.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Numbers JP20H00622, JP23H03507, JP23K02469.
We are also deeply thankful to The En Bar, English Society Bar, Doman Bar for the visits and interviews.

References

[1] Thayer, R. E. (1989). *The biopsychology of mood and arousal*. Oxford University Press.
[2] Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178.
[3] Bhattarai B, Lee J. Automatic Music Mood Detection Using Transfer Learning and Multilayer Perceptron. IJFIS 2019;19:88-96.
[4] Delbouys R, Hennequin R, Piccoli F, et al. Music mood detection based on audio and lyrics with deep neural net[J]. arXiv preprint arXiv:1809.07276, 2018.