HOW COOL IS BEBOP JAZZ? SPONTANEOUS CLUSTERING AND DECODING OF JAZZ MUSIC

Antonio RODÀ*[1], Edoardo DA LIO[a], Maddalena MURARI[b], Sergio CANAZZA[a]

[a] *Dept. of Information Engineering, University of Padova, Italy roda@dei.unipd.it*

[b] *Dept. of Pharmaceutical and Pharmacological Sciences, University of Padova, Italy*

ABSTRACT

Music is able to arouse and heighten listener's emotions and sensations. However, experimental studies on the connotative meaning of particular music repertoires, such as jazz music, are still scarce. The study uses 20 subjects to evaluate and describe verbally 25 pieces of *jazz* music that belongs to *cool* jazz and *bebop* sub-genres. Three clusters have emerged, which can be related to the well-known valence-arousal emotional space. A further analysis of the acoustic features of the tracks revealed that *bebop* tracks are mainly associated with low valence values that were characterized by a high degree of roughness.

Keywords: *music expressiveness, kansei and music, musical features, cool jazz, bebop*

1. INTRODUCTION

Various experiments have demonstrated that music can arouse the listener's sensations, such as images, colours, feelings, or emotions (Juslin & Sloboda, 2011) (Murari, Rod, Canazza, De Poli, & Da Pos, 2015). In particular, literature on the affective aspects of music describes the [1]relations between musical content and specific affective models such as the discrete emotions

[1] *Corresponding Author*

approach and the valence-arousal plan (Eerola & Vuoskoski, 2011) (Roda, Canazza, & De Poli, 2014). In addition to that, Kansei models were used to study the connotative meaning of music like Sugihara, Morimoto, & Kurokawa, (2004) that have characterized 12 music pieces from various repertoires although not including Jazz, to 40 pairs of Kansei words. Other works tried to find how emotions are related to a specific acoustic and/or musical features or a combination of them as depicted by Yang & Chen, (2012) where it was found that minor mode usually arouses affective states with low valence such as sadness or melancholy, however it is not yet clear whether this is a cross-cultural phenomenon or not.

Despite the large number of studies on this subject, very few are related to the repertoire of jazz music, and most of them are concerned towards the automatic recognition of discrete emotions using machine learning techniques, e.g. (Tao Li & Ogihara, 2004), without discussing if and which state of the art models of emotions in music are investigated. Moreover, since most studies are developed in relation to Western classical repertoire or pop/rock music, it is very difficult to hypothesise which model are more suitable to analyse jazz music.

This paper presents the first experiment of a project that aims at collecting experimental data to characterise jazz music from an affective and sensorial point of view. The objectives of this exploratory study are: a) to find the main categories which listeners apply to differentiate the emotional content of jazz pieces; b) to verify if the well-known valence-arousal model is still suitable to describe emotions in jazz music; c) to find musical-acoustic (computable) features that significantly characterise the different categories and/or dimensions. The experimental approach was proposed by Bigand, Vieillard, Madurell, Marozeau, & Dacquet, (2005), and detailed in the next section, was applied to foster a spontaneous clustering of the musical stimuli, without conditioning it by means of a predetermined list of words, such as in the semantic differential approach. Music stimuli are well-known jazz pieces chosen from the two most important and revolutionary styles since early 1940s as stated by Kernfeld, (2002):

i) *Bebop* (i.e., *bop* or *rebop*, non-sense syllables which were commonly used in *scat* singing): represents a marked increase in complexity and is mostly characterised by a highly diversified texture created by the bass player and elaborated by the drummer, with a variety of on- and off-beat punctuation added by the piano.

ii) *Cool* (i.e., *cool players*, often white musicians, named for their light, clear touch): jazz style played almost with no vibrato, placing great emphasis on simplicity and lyricism in improvisation and avoiding the upper register of the musical instruments.

2. EXPERIMENT

2.1. Participants

The experiment involved a total of 20 participants (14 males and 6 females). Of these, 11 did not have any musical training and are referred to as non-musicians; 9 had been music students for at least five years and are referred to as musicians. The participants were from 18 to 30 years old, with an average of 22 years.

2.2. Material

25 musical excerpts† were chosen as follows: 12 pieces were taken from the bebop genre; the other 13 pieces were chosen from the cool genre. The excerpts were chosen to be representative of various compositional styles and musical ensembles. Differently from their usual characteristics, some bebop pieces were chosen in order to convey a melancholic and relaxing mood (e.g., *Delilah*, by Clifford Brown, 1954, from *Brownie: the Complete Emarcy Recordings* – 1989) and some cool tunes were chosen in order to convey a happy and dynamic mood (e.g., *Jazz of Two Cities*, by Warne Marsh, 1956, from *Jazz of Two Cities, Complete 1956-1957 sessions* – 2004): in this sense, a verbal description would be very complex although a spontaneous clustering should achieve the objectives (a), (b) and (c) listed in Sect. 1. The excerpts correspond either to the beginning of a musical movement, or to the beginning of a musical theme or idea, and their average duration is 30s. The overall amplitude of each stimulus was adjusted by normalizing the maximum RMS value, in order to ensure a uniform and comfortable listening level across the experiment.

2.3. Procedure

A software interface (see Figure 1) has been developed to conduct the experiment. Participants were presented with a visual pattern of 25 loudspeakers, representing the 25 excerpts in a random order, automatically changed for each subject, in order to avoid biasness due to order effect. Participants were first required to listen to all these excerpts and to focus their attention on the affective quality of each piece. Then, they were asked to look for excerpts that induced a similar emotional experience and drag the corresponding icons in order to group these excerpts. They could listen to the excerpts as many times as they wished, and to regroup as many excerpts as they wished.

After the grouping task, participants were asked to spontaneously describe the affective characteristics of each group, by means of one or two words that were annotated on a questionnaire. This spontaneous decoding task is intended to help and guide the following clusters interpretation. The overall duration of the test was 30 minutes on average and the nature of the stimuli which are real music recordings and not artificial stimuli and ensure that fatigue effect is negligible, as confirmed by previous studies (Bigand et al., 2005) and by informal post-test interviews.

† A detailed list of the pieces with the relative audio files can be found at http://dei.unipd.it/~roda/emojazz/index.html

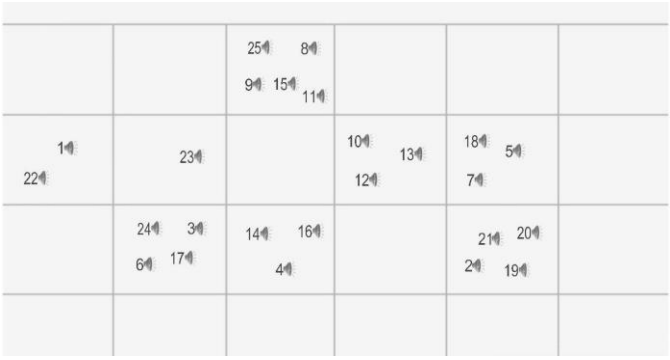

Figure 1: A screenshot of the GUI developed for the experiment.

3. RESULTS AND DISCUSSION

Participants formed an arbitrary number N of groups. Each group G_k contains the stimuli that a subject thinks similar that induces a similar affective experience. The dissimilarity matrix A is defined by counting how many times two excerpts i and j are not included in the same group:

$$A[i,j] = \begin{cases} A[i,j]+1 & \text{if } i \in G_k \wedge j \notin G_k \\ A[i,j] & \text{otherwise} \end{cases}$$

$\forall\, i,j = 1,...,25$ and $\forall\, k = 1,...,20$.

Initially, two different matrices, one for the musicians and the other for the non-musicians subjects, have been calculated. The two matrices present a high correlation value (r = .56, df = 298, p < .001), implying a high agreement between musicians and non-musicians. Then, the following results are based on a unique matrix that includes the responses of both groups.

The dissimilarity matrix was analysed by using the Multidimensional Scaling (MDS) method. In particular, given the non-metric nature of the dissimilarity matrix, the Kruskal's Non-metric Multidimensional Scaling method is adapted where a widely used ordination technique is applied. The quality of the fit of the regression was used to determine the number of dimensions to be considered. According to literature, a Kruskal's *Stress 1* greater than 0.2 indicate an insufficient adaptation of the data in relation to the number of selected dimensions. In our case, a *Stress 1* = 0.17 was obtained with two dimensions, indicating that two axes are sufficient for a good representation of our experimental data. The location of the 25 excerpts along the two principal dimensions is represented in Figure 2. The excerpts that are close in this space are those evaluated by the subjects to be more similar from an affective point of view.

The MDS solution was compared with a cluster analysis performed on the same dissimilarity matrix. The *k*-medoids algorithm was adapted and compared to the more common *k*-means

algorithm, is more robust to noise and outliers, and is able to work with an arbitrary matrix of distances between data points. Therefore, in order to decide the appropriate number of clusters and the reliability of the clustering structure, a set of values called *silhouettes* was computed. The average values of the silhouettes *S*, calculated for *k* (number of clusters) from 2 to 7, show that three clusters obtained the greatest value (*S* = 0.28) and is therefore the best choice (Figure 2).

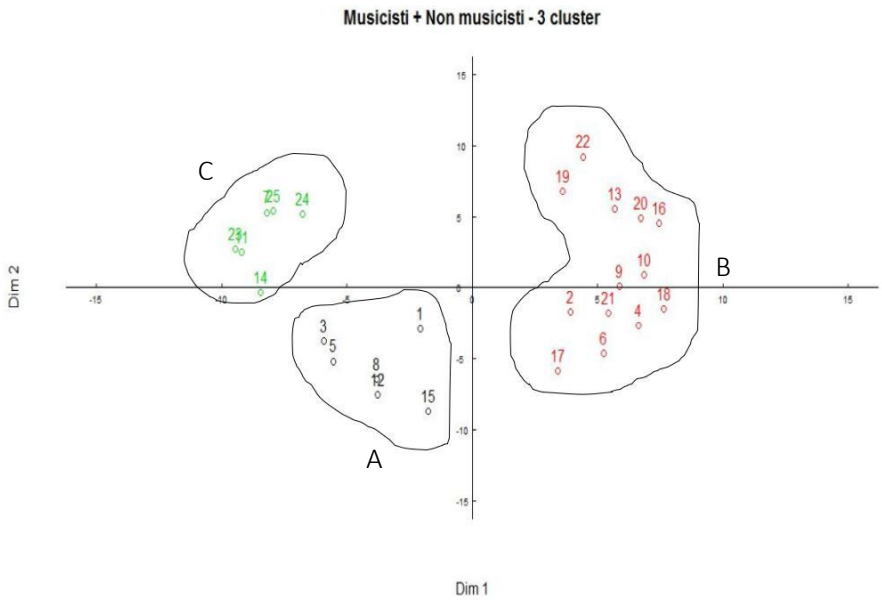

Figure 2: MDS analysis on experimental data. The colours that represents the result of the cluster analysis (black = A; red = B; green = C).

Furthermore, in order to investigate the affective meaning of the three clusters of Figure 2, the verbal responses given during the spontaneous decoding task were analysed. Data preparation of spontaneous free report terms to describe groups was based on the procedure adopted by (Augustin, Wagemans, Carbon, 2012). Spelling errors were corrected, articles for nouns and qualifiers were removed, different spellings and same-stemmed words were pooled. The word count was conducted separately for each cluster and the three most frequent terms are listed in Table 1.

Table 1: list of the three most frequent labels associated by the subjects to the three clusters (in brackets the number of occurrences).

cluster A	cluster B	cluster C
relaxing (29)	happiness (72)	melancholy (19)
happiness (16)	dynamism (57)	relaxing (15)
background (13)	empathy (34)	annoyance (13)

These data can be quite directly related to the valence-arousal plan, widespread in the study of emotions: descriptions of cluster A are related to the quadrant defined by low arousal and high valence (LAHV); cluster B is related to high arousal and high valence (HAHV); cluster C to low arousal and low valence (LALV). Observing the position of the clusters in the plan of Figure 2, it is possible to infer that x-axis is directly related to arousal and y-axis is inversely related to valence.

As the concern of the subdivision between cool and bebop pieces, cluster A is characterised by a predominant presence of cool pieces (5 cool and only 1 bebop). On the contrary, the other clusters are a mixture of the two genres (5 cool and 8 bebop for cluster B, and 3 cool and 3 bebop for cluster C). Moreover, according to the Mann-Whitney test, cool pieces have values on the y-axis (inversely related to valence) significantly lower than the bebop pieces ($U = 36$, $p < .05$). On the contrary, no significant difference can be found along the x-axis (related to arousal). Therefore, following the subjects' responses, the main affective aspect that differentiates bebop from cool pieces is valence, bebop being associated with a more negative valence than cool.

Finally, to correlate the subjects' answers to the musical features of the 25 pieces, a detailed acoustic analysis of the musical stimuli was conducted. A set of acoustic features were computed for each excerpt using the Matlab MIR Toolbox (Lartillot & Toiviainen, 2007). The set was chosen among the features that in previous listening experiments conducted by Juslin, (2001) and Rodà, (2010) were found to be important for discriminating different musical qualities. Table **2** shows the values of the features computed on the 25 pieces of the dataset. An analysis of the variance was carried out to find significant relation between features, clusters, and dimensions. Regarding the clusters subdivision, only *rolloff* (a feature related to the balance between high and low spectral frequencies) has mean values significantly different ($F(2,22) = 4.17$, $p < .05$) between clusters A (5129Hz) and B (5618Hz) and cluster C (2563Hz). Moreover, a significant correlation exists between the position of the pieces along the x-axis and *rolloff* ($r = 0.44$, $t(23) = 2.365$, $p < .05$) and *eventDensity* ($r = 0.40$, $t(23) = 2.113$, $p < .05$); and between the position along the y-axis and *tempo* feature ($r = -0.50$, $t(23) = -2.79$, $p < .05$). Regarding the subdivision between cool jazz and bebop, there is a significant difference ($F(23) = 9.58$, $p < .01$) in the mean value of *rms* ($rms_{cool} = 0.10$, $rms_{bebop} = 0.14$), and in the mean value of *roughness* ($F(23) = 3.37$, $p < .10$), having the bebop pieces a higher roughness ($1.11 \cdot 10^6$) than the cool pieces ($0.61 \cdot 10^6$).

Table 2: features computed on the 25 excerpts used in the experiment.

	brightness	rms	rolloff [Hz]	roughness	zerocross [s^{-1}]	eventDensity [s^{-1}]	lowEnergy	tempo [bpm]
1	0.53	0.09	7314	4.10E+05	812	2.35	0.53	151
2	0.63	0.19	5393	7.57E+05	1241	2.42	0.53	179
3	0.47	0.09	2529	2.57E+05	750	1.67	0.40	181
4	0.57	0.11	6134	1.10E+06	1052	1.29	0.53	109
5	0.53	0.13	5873	7.97E+05	973	2.84	0.55	118
6	0.52	0.10	5812	1.85E+05	908	4.16	0.55	119
7	0.77	0.15	2866	4.79E+05	1277	1.78	0.52	107
8	0.48	0.11	2686	6.52E+05	786	2.67	0.58	132
9	0.61	0.10	6988	2.77E+05	941	1.91	0.55	183
10	0.46	0.17	6085	3.03E+06	942	2.60	0.56	178
11	0.39	0.07	2751	6.81E+05	502	1.91	0.58	119
12	0.55	0.11	9113	7.12E+05	1402	2.74	0.58	162
13	0.50	0.09	8417	4.50E+05	1053	3.29	0.54	115
14	0.52	0.16	3546	5.73E+05	696	1.36	0.49	125
15	0.42	0.09	3257	8.01E+05	765	2.34	0.52	186
16	0.59	0.15	3945	1.81E+06	923	3.33	0.50	133
17	0.47	0.07	7370	2.94E+05	932	3.90	0.53	153
18	0.46	0.07	2464	3.64E+05	715	3.82	0.55	119
19	0.33	0.09	2417	5.16E+05	401	1.83	0.54	103
20	0.62	0.10	7845	3.86E+05	1081	3.28	0.55	129
21	0.34	0.12	1866	1.79E+06	695	3.34	0.55	191
22	0.53	0.19	8295	2.31E+06	1331	1.88	0.57	113
23	0.22	0.15	1197	1.44E+06	485	3.74	0.56	141
24	0.29	0.12	1861	9.80E+05	516	1.96	0.58	119
25	0.62	0.08	3157	1.80E+05	870	1.14	0.67	119

4. CONCLUSION

An experimental study was carried out to gain a deeper insight on the relation between jazz music and emotions. Results show that listeners tend to group the proposed songs according to three expressive categories. The first is described by words such as relaxing and happiness; the second by the words happiness and dynamism; the third by melancholy and relaxing. All these adjectives are directly related to the affective dimensions of valence (melancholy vs happiness) and arousal (relaxing vs dynamism), supporting the hypothesis that the valence-arousal plan could be a good model for this kind of stimuli, although further analysis is needed to confirm this hypothesis. Among the four quadrants of the plane, the one defined by high arousal and low valence is not represented in the data. This result differs from an analogous experiment with stimuli belonging to Western classic repertoire (Bigand et al., 2005). Further experiments are needed to verify if it is a characteristic of jazz music, or if it depends on the specific chosen stimuli. *Rolloff*, *rms*, *eventDensity*, *tempo* and *roughness* are the features that characterise the different affective categories identified by listeners' answers. These results are able to guide the design of systems for automatic emotion recognition of jazz music or can foster the development of affective multimodal interfaces, e.g. (Turchet & Rodà, 2017) and (Turchet et al., 2017). Finally, it is interesting to note that *bebop* pieces are perceived with a lower valence than cool pieces. The relationship between cool-positive valence and bebop-negative valence is consistent with the origin of the two subgenres. As mentioned above, *bebop* was born as a reaction to American musicians of European origin who were getting closer and closer to orchestral jazz. The *bebop* is therefore burdened with feelings of resentment and is generally harsh for the ears of culturally strange people. Future studies could extend the experiment to African-American culture subjects to verify to what extent the bebop-negative valence association has a cross-cultural basis.

REFERENCES

Augustin, M. D., Wagemans, J., & Carbon, C. (2012). All is beautiful? generality vs. specificity of word usage in visual aesthetics. *Acta Psychologica, 139*(1), 187-201.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion, 19*(8), 1113-1139.

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music, 39*(1), 18-49.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (Ed.), *Series in affective science. Music and emotion: Theory and research.* (pp. 309-337) Oxford University Press.

Juslin, P. N., & Sloboda, J. (2011). *Handbook of music and emotion: Theory, research, applications* Oxford University Press.

Kernfeld, B. (2002). *The new grove dictionary of jazz* Grove; London: MacMillan.

Lartillot, O., & Toiviainen, P. (2007). A matlab toolbox for musical feature extraction from audio.

Paper presented at the *International Conference on Digital Audio Effects,* 237-244.

Murari, M., Rodà, A., Canazza, S., De Poli, G., & Da Pos, O. (2015). Is vivaldi smooth and takete? non-verbal sensory scales for describing music qualities. *Journal of New Music Research, 44*(4), 359-372.

Rodà, A. (2010). Perceptual tests and feature extraction: Toward a novel methodology for the assessment of the digitization of old ethnic music records. *Signal Processing, 90*(4), 1000-1007.

Rodà, A., Canazza, S., & De Poli, G. (2014). Clustering affective qualities of classical music: Beyond the valence-arousal plane. *IEEE Transactions on Affective Computing, 5*(4), 364-376. doi:10.1109/TAFFC.2014.2343222

Sugihara, T., Morimoto, K., & Kurokawa, T. (2004). An improved kansei-based music retrieval system with a new distance in a kansei space. 141-146. doi:10.1109/ROMAN.2004.1374745

Tao Li, & Ogihara, M. (2004). Content-based music similarity search and emotion detection. *In Proc of ICASSP 2004*.

Turchet, L., Zanotto, D., Minto, S., Rodà, A., & Agrawal, S. K. (2017). Emotion rendering in plantar vibro-tactile simulations of imagined walking styles. *IEEE Transactions on Affective Computing, 8*(3), 340-354. doi:10.1109/TAFFC.2016.2552515

Turchet, L., & Rodà, A. (2017). Emotion rendering in auditory simulations of imagined walking styles. *IEEE Transactions on Affective Computing, 8*(2), 241-253. doi:10.1109/TAFFC.2016.2520924

Yang, Y., & Chen, H. H. (2012). Machine recognition of music emotion. *ACM Transactions on Intelligent Systems and Technology (TIST), 3*(3), 1-30. doi:10.1145/2168752.2168754