A Proposal for using SMS for Expression Modeling

Juan Reyes (1)
Cynthia Lawson (2)
(1) Fine Arts Department - Universidad de los Andes
jreyes@uniandes.edu.co
(2) Young Researcher Colciencias - Universidad de los Andes
lawson@colomsat.net.co

Abstract

Spectral models have proven to be a successful method for analyzing and re-synthesizing natural and acoustic sounds, to the extent that a musical phrase can also be modeled in order to extract performance parameters. This paper suggests various parameters which are embedded in the spectral analysis or .sms file and proposes some methods which can be used for expression modeling based upon compositional, perceptive, and cognition approaches. The focus is on achieving the means to construct a translation algorithm which the composer can use as a tool to analyze expressive material such as a musical phrase or a spoken sentence of text, and apply detected parameters to the vectors of a physical model or a synthesis algorithm..

1 Introduction

Our research has steered many of our points of view to the fact that there is a high degree of expression found in most performances played by Colombian performers to the point that significant research is meaningful. We believe that drum patterns from the Colombian Caribbean coast deserve attention as well as harp and accordion played also in the Caribbean, and the Eastern parts of the country. *(Figure 1)* The reason being: they portray to our knowledge very high levels of expression, and a significant degree of technical dexterity. This kind of expression can be recorded and analyzed by digital means in order to obtain time domain and spectral envelopes. Various expressive parameters can be obtained with a combination of both amplitude and harmonic envelopes depending on the approach and the parameters needed. Once some numerical values are established as correspondences to compositional hierarchies and categories, they can further serve as gesture or control means for real-time performances, for synthesis or to render a new musical phrase. We have focused on triggering signals for physical and spectral models. With this, we wish to achieve expressiveness combined or convoluted with the sound of the traditional physical model.

2 Related work in Los Andes

The Computer Music group at Los Andes has approached the tasks of sound synthesis and signal processing from the standpoints of speech and voice synthesis leading to spectral and physical models. In respect to physical modeling, a model of the flute has been proposed using waveguides [6]. Our goal at this time is to adapt this model to various circumstances depending on the context. We wish to model traditional Indian flutes and pan flutes as well as different lengths and materials on organ pipes. This implies performance parameters which we have been extracting by spectral analysis. With this procedure we have been able to feed the model by means of a parameter file. Currently, processing does not occur in real time, but this gives us the advantage of experimenting with more dynamic expression parameters such as dynamics, modulations, and time delaying activities instead of the MIDI interface approach. Furthermore, this heuristic can be benefited by the use of algorithms which enhance or transform original performance parameters. Spectral modeling has enabled us to research various text and spoken signals, as well as instrumental phrases. With a graphical interface displaying harmonic and amplitude components we have obtained research data on drumming patterns and cello performances. This has been useful for some experimenting from the Von-Helmholtz point of view and has provided the differences in gesture and haptic approaches among different traditional instruments.

3 Expression and Perception

The principles governing our scope on perception encompass those variables that deal with visual perception and image creation and some counterparts and additions in the auditive domain. From the sensorial point of view we have acknowledged a relationship between auditory and visual perception in particular to spatial and room size perception. We have experimented with stereoscopic and stereophonic imaging, resulting in the production of dynamic visual and sonic objects which resemble the traditional role of picture animation. In the case of imaging from sound sources we have been able to model virtual sound objects that, according to their position in a three-dimensional model, can portray their size, shape, and distance from the listener. This in turn has provided us some means for dealing with perception variables such as sound intensity and reverberation as expression variables. If we add subtle pitch transformation, we have relativistic effects such as Doppler. In the case of rhythm perception, most of our work has been done in the relationship among music and language phonetics. We have studied accents and phrasings from different regions in Colombia as well as differences in pronunciation and in the speed of spoken words. These changes play a crucial role in image creation in particular when we have tried to synchronize sound with video or animation. Our findings permit us to state that in most of the samples taken for this research, sound was the conducting factor providing the pace and rhythm of the video or audiovisual work. In this way sound variables as tempo and rhythm play a crucial role in the way a musical piece is perceived and hence become important expressive variables.

In relation to sound shaping or in more general terms the Critical Band approach has proven to be worthy in the case of sound imagery. In phonetic sound, by applying the technique of bank filtering with a bandwidth of a third of an octave and twenty four filters perception of spoken text can reveal gender of the sound source as well as the place and physical qualities of the soundfield. With this we can recognize if a person is talking live or over the phone. This process is very useful for spectral analysis because of the masking or tuning of adjacent frequencies to the stronger partial. In speech recognition and synthesis, the Critical Band approach becomes a sensitive issue because it gives the ear patterns for determining overall loudness and speech intelligibility [4]. Therefore substantial attention has to be paid to filtering of a sound by carefully selecting the frequency and bandwidth of the filter, if not its transfer function. Filtering has been widely used in traditional instrument performance and becomes a more relevant issue in the case of physical and spectral modeling because of their intimate association with signal processing. We take filtering as a substantial variable for expression modeling.

In more musical terms we follow the ideas of Lerdahl and Jackendoff in the Generative Theory of Tonal Music (GTTM). This theory postulates that if a listener is familiar with a particular musical style, he or she is able to hear a piece of music in that idiom as more than just a sequence of notes [5]. GTTM suggests that there is an acoustical musical surface, that there are musical semantics that pertain to the style and therefore grouping and metrical structures. Moreover the musical piece is perceived as time-span and prolongation reductions which are result of a combination of the stated rules and structures. In this sense we acknowledge style and its substructures of melody, rhythm and harmony as basic expressive quantities [3].

Figure 1.Expressive drummer from the Colombian Caribbean

4 Expression Parameters

Expression Modeling can be achieved by isolating those parameters which the composer considers relevant to the composition. In general we can isolate dynamics, rhythm, timbre and nuance. Rubato and vibrato are considered very effective means of conveying expression from a musical performance because they belong to a specific moment and to the particular performance of the musical piece. In our expression model, for a piece to convey expression we need at least a change in one of the above parameters. This can be achieved by using envelopes or by extracting parameters out of a spectral analysis file. Expression often changes in a time-domain basis and therefore makes it possible to model these values by means of deterministic or stochastic values for predicting the necessary values. Furthermore, if we apply case-based reasoning [1], or constrained-based algorithms, options might multiply and therefore be able to model a gesture performance in a more natural way. In our experiments we have extended expression parameters to rhythmic and sound articulations; time compression and expansion applied to phrase duration. Also, spatial components, room and context manipulations as well as the influence of text in expression modeling and composition. Most of them have rendered musically-meaningful results.

5 Extracting Parameters

Sound analysis and synthesis based on spectral models are useful for the extraction of high-level parameters from real sounds, their transformation and the synthesis of a modified version of the original. We have used spectral models by means of Spectral Modeling System (SMS) [8] to extract embedded information which relates to the expressive parameters described previously. In most cases SMS provides a meaningful signal representation for a sound to be transformed expressively. We obtain attributes such as attack and release times, formant structure, vibrato or average pitch and amplitude [1]. Transformation can be applied to these parameters directly in order to obtain a vector to resynthesize the original signal or to input a physical model.

6 Rules and objectives for a parameter translation algorithm

In the design of the SMS we have parameters as attack duration, decay duration and spectral envelopes as basic parameters for an analyzed note. Vibrato frequency and amplitude, amplitude envelopes, frequency envelopes can also be extracted from an analyzed musical note [8]. As a simple layout of suggestions for a translating algorithm for an .sms-like analysis file we suggest the following set of rules as directives for the design of the pseudocode: amplitude in dB should translate into dynamics; spectral components of the sound can be analyzed as timbre and its related derivatives such as nuance and spectral change; partial trajectories reveal timbre change and become very expressive because of its dynamic constrains; time envelopes in seconds and milliseconds are time-domain envelopes which change rhythm and note duration and if used proportionally can affect articulation and overall tempo. We also consider rhythm as the relationship between sound and silence. In addition to sound shaping, nuance can also be regarded as the relationship between one timbre and another. As for the graphical interface we would like the flexibility of tcl/tk to be port on different platforms but we will be focusing mainly on UNIX operating systems. Our design of this sort of GUI should suggest composers to explore expression modeling from the standpoints of spectral analysis to modeling of sounds by means of SMS, physical models and traditional Music V synthesis.

7 Expressive Parameters for a Physical Model

We have focused a significant part of our research to the synthesis of electronic sounds and in particular to Physical Modeling. We have seen the advantage of working with controllers, time-domain envelopes and control files as the more effective ways of controlling the sound. For a model, in order to take advantage of expressive parameters it must be designed in such a way. In this proposal we want to avoid the use of physical or haptic interfaces but nevertheless we want to obtain similar results. We believe the more flexible the model, the more expressive. However, as more parameters are introduced more complexity in design and execution is added. Therefore with flexibility as an issue, the composer or the user should be able to control the degree of flexibility for the expressive parameters and the designer should bound manageable degrees of complexity. For instance, in our flute model we would like to change the size of the bore of the flute in a time-domain basis. This single parameter adds more complexity to the model and perhaps new design and new transfer functions, but bore changing becomes part of an expressive performance which is mandated by the connection between the .sms file and our proposed expression algorithm. These sort of expressive controls can also be achieved by the use of modal synthesis parameters as proposed by Iovino et al in Modalys [2]. In general we have successfully experimented expression parameters such attack variation, breath pressure manipulation, and bore manipulation on existing brass and woodwind models as suggested by Scavone in his saxophone model [7].

8 Conclusions

The analysis and extraction of expressive parameters by means of SMS is a very important step for obtaining meaningful musical information from a digital soundfile. In many cases this analysis does not convey useful results and thus we believe that substantial attention, knowledge and understanding has to be applied at this crucial step. It is imperative for the success of a translation algorithm that there should be pertaining information from the analysis of the soundfile to be able to detect the necessary values. We have obtained results in which deterministic values are not enough parameters to model expression, therefore non-linearities in a musical phrase should not be taken slightly. In the case of concrete music composition the combination of stochastic and deterministic signals gives good results in unpitched sounds but this sort of response is barely useful for musical phrase modeling. Hence, a waveform visualization of both time and spectral envelopes is almost mandatory before any analysis procedures can take place. This provides the relevant information for the analysis parameters in SMS. Once the analysis stage has been successfully overcome most of our procedures described above can be performed in a reliable way. Our only concern at this stage is managing the amount of sampling rate resolution information in order to obtain better bench-marking in the synthesis algorithms of our models. We are in the process of analyzing redundancy and data compression techniques for improving this situation, but also constrained-based algorithms for functions in expression modeling has been suggested to us. Finally we acknowledge the feasibility of expression modeling by means of SMS and imply the need for more research in this direction and a useful musical implementation for testing with music users with little computer knowledge.

References

[1] Arcos J.L., de Mantaras R.L., and X. Serra, 1997. “SaxEx: a case-based reasoning system for generating expressive musical performances”, in *Proceedings of 1997 International Computer Music Conference*, pp 329 - 336, Thessaloniki, Greece.

[2] Iovino, F., Causse R., and Dudas, R., 1997. “Recent work around Modalys and Modal Synthesis”, in *Proceedings of 1997 International Computer Music Conference*, pp 356-359.

[3] Jackendoff, R., 1993. *Languages of the Mind*, Cambridge Mass, USA, MIT Press.

[4] Julesz, B. ed. 1995, “Auditory and Visual Perception Compared”, *in Dialogues on Perception*, Cambridge Mass, USA, MIT Press.

[5] Lerdahl, F. and R. Jackendoff, 1985. "A generative Theory of Tonal Music", Cambridge Mass, USA, MIT Press.

[6] Reyes, J. and Lawson C., 1997. "Another Approach to Expression With Algorithms", Submission to ICMC98, Santafe de Bogota, Universidad de Los Andes.

[7] Scavove G.P., 1996. "Modeling and Control of Performance Expression in Digital Waveguide Models of Woodwind Instruments", in Proceedings of 1996 International Computer Music Conference, pp 224 - 227, Hong Kong.

[8] Serra, X., 1997. “Musical Sound Modeling with Sinusoids plus Noise”. G. D. Poli and others (eds.), *Musical Signal Processing*, Swets & Zeitlinger Publishers.