FIRST STEPS TOWARDS A DIGITAL ASSISTANT FOR PERFORMERS AND STAGE DIRECTORS

Alain Bonardi, Isis Truck

FIRST STEPS TOWARDS A DIGITAL ASSISTANT FOR PERFORMERS AND STAGE DIRECTORS

Alain Bonardi
MSH-Paris Nord - UMS 2553
Groupe de travail "Visualisation du son"
de l'AFIM
4, rue de la Croix-Faron
93210 La Plaine Saint-Denis
alain.bonardi@wanadoo.fr

Isis Truck
MSH-Paris Nord - UMS 2553
Groupe de travail "Visualisation du son"
de l'AFIM
4, rue de la Croix-Faron
93210 La Plaine Saint-Denis
truck@ai.univ-paris8.fr

ABSTRACT

In this article, we present the first steps of our research work to design a Virtual Assistant for Performers and Stage Directors. Our aim is to be able to give an automatic feedback in stage performances.
We collect video and sound data from numerous performances of the same show from which it should be possible to visualize the emotions and intents or more precisely "intent graphs". To perform this, the collected data defining low-level descriptors are aggregated and converted into high-level characterizations. Then, depending on the retrieved data and on their distribution on the axis, we partition the universes into classes. The last step is the building of the fuzzy rules that are obtained from the classes and that permit the detecting of emotion states.

1. INTRODUCTION

Thinking of the collaborative work of stage directors with their assistants, we are working on a research project to design a virtual assistant for performers that gives feedback to performers and stage directors shortly after the sounding of a part. This is all the more useful as productions contain a part of improvisation. Performers of these productions need to be directed since the improvisations always follow some rules, and a computerized assistant could provide a strong basis of work and parameters within a variable environment.

Thus, computers may probably be of great help in assisting the stage director and/or the performer. Indeed, we think it is important to conceive several tools to help the stage director in his task of actors' performance supervision. The tool we propose here is a kind of assistant that gives a visual representation – through a graph – of a set of complex data for the exploration and observation of a show. It also permits to understand better the creation and execution of the show. One important point is that, as every computer program, the assistant must be deterministic and systematic, i.e. it should always give the same results for a given entry and it must look over the whole data. Like high-level sportsmen that have tools to analyze, correct and improve their gesture, we want to propose a tool to assist the creative artists, to let them better understand the phases of a performance and to help them in their creation process.

For our practical experiments we have chosen to work on a digital opera called *Alma Sola* written by Bonardi (music composer) and Zeppenfeld (librettist) where a performer plays (sings and moves) different blocks from different *universes* (such as Prologue, Love, Pleasure, etc.). *Alma Sola* is an open form of opera [1].

Figure 1. Singer Caroline Chassany performing Faust in *Alma Sola* (photography by Philippe Monges).

The performer embodies a feminine Faust and wanders through the various *universes* split into blocks. She therefore interprets an opera playlist that she selects during the show itself. For instance, a performance can be : Love-3, then Wealth-5, then Pleasure-3, etc. Thanks to Hidden Markov Models, the computer offers continuations to the performer and suggests the next block to be performed (cf. figure 2)[1].

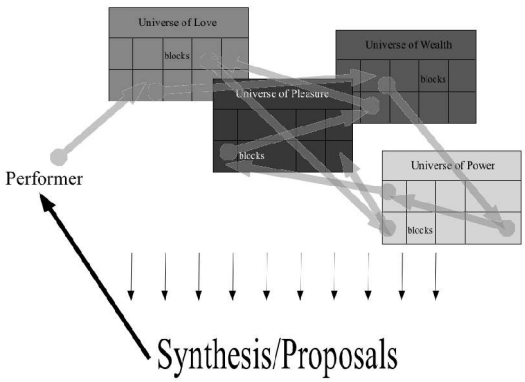

Figure 2. Navigation in *Alma Sola* digital opera.

[1]http://www.musimediane.com/Bonardi-analyse/almaSolaInit.htm

The organization of this opera in blocks therefore fits the requirements of our assistant, enabling to perform and capture small but significative fragments several times with various intents. Another reason for this choice is that it is relatively easier to work with operas than with theatre plays, since time and expression is generally driven by music and therefore rather simple, whereas theatre can use such very subtle expressions as face movements, or silent and slow gestures.

We examine the actor's performance on video and audio files and look for emotions in the data. A distinction must be set between intent and emotion: indeed intents correspond to the conscious part of the emotion. Thus the assistant enables the comparison between the performer's intent and the rendered emotions in the local context of *Alma Sola*. It is widely known that fuzzy logic offers good tools to deal with such subjective concepts [10], emotions here, this is why we shall use a fuzzy rule-based system (FRBS) to detect the performer's emotions.
The article is organized as follows: first we give an overview of the existing research about assistants in art, then we describe our assistant and notably the way we retrieve emotion descriptors. In the third section we show how the descriptors are partitioned into classes. This step is necessary to build correctly the fuzzy rules that form the inference system. Finally, section 4 concludes this study.

2. RESEARCH ABOUT PERFORMER'S ASSISTANTS

In a way, research about performer's assistants has existed for centuries. One immediately thinks of the use of mirrors in dance. At the time when mirrors become commonplace in Europe (Renaissance), the first treatise about dance is released by Thoinot Arbeau in 1589. The issues raised are still the same :

- first point, to be able to state an ideal prescription of the performance. This generally starts from scores and notations, which have been developed for centuries in music and dance. They include both implicit gestures (you have to play an F, but the score does not tell how to do it) and explicit gestures (cross hands when playing the piano, for instance) to be achieved. From these structured indications, a dancer or a musician tries to infer some of the author's intentions[5]. His/her representation of the author's intentions become the ideal prescription of the performance. In an opera production, this prescription is completed by the stage director's indications.
- second point, to be able to measure a kind of difference between the realized performance and its ideal prescription. For instance, in his approach of "Virtual scores", Manoury has considered [7] in his pieces for solo instrument and live electronics the computation of this difference (*Jupiter* for solo flute and live electronics, 1987, *En Echo* for voice and live electronics, 1991) as a basis to generate electronic sounds.
- third point, which is correlated to the second one, is the method to measure this difference. The first approach consists in directly measuring various aspects of the performance, using captors in a broad meaning [11]. Various captors are nowadays available: video camera, wireless microphone, ultrasound device, carpet detectors, digital compass, etc. The second approach consists in collecting human appreciations of this difference between the prescription and the realization. Composer Roger Reynolds has for instance imagined psychological testing [9] with listeners for his piece *The Angel of Death* (2001).

From a technical point of view, dedicated software platforms have been developped. Recently, Camurri and his team have developed the first robust platform for the analysis of gestures and consequently of performer's emotions. It is named EyesWeb [3], [6]. It is not based on captors implemented on the performer's body (with heavy batteries and radio transmission), but on video capture with a static shot. It is based on a graphical language that implements many descriptors of gestures: quantity of motion, stability, etc. EyesWeb has become a worldwide standard for performance analysis.

Roughly ten years before, the Ircam institute (and Cycling 74 company) had developed the famous real-time digital sound analysis and synthesis platform to complete Max software. It is named MSP (Methods for Sound Processing) and is now included in Max software (Max/MSP). This software is a worldwide standard in real-time sound analysis. At the present time the state-of-the-art consists mainly in inferring a few emotional states from the raw data delivered by EyesWeb and/or Max/MSP.

For instance, a project relatively similar to ours is developed by Friberg and his team. They have conceived a real-time algorithm to analyze emotional expression in musical performance and body movement [4]. In the framework of a game named "Ghost in the cave", the player has to express different emotions using his/her body or his/her voice, and these emotions are the input values of the software. They use EyesWeb to recover body movements and sound descriptors (sound level, instant tempo, articulation, attack rate, high-frequency content). But in this game, there is an immediate feedback, and the player has to move constantly and talk until the software reacts according to his/her wishes. Our aim is not the same, i.e. the assistant adapts to the performer and not the contrary.

3. PROJECT DESCRIPTION

As explained above, we have chosen to work on the interactive opera *Alma Sola* written by Bonardi and Zeppenfeld. Two dissimilar scenes have been extracted to be used as "sample scenes": they are the chanted Prologue (which is improvised, unwritten) and the Love Universe (which is "strictly written"). The performer

(singer Claire Maupetit here) is filmed by a camera in wide and static shot (cf. figure 3, left) and his/her voice is recorded in a separate file, in order to handle both sources of data separately. The project is centered on two main phases: the acquired scene processing (with the performer) and the sound capture processing. In this article we focus on the first phase, where EyesWeb has been used (through a dedicated patch we have written) to analyze accurately, understand and exploit non-verbal expressive gestures. Several parameters can be extracted from the video file: quantity of motion, stability, motion duration, pause duration in a scene, contraction index and surface of the performer in the image, convex hull of the body silhouette, velocity, acceleration, etc. This patch construction step is not trivial and represents hours of tests with sample videos and it implies a concertation with the stage director.

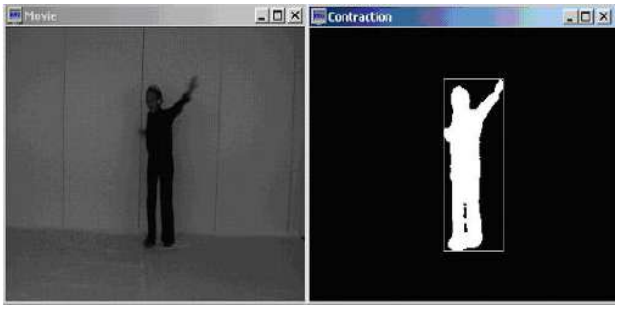

Figure 3. Left, an image taken form the video file; right, the corresponding convex of the body silhouette (performer : Claire Maupetit).

First of all, the background of the image must be removed — using the difference between two frames — in order to keep only the performer's movement, since the camera is static. Next, the parameters can be easily extracted. Here is the description of some of the most interesting parameters for our problem :

- the quantity of motion is computed by the number of pixels changing position between two instants (the *white pixels* in figure 3);
- the convex hull of the body silhouette is the bounding rectangle of the *white pixels* (cf. figure 3);
- the stability is the ratio of the height of the silhouette's center of gravity on the length of the segment connecting the lower points of the silhouette;
- the contraction index is the ratio of the silhouette's surface over the surface of the convex hull.
- the stability is an important descriptor since it gives good insight on whether the performer is near the ground or not, if the performer puts himself at risk or not.
- the contraction index is also very useful and reflects whether the performer is effusive or not.

Choosing these parameters judiciously (called video descriptors) allows us to compute various aggregations of each set of values for each descriptor. The chosen aggregators are: partial and general means, standard deviation, covariance, etc. Then, to a meta-level, we characterize and categorize the sequences of each scene thanks to an FRBS. The number of categories used in this opera is five (Sleepy, Angry, Happy, LoveBeliever (could be also called Effusive), LittleEffusive). Figure 4 sums the whole process up. (We capture 25 frames per second, i.e., 25 values per second for each gesture descriptor.)

As can be seen in figure 4, once the descriptors are extracted from the video files, they are aggregated in order to give a description of each emotion we want to recognize. Figure 5 shows graphically the results for the Universe of Love.

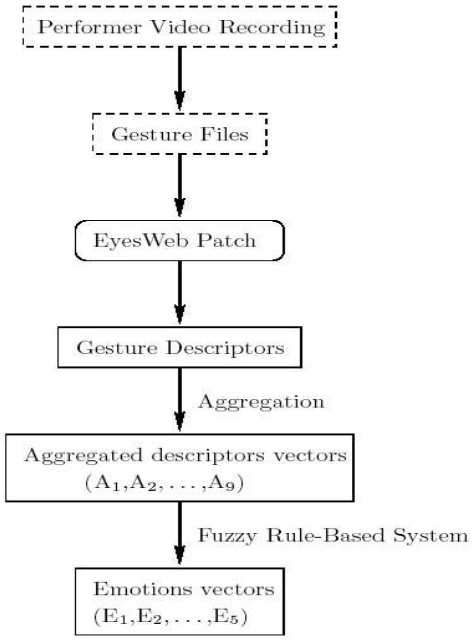

Figure 4. Our system components.

The next step is to classify in partitions the aggregation results. For example, when the performer expresses an emotion such as happiness, he/she moves a lot. But this is not always easy to guess even if the emotions share some general patterns, event in the restricted context of *Alma Sola*.

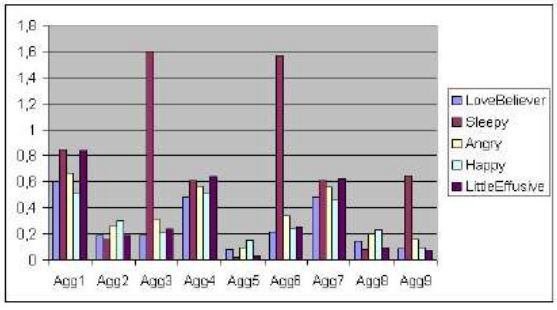

Figure 5. Aggregated descriptors for each emotion in the Universe of Love.

For sounds, we have achieved the same process. During capture sessions, the performer receives the musical accompaniment thanks to wireless headphones (cf. figure 3). Her solo voice is recorded. We first

extract sound files from captures, then compute descriptors thanks to a Max/MSP patch (cf. figure 6), then aggregated sound descriptors, and finally calculate emotion vectors, using a FRBS. The Max/MSP patch is mainly based on Tristan Jehan's *Analyzer* object[1].

In both cases, for video and sound computation and extraction, we come to an important notion that is the division of time between movement/fixed phases, and silent/sung phases. The aggregators are computed globally for each fragment, but also computed separatedly for movement and fixed phases, and silent and sung phases.

The fuzzy partitioning allowing the construction of the rules is now described.

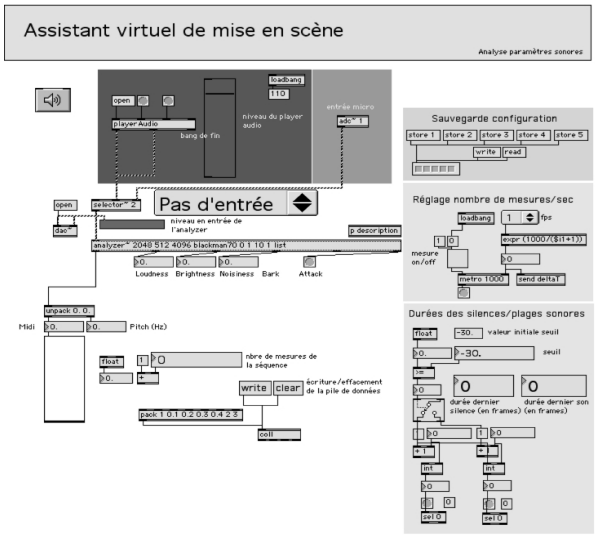

Figure 6. The Max/MSP patch used to compute sound descriptors from performer's voice.

4. BUILDING THE FUZZY RULES

Fuzzy partitioning needs to spread the various values taken by the aggregation results suitably. However a simple uniform distribution of the classes on the axis is not appropriate since sometimes small, other times average or in other cases big variations lead to an emotion change, depending on the aggregated vectors. Like Martinez & al. in [8], we propose a categorization depending on the data distribution. First of all, five classes are considered: Very Low, Low, Average, High and Very High values (denoted *VL*, *L*, *A*, *H*, *VH*). For the moment, we consider the classes as simple intervals. The fuzzy sets construction will be explained below. For each aggregator and for all the records of each scene, the building of class Low is dynamically performed as follows: the interval's left bound is the minimum value of the whole data denoted $x0$. This way, when analyzing a new record of the same scene, if there are smaller values than $x0$, they will be categorized as Very Low values. In the same way, for each aggregator and for all the records of each scene, the building of class High is dynamically performed as follows: the interval's right bound is the maximum value of the whole data denoted $x2$. Then, for each aggregator and for all the records of each scene, the average value of the whole data is denoted $x1$. Both intervals $[x0, x1]$ and $[x1, x2]$ are split into three equal sub-intervals, each. Finally, the interval for class *L* is the concatenation of the first two sub-intervals, the interval for class *A* is the concatenation of the next two sub-intervals and the interval for class *H* is the concatenation of the last two subintervals. Figure 7 shows the intervals split.

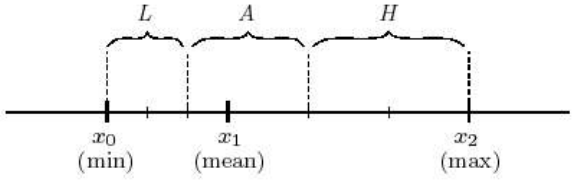

Figure 7. The split intervals for the five classes.

Thus it is possible to classify the aggregation values for the five emotions, as shown in table 1 where A_i is the i^{th} aggregator, in the case of the Love Universe.

Emotions	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9
LoveBel.	*L*	*A*	*L*	*L*	*A*	*L*	*L*	*A*	*L*
Sleepy	*H*	*L*	*H*	*H*	*L*	*H*	*H*	*L*	*H*
Angry	*A*	*H*	*L*	*A*	*A*	*L*	*A*	*H*	*L*
Happy	*L*	*H*	*L*	*L*	*H*	*L*	*L*	*H*	*L*
LittleEff.	*H*	*L*	*L*	*H*	*L*	*L*	*H*	*L*	*L*

Table 1. Classification for the Universe of Love.

The next step consists in building the fuzzy subsets representing the five classes for the FRBS. Figure 7 shows how this is performed. The smallest class-interval, *L* in our example, is defined as an L-R fuzzy number and sets the construction unit for the other intervals (cf. figure 8).

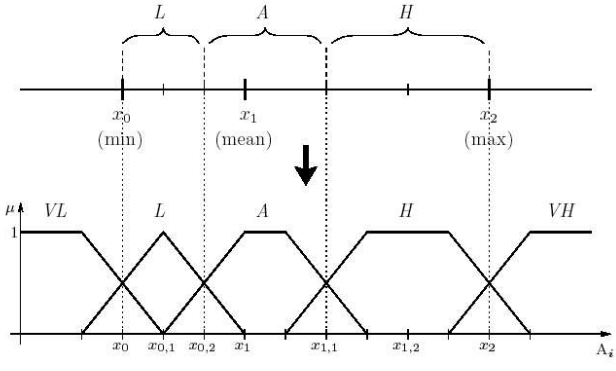

Figure 8. Building the fuzzy classes.

Figure 9 shows a screenshot of our software (see section 5) where the fuzzy subsets used are displayed.

[1]http://web.media.mit.edu/~tristan/

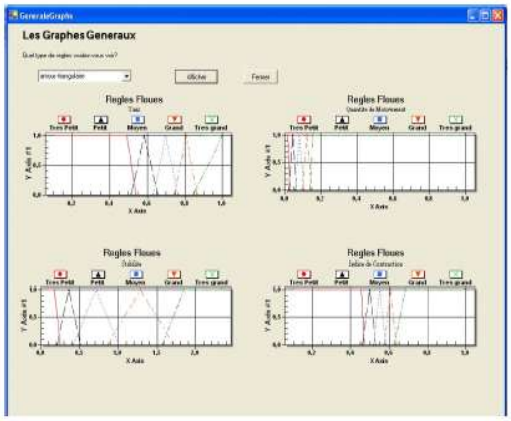

Figure 9. Screenshot of our software (window for fuzzy rules setting)

Finally, it is easy to establish the fuzzy rules according to Table 1, one rule per line. Here is an example for the Universe of Love (figure 10) :

```
Rule 1:
If A1 is L & A2 is A & A3 is L &
   A4 is L & A5 is A & A6 is L &
   A7 is A & A8 is L & A9 is L
Then the emotion is LoveBeliever
```

Figure 10. An example of fuzzy rule.

5. APPLICATION

The application we have developed implements the concepts explained above. We have worked on two universes: Love and Prologue. The performer has played several times each universe and the software has tried each time to detect the emotions perceived. The video files obtained for the performances are split into smaller files that are given to the software. After having chosen the fuzzy subsets that will be used for the partitioning (cf. Section 4), the values of the aggregated descriptors are displayed and a graphical result is also proposed. Figure 11 shows that both Love Universe and Prologue have rather been performed with Sleepy emotion.

An interesting point is that results computed from video and from audio sources converge here.

6. CONCLUSION

In this paper we have presented a system that is able to give clues to the stage director in order to evaluate a performer's rendition. This is done thanks to a fuzzy rule-based system that detects the actor's emotions during a performance show. One originality is the way we construct the fuzzy classes when partitioning the universes before the rule construction. They are dynamically built according to the values characterizing the performance.

As a future work, we will try our assistant on other test sets, i.e. with more records from *Alma Sola* (with the same performer or not) but also with records from other shows. Moreover, it would be very interesting to include a back propagation in the software: when an unexpected emotion is detected, the assistant should suggest modifications of his behaviour to the performer in order to obtain best results during the next detection.

Concerning *Alma Sola* opera itself, we can also imagine in the future that the assistant could be used to classify the blocks performed according to the detected emotions and then contribute to the design of the open form.

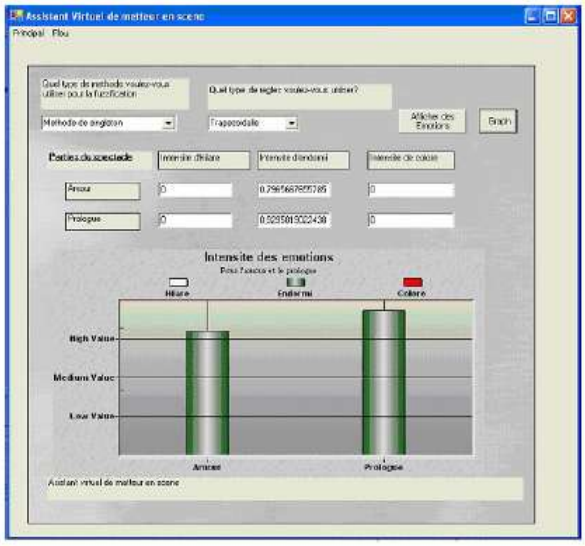

Figure 11. Computation of emotion intensities.

7. ACKNOWLEDGMENTS

We thank Adrien Revault d'Allonnes and Murat Goksedef for their collaboration and help on this work. This project was funded by the Maison des Sciences de l'Homme Paris Nord.

8. REFERENCES

1. A. Bonardi & F. Rousseaux. New Approaches of Theatre and Opera Directly Inspired by Interactive Data-Mining. In *Proceedings of the Int. Conf. Sound & Music Computing (SMC'04)*, pages 1-4, Paris, France, 2004.
2. B. Bouchon–Meunier. *La logique floue et ses applications*. Addison–Wesley, 1995.
3. A. Camurri, M. Ricchetti & R. Trocca. EyesWeb - toward gesture and affect recognition in dance/music interactive systems. In *Proceedings of the Int. Conf. IEEE Multimedia Systems*, Firenze, Italy, 1999.
4. A. Friberg. A fuzzy analyzer of emotional expression in music performance and body motion. In *Proceedings of Music and Music Science*, Stockholm, 2004.
5. K. Kahol, P. Tripathi & S. Panchanathan. Automated Gesture Segmentation From Dance Sequences. In *Proceedingsof the Sixth IEEE International Conference on Automatic Face and*

Gesture Recognition (FGR 2004), pages 883- 888, Korea, 2004.

6. E. Lindström, A. Camurri, A. Friberg, G. Volpe & M.-L. Rinman. Affect, attitude and evaluation of multi-sensory performances. In *Journal of New Music Research*, 34(1): 69-86, Taylor & Francis, 2005.
7. P. Manoury & M. Battier. Les partitions virtuelles. In *Ircam documentation*, Paris, 1987.
8. L. Martinez, J. Liu, Da Ruan & J.B. Yang. Fuzzy Tools to Deal with Heterogeneous Information in Engineering Evaluation Processes. In *Information Sciences*, In Press, 2006.
9. R. Reynolds. Epilog:Reflections on psychological testing with The Angel of Death. In *Music Perception* n ° 22, pages 351-355, 2004.
10. L.A. Zadeh. Toward a generalized theory of uncertainty (GTU) — an outline. In *Information Sciences*, Elsevier, 172 (1-2), pages 1-40, 2005.
11. C. Zeppenfeld. *L'acteur face aux technologies*, Master dissertation, University Paris 3, 2004.