pDaniM: a case study around interactive processes for expressive music generation in the computer animation production pipeline

pDaniM: A Case Study around Interactive Processes for Expressive Music Generation in the Computer Animation Production Pipeline.

João Cordeiro

Escola das Artes, Universidade Católica Portuguesa, Porto, Portugal

***Abstract* — This paper presents an experimental application that shows how an interactive process can save time and production resources in the process pipeline of music scoring for a 3D animation movie. This investigation uses as basis several rules (about 30) that performers use to add certain expressive emotions to a score, covering different aspects of music performance. Basically, applying the rules produces a rearrange in time and velocity parameters such as "microtiming", articulation, tempo and sound level. The resulting application should take as input a basic music score (MIDI) quantized linearly, and imprint expressiveness and emotion to this music in sync with the expressiveness cues from the timeline exported from the animation project. Acoustic cues are driven by facial and gesture expression from the characters.**

***Index Terms* — Expressiveness, Interactive Systems, Multimedia computing, Music.**

I. Introduction

Computer animation is one of the most significant clusters of today's entertainment industry. There is a huge commercial market leaded by major companies that somehow define the standards of the industry. However, this media is getting popular in several other fields such as independent animation films, artistic installations, publicity industry and even in interactive applications like pedagogical agents in tutoring systems or companions in public places like museums. What they all have in common is the complex and time consuming production process involved. Therefore, the use of interactive software tools is regarded in the industry as a valuable method to boost productivity in different phases of the production pipeline. The motion capture technique with infrared technology uses real movements of actors or dancers to animate the characters in the movie. What normally would take months of hard work, by animating frame by frame (12 or more per second), is now simplified. This technique captures real movement coordinates from an actor's performance and uses them to animate a 3D model. Here is a good example how an interactive process can save time and productions resources in the animation pipeline. On the other hand, sound design and music composition for computer animation is also a very challenging field in terms of creativity and technical knowledge requirements. In this type of audiovisual project the psychoacoustic experience of the audience is always completely created from scratch. Since all the visuals are usually computer generated, sound (and especially musical sounds) often assumes an added importance in the achievement of the overall expressiveness conveyed in the piece. Furthermore, in order to accomplish a strong sense of realism and expressiveness in the key dramatic points of a narrative it is fundamental to guarantee a coherent correspondence between the animated characters, created by the animation artists, and the music created by the composer. Also, in this matter it is important to overcome issues that somehow consume energy and time to whom is working on them without improving the final result and that is the goal of this research: find ways to overcome time and money issues, in order to produce creative hi-end solutions for sound design in computer animations. To help with this problem, a software application based on an existing one (both described later), was developed by the author.

Fig. 1. A frame of "O Trovão" initial scene. A 3D animation movie used as case study on this project.

The aim of this research is to start off from the work realized at the KTH Department of Speech, Music and Hearing in the field of musical emotion and expression parameterization [1] and a computer animation movie production (case study) undergoing at UCP called "O Trovão"[1], Porto (Fig. 1) and create an experimental

[1] Barbosa, A. 2008 "O Trovão" (www.abarbosa.org/work.html)

framework for expressive music generation in the computer animation production pipeline.

II. MUSIC PERFORMANCE RULES

Regarding previous studies concerning music performance and communication between performer and listener [2]-[3] it is possible to enounce that the performer is an important variable in the final audible result of a given composition. He has the power to provide expressive intention to the score depending on his own interpretation, enhancing the musical structure and introducing gestural qualities. Without it, music performed exactly as written would sound life-less and very mechanical. That is one of the issues when a computer performs a score, a dull and sometimes flawed result happens. Considering this fact the research at KTH Speech, Music and Hearing department analysed the characteristics that where able to change in a performance of a given score and related the quality and quantity of this changes with the expression of given emotions. The results of this investigation produced a system of several rules (about 30) that performers use to add certain expressive emotions to a score, covering different aspects of music performance. Basically the application of the rules produces a rearrange in time and velocity parameters such as "microtiming", articulation, tempo and sound level[2].

III. DIRECTOR MUSICES: APPLYING THE RULES

Director Musices (DM) is the main application that implements the rule system [4]-[5]. DM opens score files with the native .mus extension and MIDI files (type 1 or 0). After opening one of these files one should apply the rules. Those can be selected from a list of preset files containing groups of rules (Fig. 2.), or can be added individually. After selecting the rules, it is important to choose the quantity of each rule present in the performance. This step should be carefully executed. For non-expert users the result of each rule should be tested using maximum and minimum values in order to understand its real action on the performance. If different rules have an effect on the same parameter of the performance (ex: amplification or duration of notes) the result is a sum of both quantities. When all rules are applied the result can be listened immediately by clicking the "play performed" button. The results can also be saved in different formats like .mus – saves score and tracks variables; .per – like .mus but also saves performance variables; midi 1 and 0 – like .per but in MIDI format; and .pdm – a text file containing the application of most rules, along with the score and the deviation of those rules. This last format will allow applying the rules in real-time using another software to change their quantities.

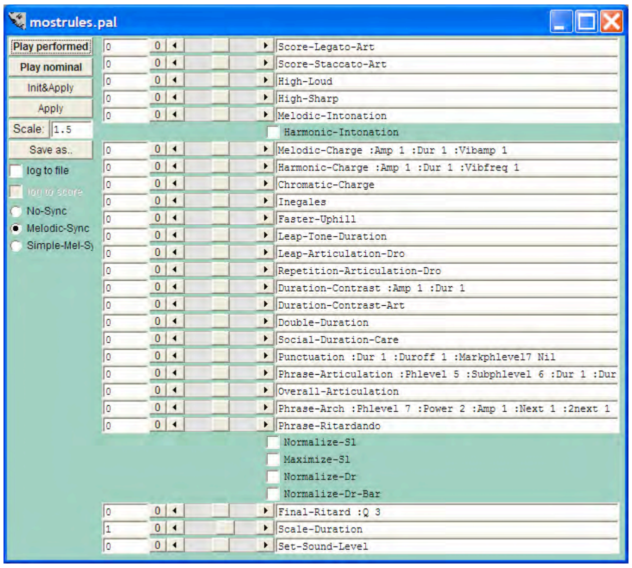

Fig. 2. DM window corresponding to the pallet preset called "mostrules". With this pallet configuration it is possible to apply and change the individual quantity for the majority of KTH performance rules.

IV. pDM: REAL TIME SUITE

pDM is a cross-platform Pure Data[3] (PD) patch used to expand the possibilities of the KTH rule system to real-time purpose. Its application is only possible in files that have been properly prepared in DM.

Realtime control in pDM can be done in two different ways: by manually changing the k factor (quantity) of each rule; or by using a more user-friendly interface based on a 2-dimensional space where the user can change the position of a single point with the use of a mouse. This second choice is rather imprecise, but definitely useful and necessary in real-time applications. Associated with the position of the point are different quantities of each rule reflecting different emotional expressions. The square (Fig. 3) is a 2 axis graphic used to map emotional expressions according to the activity-valence pair. The x axis stands for activity - balance between high and low energy; and the y axis for valence – positive or negative

[2] More information about the KTH performance rules: www.speech.kth.se/music/performance (01/07/08)

[3] http://puredata.info/

emotions. By the conjugation of these two axes the quadrants of the square can be characterized as happy, sad, tender and angry. Beside this mapping scheme, there are others present in the program like the Kinematics-Energy and the Musical Mapping. These are nothing but different combinations between the rule parameters and the 2D graphic. Other relations can be suggested regarding the target application, the interface and other factors related with its use.

Fig. 3. pDM Activity-Valence graphic interface used to change the emotional expression in real-time performance.

Sound design, like 3D animation, is today very dependent from digital technology. For those who deal with these issues the processes are now available in a very flexible and affordable way. The democratization of technology allowed people and groups to accomplish their objectives with minor investments and in a reduced amount of time. Regarding this matter there is the interactivity issue where automations can create even more “ready to use” tools.

V. pDaniM: The Soundtrack Automation

Producing an animation film is always an exhausting process that does not end when the characters are animated. After finishing the animation process there is, usually, the sound design work that can include: music, speech, sound effects, *foleys* and others sound groups. Ideally the link between image and sound should start right from the beginning of the production process. This is especially true when the animation involves speech, because the lip synchrony always start with sound recording prior to animation and not the opposite. Beside the lip sync, other good results can be achieved combining sound an image from the beginning. The mood of the characters can be positively manipulated by the music intension, so as rhythm in the edition process. Once the result is always a combination between sound and image it is relevant to think about new ways to improve both medias connection [7].

pDaniM (Performance Director Animation) it is a PD patch that attempts to answer this concern. The essence of the patch is to modulate the emotional expression of the MIDI soundtrack according to the narrative and characters emotions, concerning an interactive procedure between the movie script and the soundtrack. The core of pDaniM patch consists of an alteration to pDM, adding it a new feature: to read a text document with information regarding the emotional expression of a character (or global mood of the narrative) along a given animation movie.

In fact, pDaniM is only one part of the process for the interaction used in this work. First it is necessary to fill an Excel table with data extracted from the original movie script (Tab. 1). This data is divided into three different parameters: emotional condition of the scene/characters (happy, sad, angry, and tender), the exact time the emotion reaches its “peak” (hour: minute: second) and the amount of time the emotion takes to change form the previous one until reaching the peak (called ‘ramp’).

For example, assuming that the character is feeling happy the moment film starts, one should fill the first column of the table with the “happy” parameter, and the next three columns (corresponding to the moment when the character assumes that emotion) with 0, 0 and 0 values (hour: minute: second). The “ramp” parameter should be 0, 0 because the character starts with the defined emotion, not changing from a previous one. Finally, the coordinates (X,Y) for the desired emotion represented on the graphic interface (Fig. 3) are self-filled according to the emotion of the character, pré-selected on first column (Tab. 1). After completing this step it is necessary to export the table in the text file format (.txt) in order to be opened by the pDaniM. Graphic 1 is a bi dimensional graphic extracted from Table 1. used to recognize more easily the overall mood of the test movie.

Once this new emotional map (according with the movie script) is completed, it is necessary to load it into the application along with the score file (.pdm), previously updated in DM, containing the main theme. This is done similarly to the pDM application, by clicking in the open

	peak			ramp		coordinates	
Emotion	hour	minute	second	minute	second	x	y
happy	0	0	0	0	0	0	0
sad	0	1	0	0	15	0	400
tender	0	1	25	0	8	400	0
angry	0	2	50	0	3	400	400

Table. 1. Example of an Excel table used to map characters and narrative emotions during timeline. This table is the first step of the process designed to achieve interaction between sound and narrative/characters emotions. It should be filled according to the movie script.
The first column on the right determines the emotion of the character/narrative, the next three columns (under the "peak" topic) should be filled with the exact time (hour:minute:second) the emotion reaches its peak, the next two columns under the "ramp" topic, correspond to the amount of time the emotion takes to reach its higher level (peak) and the last two columns are self-filled.
In this example, the tabled is filled with only four emotions extracted from a test movie, specially created for the occasion by the responsible animator of "O Trovão", Dr. João Rema (Escola das Artes – Universidade Católica Portuguesa).

button in the main window. In pDaniM, two windows asking for a file (one after another) show up (and not only one like in pDM). The first is concerned with the score file and the second with the text file. The procedure to hear the main theme with emotional expression modulation is now completed. The user only needs to push the play button and tick the repeat one (in case of the theme being shorter then the film's total duration). The score will sound with the application of the rules, and those will change according with the timecode defined previously in the table (Fig. 5). The sounds used to interpret the score are the ones selected in DM before export to pDM format. Those sounds can be selected form the available MIDI devices connected or installed (if virtual instruments) on the computer. The DM has already some virtual instruments that can be used as a rough example for the animator to perceive the global sound mood. The sound designer or the musician/composer responsible for the movie should use more specific music production software (like Cubase, Pro-Tools, Logic or other multitrack applications) because those allow dealing with MIDI virtual instruments in a more efficient way. To accomplish that, a simple but effective tool has been used in order to record several midi channels at once directly to the production software. This tool is a free open source software called Midi OX[4] that works as a MIDI manager for monitoring, routing and other sort of operations related with MIDI. Using a patch bay feature makes possible to route the different channels of the score to different tracks in the production software. This way, several voices can be recorded separately in order to apply different sounds to them and do other kind of editing like muting, applying effects, transpositions, add voices, etc.

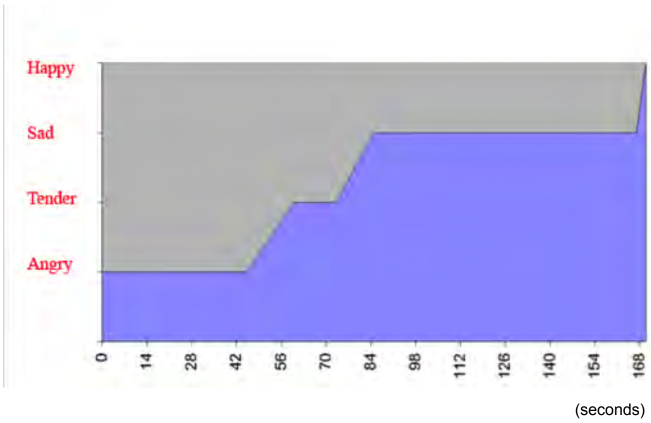

Fig. 4. A bi dimensional graphic extracted from of the table above. On X axis is represented the overall time of the movie (in seconds) and on Y axis the four different emotions. With this view it is possible to recognize more easily the overall mood of the movie, and how it changes along time.

VI. Possible Applications For pDaniM

pDaniM was created with interactive proposes in order to free the creative work from time and technique constrains that can sometimes overcome the first ones.

With its features pDaniM constitute an important tool for Directors, Producers, Sound Designers, Musicians and Animators. It particular suits the needs of the production process, providing time base for everyone involved. In a first moment can be seen as a sonic parallel of the animatic, a technique used in animation to give the notion of time and rhythm in the movie. It's basically a succession of non-animated key drawings of the animation, edited in a way that somehow respects the overall timing.

For Sound Designers it could be a good help to perceive how the soundtrack will sound or, at least, to have an idea about it. The musician can use it the same way; to get a first audition of his theme performed according the different emotional expression of the narrative. For animators this should be important since they can start to adapt their drawings and animation style to the movements of music. For Directors, Producers and in general all those involved in the process could constitute

[4] www.midiox.com - official web page of the application.

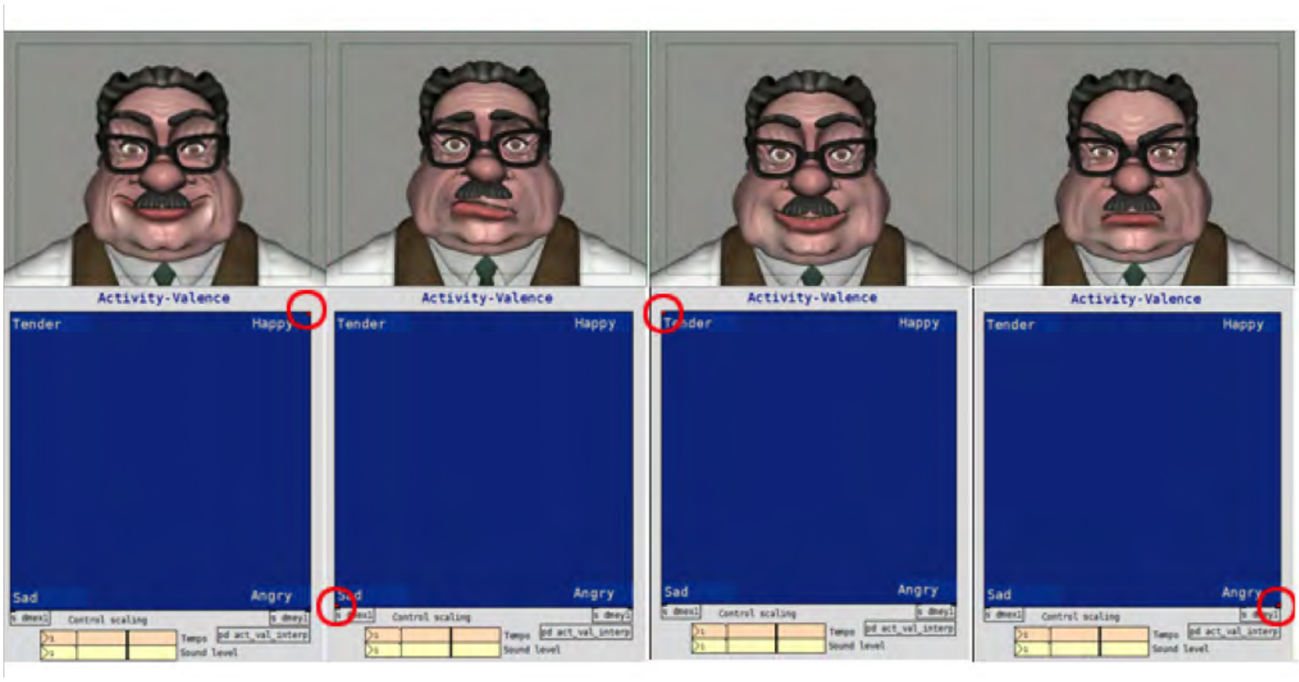

Fig. 5. Relation between character expressions and the point position in the pDM (or pDaniM) interface.

one more way to dialogue about the movie, to preview the final result.

VII. Future Developments

The development of this tool for sound design applications is far from being finished. Much more features could be added in order to improve the production pipeline, certainly regarding more interactive proposes.

One of those features could be a total automatic interaction between the character emotions and the music expression. Attending to the fact that it is possible to extract a logic description of the characters body and facial expressions (produced with Blend Shapes Techniques) from the animation timeline, it is also possible to use this data to transform algorithmically a basic music score that matches these expressions, resulting in an interactive paradigm for assisted music scoring.

This step would replace the task of input by hand the emotions expressed in the movie script, increasing interactivity between sound and image. Some directors choose not to use the traditional process of film production (like writing a script, storyboards, light plan, etc.), and just animate and make decisions along the process. This is more obvious in artistic contexts where the commercial intentions are not, usually, the most significant ones. In one case or another it seems important to make the process more interactive and dependent from the animation itself.

Another essential feature is the possibility to extract a MIDI and an audio file like wav or aiff (with the resultant emotional shape), directly from the pDaniM, instead of recording it in real-time. This option would make it easier for non-professional sound designers to access a file on a friendlier format in order to listen it on regular audio player, without the use of complex DAW (digital audio workstations).

VIII. Summary And Conclusions

A prototype of a software application for soundtrack creation was described. The application should take as input a basic table describing the dramatic charge of the narrative along time and a music score prepared in DM. The output should result in the performance of the same score with a new expressiveness and emotion in sync with the expressiveness cues from the table. The performance expressiveness is based on the performance rule system developed at KTH and uses the Director Musics (DM) and pDM applications to apply them.

A test movie with basic facial animations of a character extracted from the “O Trovão” cast, and a MIDI file with the main theme of the movie were used to prove the efficiency of the application. Preliminary listening evaluations demonstrate the relevancy of the investigation.

A correct correspondence between the character emotional expressions and the music performance emotions has been achieved. Although the subjectivity of aesthetic evaluation, it clear that the music performance emotions follows, coherently and harmoniously, the character facial expressions.

ACKNOWLEDGMENTS

This research was developed by the author during his stay at the Speech, Music and Hearing department at the Royal Institute of Technology, Stockholm – Sweden, under a Short Time Scientific Mission supported by the COST Action IC0601 on Sonic Interaction Design (SID).. Part of the work was also developed at Escola das Artes da Universidade Católica Portuguesa, Porto – Portugal and was tested and validated using an on going production of an animation movie, directed by Álvaro Barbosa at the same institution.

REFERENCES

[1] Bresin, R., A. Friberg, and J. Sundberg. 2002. "Director Musices: The KTH Performance Rules System." Proceedings of Sigmus-46, Kyoto. Japan.

[2] Palmer, C. 1997. "Music Performance." Annual Review of Psychology 48:115-138.

[3] Gabrielsson, A. 1999. "Music Performance". In D. Deutsch, ed. The Psychology of Music, 2nd ed. New York.

[4] Friberg, A., et al. 2000a. "Generating Musical Performance with Director Musices." Computer Musical Journal 24(3): 23-29.

[5] Friberg, A. 2005. Director Musices 2.6 Manual. (www.speech.kth.se/music/performance/download)

[6] Friberg, "pDM: An Expressive Sequencer with Real-Time Control of the KTH Music-Performance Rules". Computer Music Journal, 30:1, pp. 37–48,

[7] Mancini, M., Bresin, R., & Pelachaud, C. (2007). A virtual head driven by music expressivity. IEEE Transactions on Audio, Speech and Language Processing, 15(6), 1833-1841.