One Micro Song, Three Ends: an approach for musical composition and an interactive decision machine based on expressive live performance.

Rodrigo Schramm[1], Helena de Souza Nunes[1], Leonardo de Assis Nunes[2], Federico Visi[3], and Eduardo Reck Miranda[3]

[1] Universidade Federal do Rio Grande do Sul
[2] Universidade Federal da Bahia
[3] Plymouth University
rodrigos@caef.ufrgs.br, helena@caef.ufrgs.br, assisnunes@gmail.com, federico.visi@plymouth.ac.uk, eduardo.miranda@plymouth.ac.uk

Abstract. A machine system is designed to analyze the musical aspects during the live performance, allowing an interactive and dynamic flow of new expressions and also opening new compositional forms and multimodal methods. The focus of this approach is to measure the expressiveness from distinct characters during the performance of the musical piece while decisions are made by the machine. This multimodal approach is motivated by the Three Chamber Micro Songs composition, where measures of loudness and body motion are used by the algorithm to choose a particular musical end.

Keywords: multimodal system, machine decision, micro song

1 Introduction

The composition Three Chamber Micro Songs - *Essência Pierrot, Atitude Arlequim, (In)Decisão Colombina* is a musical piece for singer, piano and computer, composed based on the compositional method *Microcanções CDG* [18]. This work is a kind of Brazilian chamber song [1], containing many situations of love tensions in its narrative line. They are represented by a romance triangle among *Pierrot - Arlequim - Colombina*, in a contemporary perspective of *Commedia dell'Arte* [19]. These aspects inspired a conceptual map used in the developing process of the musical composition. Also, the Brazilian modernist poem *Máscaras* [5] was an inspiration for the words and linguistic expressions, which, after included in the conceptual map, provided the maturation of a new poetic text. Subsequently, this poetic text was transformed into music by the use of the concept of *Melodias Embrionárias* [11]. These aspects supplied the music composition process, including choreography and character expressiveness. All these elements are formalized in songs of small dimensions, but also represented by a dense amount of information and rich performance possibilities. The piece Three Chamber Micro Songs arises from the labor of the composer, which defines the referential elements and the compositional framework. From this components basis, the interpreters/performers also improvise and create. The story line in this composition is presented in a specific form:

(1) Introduction, which presents the characters; (2) Three micro movements, where each of the characters are put in evidence, indicating their particular way to reach the love; and (3) Coda, where finally their fates are sealed. The characters' destiny is not always the same because it depends on the interrelationship of expressiveness among the characters during the musical performance. The machine will indicate the one who the expressiveness is more prominent, and then a real-time algorithm will place the final Coda to the interpreters. The Coda contains the music score from one of the three possible denouements based on the romantic relationship among the characters. Thus, the machine choice is presented in live to the performers at the last music score page on a screen display. The music score and the artistic interpretation of this version of the piece (2.0) are played by a pianist, a solo singer and a computer. The performers are free to incorporate each of the characters, without the duty to defend or support one of them [17]. This means that the performers should enjoy the relationship among the three characters, expressing their feelings through the dynamics of sound and movement. As the artists aim for unique aesthetic experiences, which are revealed by deep immersion in the musical piece context, each particular musical ending is a consequence of the expressiveness performance of each character along the whole storyline. In this case, it is a consequence of the expressiveness evolution of the performers on the stage.

This work is situated in the vast, interdisciplinary research context of expressiveness in music. There are several instances of analysis aided by computational algorithms as well as examples of artistic practice based on academic research. Camurri and colleagues carried out extensive research on using multimodal data for expressive content analysis and interactive implementations [3], with the particular focus on technologies and methods for the analysis of expressiveness in music and interactive performances involving body movement [2]. In terms of composition, an open-ended structure where multiple endings are selected during the performance through live EEG data processing was used by Miranda and Eaton [7]. Brain Computer Interfacing is also adopted for a variety of different tasks related to expressivity, such as detecting affective states and controlling electronic musical instruments [16]. Classification algorithms and machine learning have also been employed to interpret movement trajectories in musical performance [4], and several automated computer systems for expressive music performance have been developed over the past three decades [12]. In addition, movement and gestures are often taken into consideration in the analysis of musicians' expressivity [10] and the performer's movements are increasingly regarded as an expressive feature intertwined with other musical elements in the score [20]. In the next section, the algorithm and computer model are explained in details. In section 3 is described the experiment realized with the Three Chamber Micro Songs composition, and in the last section are traced some considerations about this developed work.

2 Computing the Character Expressiveness

The proposed machine system is designed to analyze the musical aspects during the live performance, allowing an interactive and dynamic flow of new expressions and also opening new compositional forms and multimodal methods. The focus of this approach is to measure the expressiveness from distinct characters during the performance of a

musical piece. In the end, a decision is made by comparing the amount of expressiveness related to each of possible characters. The decision can have several forms and depends strictly on the compositional process. In this approach, the machine acts as an external evaluator, which regards to several features captured from the musicians, dancers and actors during the live performance on the stage. These features are extracted from audio and video signals by the use of a set of signal processing techniques, which maps three standard measures to the expressiveness hypotheses:

1. RMS - Root Mean Square: reveals a notion of energy used to play a musical instrument or sing. This measure is extracted from the audio signal, and it specially maps the amount of energy used at the attack points of each musical note (onsets) [14]. The RMS gives estimates of the loudness level, revealing the dynamics of the music, which have been widely used as a parameter of expressiveness in several computer music approaches [13,15,9].
2. CI - Contraction Index: quantifies how much is the contraction of the human body during the performance. It is extracted using a motion capture device.
3. QoM - Quantity of Motion: measures the amount of body motion during the musical performance. It is also extracted by a motion capture apparatus. Both CI and QoM measures[4] have been widely used in multimodal embodied approaches to gesture mapping [21,22], including 3D expressive movements [8].

2.1 Mapping the Character Expressiveness

The key idea of this system is to compare the expressiveness of distinct characters during the musical performance. This comparison serves to make choices at the live performance, breaking the linearity of the musical composition. Basically, this approach is a causal system, where the past history provides information for the algorithm to make new choices. Also, the goal of this approach is to catch a variation quantity of expressiveness from a set of musical and choreographic patterns. The measures of expressiveness used in this approach might be strongly affected by extrinsic and intrinsic factors like noise, camera calibration and, especially, by the nature of these measures. In fact, the measures (CI, QoM and RMS) have distinct and incompatible magnitude values, which makes a direct comparison impractical. For instance, two choreograph shapes might have high difference values in CI and QoM measures, but both might convey the same expressive meaning. To overcome this issue, it is proposed a probabilistic model, which each measure is related to a random variable estimated using the data captured from the rehearsal sessions. Thus, instead of using the absolute values, the system implements a function that links the measures (RMS, CI, QoM) to a probabilistic estimate that falls within the interval $[0,1]$. To make the notation easier, henceforth, the RMS, CI, QoM measures will be expressed by the ψ,λ,γ variables, respectively.

The expressiveness is defined by the cumulative distribution function $P_b^{\omega}(\lambda,\gamma,\psi)$, where b specifies the music part (a window time) where the measures are taken. The superscript ω indicates one of the possible targets, i.e., $b_i \in \boldsymbol{B}^{\omega}$ indicates the music parts from where come the extracted measures λ, γ, ψ. Thus, it is possible to define

[4] Details about how the CI and QoM measures are calculated can be found in [8].

distinct ω targets (characters) by defining distinct sets $\boldsymbol{B}^{\omega}$. In this way, the composer can use empirical heuristics to specify points (or parts) in the musical composition where are more convenient to extract the CI, QoM and RMS measures. A set $\boldsymbol{B}^{\omega}$ can be interpreted as the character path. Because the CI and QoM are derived from the body motion, they are inherently dependent[5]. Thus, in this approach, these two variables are measured as joint probabilities, and the RMS estimate is assumed to be an independent random variable. Hence, the probability function can be rewritten in the form (ω is omitted for simplicity):

$$P_b(\lambda,\gamma)P_b(\psi), \tag{1}$$

and the character expressiveness at a point b can be estimated in terms of the cumulative distribution function from the CI and QoM measures:

$$P_b(\lambda,\gamma) = \int_{-\infty}^{\lambda}\int_{-\infty}^{\gamma} f_{\lambda_b,\gamma_b}(x,y)\,dx\,dy. \tag{2}$$

where $f_{\lambda_b,\gamma_b} \sim \mathcal{N}(\mu_b,\Sigma_b)$ is a bivariate normal density function [6] defined by the mean vector $\boldsymbol{\mu}_b$ and the covariance matrix $\boldsymbol{\Sigma}_b$, both estimated from the training data. Analogously, the cumulative distribution function estimated from the RMS measure is given by:

$$P_b(\psi) = \int_{-\infty}^{\psi} f_{\psi_b}(t)\,dt. \tag{3}$$

where $f_{\psi} \sim \mathcal{N}(\nu_b,\sigma_b)$ is a normal density function defined by the mean ν_b and the standard deviation σ_b. The final model is then parameterized by $\boldsymbol{\mu}_b$, Σ_b, ν_b, σ_b, which values are estimated from the data captured at the training phase, after the evaluation of several rehearsals of the musical performance. Figure 1 illustrates the normal density and the respective cumulative distribution function generated from the data obtained on the training phase at a specific time window b.

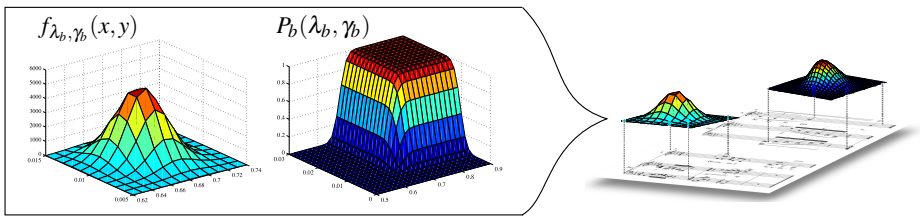

Fig. 1. Normal density $f_{\lambda_b,\gamma_b} \sim \mathcal{N}(\mu_b,\Sigma_b)$, and respective cumulative distribution function estimated from the CI and QoM extracted data.

[5] Despite the resulting sound is produced by the movement of the performer, we decide to treat sound and movement independently. This choice is based on the fact that, in our approach, motion and audio data is captured from distinct and autonomous sources.

2.2 Machine Decision

In previous section, it was described how the real-valued data captured by the sensors are mapped by the functions $P_b^{\omega}(\lambda,\gamma)$ and $P_b^{\omega}(\psi)$ into probabilities (real interval $[0,1]$), which are more meaningful and denote the amount of expressiveness of one character during the musical performance. Still need a procedure that will be used by the computer system to make a decision. This procedure will allow the system to act as a state machine, changing the flow of musical execution or triggering new events. Since there is a direct mapping between the features and the final probabilities, the most expressive music sequence part mE is defined by the average of the estimated probabilities along the set of evaluated points in each path B^{ω}:

$$mE = \underset{\omega}{\operatorname{argmax}} \frac{1}{|B^{\omega}|} \sum_{b \in \{B^{\omega}\}} P(\lambda_b,\gamma_b)P(\psi_b) \tag{4}$$

where B^{ω} is the sequence of points which are evaluated for a specific path (character) ω. It is worth noting that the probabilities estimates on Equation 4 depends only on the set of music points belonging to B^{ω}. Thus, the classification depends specifically on the set of points used to represent the character. In fact, one central idea of this approach is the uncomplicated possibility to map the expressiveness of each character along the musical piece.

3 Experiment

Aiming to illustrate some of the revealed possibilities by this approach, it will be described in this section the use of the proposed algorithm in the context of the musical work Three Chamber Micro Songs. In this approach, the audio signal is captured by microphones that are located in strategic positions on the stage. The gestures, performed by the singer[6], are captured by an RGB-D camera sensor (Microsoft Kinect). Using this configuration, the data was extracted from several rehearsals of the piece. At each rehearsal session, the motion and audio signal values, which are continually evaluated along the whole musical piece, were stored into a training database. The algorithm maps the set of features that are previously extracted from the skeleton tracking and the audio signal (CI, QoM and RMS) into probabilities, as described in the Section 2.1. These features are obtained from several points of the musical performance and are represented by B^{ω}, where $\omega \in \{a,p,c\}$ means the characters *Arlequim*, *Pierrot* and *Colombina*, respectively. Each set B^{ω} is represented by a notation scheme on a particular staff in the music score, allowing the composer to experiment different combinations of paths along the musical performance. Figure 3 shows an excerpt of the music score used in this example. How can be seen in this figure, the computer algorithm has its own staff, where are placed the temporal windows of evaluation as musical notes. Each musical note is a $b \in B^{\omega}$ and each line/space in the "Machine" staff (the note pitch) is a path ω connected to a character. On Figure 3 is shown one example of the acquired data from

[6] In this version of the system, only one body skeleton was used by the algorithm. Nevertheless, other bodies can be used if required.

eight distinct temporal windows along the musical piece, while one performance session. In this experimental approach, it was used an electronic metronome to synchronize the musicians and the computational system.

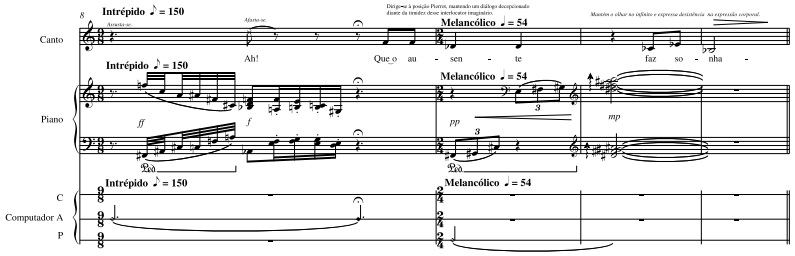

Fig. 2. Excerpt of the music score of the Three Chamber Micro Songs composition. There is a staff "Machine", which specifies the points where the CI, QoM and RMS measures will be taken.

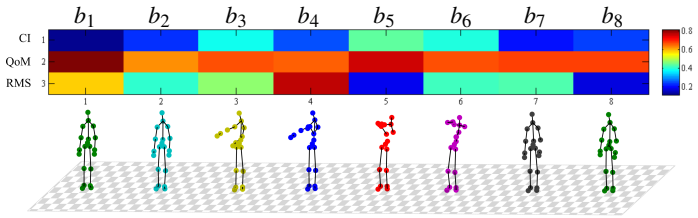

Fig. 3. Example of input data used to train the proposed model. The grid on the top shows the CI, QoM and RMS measures at each window b_i, recorded during one performance session (the data was normalized for better visualization). Full skeleton bodies from the central frame of each temporal window are illustrated on the bottom.

The notation on the music score for the proposed example generates the sets: $B^p = \{b_1, b_4, b_8\}$, $B^a = \{b_2, b_5, b_7\}$, $B^c = \{b_3, b_6\}$, which define the path of evaluation of the characters *Arlequim*, *Pierrot* and *Colombina*, respectively. For each temporal window b_i were extracted the average and the standard deviation of the CI, QoM and RMS measures. Table 1 shows all estimates computed for the Three Chamber Micro Songs composition.

3.1 Character Expressiveness Evaluation

In this evaluation process, the samples in the training database were labeled according to the expressiveness observed by an external committee (human evaluators). The expressiveness analysis was made considering the set of temporal windows and also were taken into account features which include the contraction and expansion of the singer body, as well the loudness of the musical performance. Thus, the samples in the dataset were labeled as one of the possible targets: *Arlequim*, *Pierrot* or *Colombina*. After, the model was trained with the entire dataset and each sample was subsequently classified using the proposed system. Finally, the machine output was compared with the labels proposed by the external committee. Despite the low number

Window of Evaluation	Statistics from Audio and Motion Capture Data							
	Position (in seconds)		CI		QoM		RMS	
	Start	End	μ	σ	μ	σ	μ	σ
b_1	0	5	81.357	2.549	3.336	1.255	0.054	0.038
b_2	6	8	80.895	4.166	3.122	3.330	0.047	0.030
b_3	9	13	81.361	2.113	2.503	1.154	0.048	0.032
b_4	22	51	69.292	2.570	7.984	1.372	0.081	0.051
b_5	56	80	68.225	3.262	14.982	3.900	0.085	0.052
b_6	82	103	70.426	4.389	14.982	4.759	0.062	0.038
b_7	104	112	74.652	5.195	8.567	4.312	0.054	0.035
b_8	114	125	66.451	4.537	8.030	2.770	0.036	0.020

Table 1. Average and Standard Deviation from the captured data. Each line represents one window time of evaluation, including the CI, QoM and RMS measures. This table is used to estimate the normal densities required by the Equations 2 and 3.

of examples, these preliminary results have shown that the system classification was coherent with the evaluation of the committee on more than 70% of the trials. This means that the results indicate the probabilistic model is working as initially desired. It is already planned as future work an expansion of the training database, including more rehearsal sessions and also other musical compositions with similar configuration. The music score and the recorded data used on the experiments can be accessed at `http://inf.ufrgs.br/~rschramm/projects/music/apc`.

4 Conclusion

The form of the musical composition *Three Chamber Micro Songs* includes an introduction, three movements, and a Coda. On each movement, a same singer acts as one of the three characters: *Pierrot*, *Arlequim* or *Colombina*. The piece has also three possible ends, each one corresponding to a different character, according his/her expressive power in the performance. The choice of the last part is not made by the performers on the stage, but by an external evaluator machine who measures the degree of expressiveness between these characters. This is an ideal artistic condition for the use of multimodal features and a machine learning system. The described approach uses RMS, CI and QoM as measures of expressiveness variation during the performance of the musical piece. A probabilistic model maps the expressiveness variation into probabilities values, which are more suitable to integrate distinct sort of measures from this multimodal analysis. In this case study, experiments had shown the system responses agreed with the expert committee opinions on more than 70% of the trials. In future work, it is planed to be incorporated other audio and visual features to the system. Furthermore, aiming to improve the final decision algorithm, it is also intended a deep study about the correlation between those features and the notion of expressiveness.

References

1. BORÉM, F.; CAVAZOTTI, A.: Entrevista com Luciana Monteiro de Castro, Mônica Pedrosa e Margarida Borghoff sobre o Projeto “Resgate da Canção Brasileira”. Per Musi (15), 78–86 (Jan 2007)
2. Camurri, A., Volpe, G.: Multimodal analysis of expressive gesture in music performance. In: Solis, J., Ng, K. (eds.) Musical Robots and Interactive Multimodal Systems, pp. 47–66. Springer Berlin Heidelberg

3. Camurri, A., Volpe, G., De Poli, G., Leman, M.: Communicating expressiveness and affect in multimodal interactive systems. IEEE Multimedia 12(1), 43–53 (Jan 2005)
4. Caramiaux, B., Wanderley, M.M., Bevilacqua, F.: Segmenting and Parsing Instrumentalists' Gestures. Journal of New Music Research 41(1), 13–29 (Mar 2012)
5. DEL PICCHIA, M.: Mascaras: o amor de Dulcinéia. Ediouro, Rio de Janeiro (1987)
6. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification (2nd Edition). Wiley-Interscience (2001)
7. Eaton, J., Miranda, E.: Real-Time Notation using Brainwave Control. In: Sound and Music Computing Conference 2013. Stockholm, Sweden (2013)
8. Fenza, D., Mion, L., Canazza, S., Rodà, A.: Physical movement and musical gestures: a multilevel mapping strategy. In: Proceedings of Sound and Music Computing Conference. Salerno (2005)
9. Gang, R., Lundberg, J., Bocko, G., Headlam, D., Bocko, M.: What makes music musical? a framework for extracting performance expression and emotion in musical sound. In: Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), 2011 IEEE. pp. 301–306 (Jan 2011)
10. Glowinski, D., Baron, N., Grandjean, D., Ott, T., Shirole, K., Torres-Eliard, K., Rappaz, M.A.: Analyzing expressive styles and functions of bodily movement in violinist performance. Proceedings of the 2014 International Workshop on Movement and Computing - MOCO '14 pp. 154–155 (2014)
11. KIEFER, B.: Elementos da linguagem musical. Movimento, Porto Alegre, 3 edn. (1979)
12. Kirke, A., Miranda, E.R.: A survey of computer systems for expressive music performance. ACM Computing Surveys 42(1), 1–41 (Dec 2009)
13. Langner, J., Goebl, W.: Visualizing expressive performance in tempo-loudness space. Computer Music Journal 27, 2003 (2003)
14. Lartillot, O., Toiviainen, P.: Mir in matlab (ii): A toolbox for musical feature extraction from audio. In: Dixon, S., Bainbridge, D., Typke, R. (eds.) ISMIR. pp. 127–130. Austrian Computer Society (2007)
15. Mion, L., De Poli, G.: Score-independent audio features for description of music expression. Audio, Speech, and Language Processing, IEEE Transactions on 16(2), 458–466 (Feb 2008)
16. Miranda, E.R., Castet, J.: Guide to Brain-Computer Music Interfacing. Springer (2014)
17. NUNES, Helena de Souza; NUNES, L.d.A.: Três microcanções de câmara – essência pierrot, atitude arlequim, (in)decisão colombina (partitura) (2015)
18. de Souza Nunes, H.: A canção brasileira infantil na perspectiva da ficha cdg para análise e composição de canções. Revista Brasileira de Estudos da Canção 1(1), 151–173 (Jan 2012)
19. VENDRAMINI, J.E.: A commedia dell'arte e sua reoperacionalização. Trans/Form/Ação 24(1), 57–83 (2001)
20. Visi, F., Coorevits, E., Miranda, E., Leman, M.: Effects of different bow stroke styles on body movements of a viola player: an exploratory study. In: Proceedings of The joint ICMC|SMC|2014 Conference. Athens, Greece (2014)
21. Visi, F., Schramm, R., Miranda, E.: Use of body motion to enhance traditional musical instruments. In: Caramiaux, B., Tahiroglu, K., Fiebrink, R., Tanaka, A. (eds.) Proceedings of the International Conference on New Interfaces for Musical Expression. pp. 601–604. Goldsmiths, University of London, London, United Kingdom (June 30 – July 03 2014)
22. Visi, F., Schramm, R., Miranda, E.R.: Gesture in performance with traditional musical instruments and electronics: Use of embodied music cognition and multimodal motion capture to design gestural mapping strategies. In: Bevilacqua, F., Alaoui, S.F., Françoise, J., Pasquier, P., Schiphorst, T. (eds.) International Workshop on Movement and Computing, MOCO '14, Paris, France, June 16-17, 2014. p. 100. ACM (2014)