A compositional approach to analysis re-synthesis

Report on composer in research project

Marco Antonio Suárez Cifuentes
Composer
marcosuarezcb@hotmail.com

Norbert Schnell
Researcher/Developer
Norbert.Schnell@ircam.fr

ABSTRACT

This report summarises the work achieved in the framework of a *composer in research* project in the IMTR team at IRCAM.
In this project we have experimented with different real-time processing techniques for the analysis, transformation, and resynthesis of particular morphological aspects of live performed voice and instrumental sounds.
Based on segment-wise sound descriptions, the developed tools allow for re-orchestrating and re-organising particular morphological aspects of the live performer's sound using concatenative synthesis of pre-analysed sound materials driven by the real-time analysis.
The relationship between the morphology of live performance and the synthesised sound materials is controlled by a set of compositional tools manipulating the extracted sound descriptions in real-time.

INTRODUCTION

Although, first collaborations and reflections that have considerably influenced this project have been started in 2009, the project has been essentially conducted between January and December 2010. Over the year we have alternated more intensive periods of collaboration on the project with periods of occasional interaction.
While the essential of the project was finalised in Autumn 2010, the last period was dedicated to the consolidation and documentation of the developed tools as well as the experimentation with further prospective approaches.

We have had the possibility to experiment with some of the developed ideas and tools in three musical projects:

- *Caméléon Kaléidoscope,* March 2010
- *Installation sonore sur des fragments de Caméléon Kaléidoscope*, May 2010
- *Plis,* September 2010

The following calender gives an overview of the work and its milestones:

During 2009

- First reflections and collaboration with an ATIAM internship on "*Morphological Segmentation*".

End of 2009, beginning of 2010

- Development of a first segment-wise morphological sound description based on sound materials of an early version of the piece *Cameleon Kaleidoscope*.
- Development of a first real-time implementation of the extraction, concatenate synthesis, and compositional tools using FTM & Co.

March 2010

- Premiere of the piece *Cameleon Kaleidoscope* at the Biennale Musiques en Scène Lyon using the first set of tools.

Spring 2010

- Reimplementation of the analysis and synthesis around the MuBu container.
- Revision of the description (loudness and pitch distribution).
- Experiments with hand drum performances using materials from the pieces *Cameleon Kaleidoscope* and *Plis*.
- Experimental integration of ambisonics spatialisation into the system (*Installation sonore sur des fragments de Caméléon Kaléidoscope).*

May 2010

- Public performance of an experimental study for hand drum at the Ambisonics 2010 conference at IRCAM.
- Exploration of fuzzy-logic (with Alain Bonardi) as a means of creating relationships between the morphology of analysed and synthesised sounds.
-

Summer/Autumn 2010

- Extension of the segment-wise description by timbral coefficients (MFCC).

- Development of an alternative segmentation and description adapted to continuous sound textures ("tiling").

September 2010

- Premiere of the piece *Plis* at the *Abbaye de Royaumont*, using a part of the developed tools.

End of 2010

- Consolidation of the developed tools and their documentation.
- Experimentation with classification tools based on segment descriptions.
- Elaboration and experimentation of further applications transforming realtime analysis descriptors.

BACKGROUND

The reflections on composition and audio processing techniques that led to this research project had their origin in a former collaboration on performance *Poetry for //dark-/ dolls* elaborated in the framework of the the Cursus II project at IRCAM in 2009. This musical work is conceived to be performed in two synchronous separated spaces. It was premiered by Valérie Phillipin and the *Ensemble Intercontemporain* at the *Espace de Projection* at IRCAM. Motivated by the research of extended vocal techniques, the piece for voice, five instruments and live electronics creates inside one concert hall two acoustically isolated and interdependent performance spaces also representing two interdependent musical faces of one unique piece.

Part of the live electronics developed for this musical project involved a particular analyse/resynthesis technique founded in the idea of using granular synthesis and real time transformations of sound for acoustical masking and perceptual illusions concerning the localisation of sound sources. The implementation, essentially based on frame-wise audio mosaicking using the extraction of Mel bands, was realised with FTM & Co (Gabor).

The experimentation with the audio processing developed for *Poetry for //dark-/ dolls* and the resulting sounds was partly inspiring the general approach of instrumental scoring in the piece as well as new ideas conducting re-orchestration of vocal gestures by instrumental sound segments. The real-time environment allowed for the composition and high-level control of the rhythmical structures generated by the audio mosaicking using granular synthesis. The precise rhythmical control over the synthesised sounds was as important as the timbral aspects for establishing a coherent musical relationship between the live electronics and the vocal and ensemble performance.

While the timbre of the synthesised sounds in this system was derived (by spectral matching) from the performers sounds, the temporal aspects were generated by pre-composed rhythmical structures. The idea for a next generation of tools was to derive also the temporal morphology and rhythmic articulation by analysis from the live performance. This motivated the research for a adequate morphological descriptions as well as compositional tools to apply transformations and permutations to the represented musical gestures.

At the beginning of the research project we started a collaboration with an ATIAM student in the framework of his masters project on "Morphological Segmentation". In this project, Pierre Machart developed a non real-time (MatLab) application engaging semi-automatic segmentation and classification based on *Segmental Hidden Markov Models* (SHMM) that implied the generalisation of a partial segmentation and classification made by a user. In this collaboration, the system was partially evaluated using musical materials (i.e. recordings of instrument vocal performances) from *Poetry for //dark-/ dolls* as well as further pieces.

Although the real-time techniques described in the following are not technically following up on this work, this rich and fruitful experience has inspired new compositional ideas as well as other aspects of the techniques for the morphological segmentation and description developed in this project.

RESEARCH AND DEVELOPMENT

Within this project, for each development we have engaged previous exchanges concerning musical concepts and realisation criteria. We have noticed from the beginning that one main problem with high-level descriptors usage was their excess of information and their non-easy application to interpret a musical or instrumental gesture with its temporal development.

Each composer has it's own conception of what “gesture” means, how it contributes to the elaboration of his own music, therefore what a gesture needs to be recognised by a technological environment. It is not possible to establish a one rule making automated morphological segmentation adapted for every musical context neither to the aesthetics of each composer.

One statement during our research project, was to implement tools and methods for specific musical applications and materials and not generic descriptors which often force the user to adapt his artistic needs and language of a particular technology.

This principle also defines the development approach for the live electronics of *Caméléon Kaléidoscope.* Our expectance for this ensemble piece was to use real-time analysis and morphological segmentation to control concatenative synthesis in the context of a dynamic interaction between live performance, and a pre-recorded and analysed instrumental parts.

The idea was to realise a spatialised virtual ensemble controlled by an environment reacting to individual or collective live musical gestures. Thus, allowing a strong formalised relation between the written score, the live performance and live electronics. Other main motivation was to conduct this interaction with compositional strategies engaging the analysis of logical sequential patterns and the generation of similar semiautomatic musical constructions driving gestural types or rhythmical developed sequences.

This two approaches allow for playing with perceptual limits between acoustical sounds and the spatialised resynthesis of the pre-recorded sound materials.

During our research period we have advanced in the knowledge and realisation of *Max/MSP* patches involving real-time analysis and morphological segmentation based on selected descriptors. In addition, we have revisited instrumental timbre analysis on a *MuBu* version of audio the pre-existing developing and temporally integrated descriptors MFCC and Mel based spectral analysis.

This work covers essentially two aspects:

- Development of a segment-based morphological description adapted to specific sound materials and related real-time and offline analysis tools
- Development of compositional tools in Max/MSP and experimentation in different musical contexts

In addition, we have conducted further experimental developments such as an experimental integration of ambisonics spatialisation into the *mubu.concat~* concatenative synthesis module and an experimental environment for the classification of sound segments.

Great parts of these developments, although driven by the implementation of specific functionalities in a specific musical context, are completely generic. The project has considerably contributed to the design and validation of numerous features of the *MuBu* container as well as the related access and synthesis modules (i.e. *mubu*, *imubu*, *mubu.knn*, *mubu.concat~*, *mubu.granular*) and integrated since into the distribution and documentation of these tools.

Morphological Segment-wise Sound Descriptions

The details of the first version of the segment description have been published at the 2010 *Sound and Music Computing* conference [1]. For this description a set of coefficients is calculated on each segment. The segmentation is based on an algorithm initially developed in the *VoxStruments* project. Even if the description has been specifically developed for sound materials from the piece *Caméléon Kaléidoscope*, it also can be applied to similar strongly articulated materials and speech.

For a second version of the description, the set of coefficients has been revised. The resulting description calculated for each segment can be summarised as follows:

- Description of the loudness envelope (duration, maximum loudness, slope)
- A pitch distribution over the segment (mode, centroid, standard deviation, energy weighted pitchness)
- Eight loudness weighted MFCC coefficients (2^{nd} to 9^{th} out of 24 calculated coefficients)

Other than in the first version or the description published in the *SMC* paper, the slope of the loudness envelope in the second version is calculated by a linear regression over the previously calculated envelope. This description has been developed for strongly articulated sound materials of voice and instrumental performance and also gives interesting results when applied to speech.

Starting to work with very different sound materials performed with hand drums, we had to develop a new description adapted to slowly evolving unvoiced sound textures. This description is calculated in regular periods on temporal windows between 500 and 1500 milliseconds. The coefficients calculated for each window are the maximum loudness, the standard deviation of the loudness, the maximum and standard deviation of an onset detection function, and eight loudness weighted MFCC coefficients. The standard deviation of the loudness as well as the coefficients derived from the onset detection function represent different temporal aspects of the texture (i.e. its roughness). The re-synthesis of sound textures corresponding to this analysis uses granular synthesis with grains of the size of the analysis windows. We have called this technique “tiling”.

The *segment-wise* description, the *tiling* and previously developed *frame-wise* audio mosaicking three complementary temporal representations adapted to sound materials of different characteristics. For performances using inhomogeneous sound materials it makes sense to use different techniques in parallel.

A set of Max/MSP patches based on MuBu and FTM & Co implements the extraction of the developed descriptions in real-time as well as asynchronously on pre-recorded audio files loaded into a buffer.

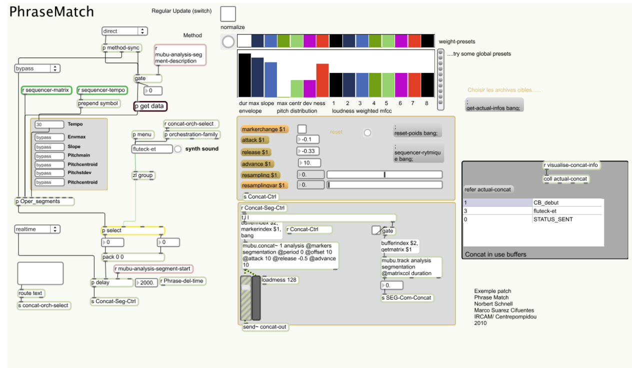
Segment-wise Matching

A musical phrase can be defined as a coherent sequence of musical gestures, sound qualities and durations. In this research a representation of gesture is realised by the description of selected acoustical parameters and their evolution in time.
In the context of *Caméléon Kaléidoscope and Plis* we were able to use a set of nine descriptors providing informations of duration, energy, envelope and pitch for each morphological segment. As described above, later, on our research project we have implemented a new description, integrating Mel's or MFCC's as morphological segment analysis parameters.

The interactive control of analysis and synthesis parameters in *Caméléon Kaléidoscope* and *Plis*, helped to understand the constraints and advantages of segment-wise matching, and to observe how different parameters engage musical consequences. Which is the better architecture on the integration of this tools on a concert patch? How a composer should approach the formalisation of the electroacoustic score for an optimal application of this environment?

One relevant possibility on the actual version of k-NN segment-wise matching is the real-time control of analysis parameters weights with normalised descriptors. As it can be verified on the related patch *_concat-research.pat* different combinations of descriptors weights conduct to significant variations of resulting resynthesis. Some parameters such as duration, maximal loudness and slope have a mayor influence on our perception of source and synthesis correspondences on a first level. Nevertheless using strongly weighted MFCC descriptors on segmental matching results on further convincing results, in particular when associating instruments to human voice.

We have also observed in the practice of segment-wise normalised matching that extremely differentiated kind of sound sources between live performance and synthesis enabled richest musical interaction and further interesting results.
The use of weighted descriptors and k-NN allows a flexible musical or gestural interaction between a source and it's synthesis instead of a mechanical representation of a sound.



Articulation of sound and rhythmical matching, have demonstrated to be fundamental parameters in the perception of the relationship between the live performance and the electronics.
Segment-wise matching reveals a constraint, it is limited to specific prerecorded materials related to a piece, therefore in difference to other similar approaches such as *Orchidée* (IRCAM) this technique doesn't pretends to generate a best absolute response , but the best possibility within a relative controllable context. We have confirmed that important collections in a container including a large palette of sounds, instrumental techniques, durations, dynamics and gestural types provide interesting and convincing results.

Relaxed Real-Time into Musical Time

The use of morphological segment-wise matching has one consequence in real-time usage, it is not possible to get the temporal description of a segment before its end has been recognised by the environment, therefore, synthesis sound output will always be delayed. To preserve the source musical phrase structure on segment-wise matching the retard should be controlled and composed precisely, this technique has been experimented in *Caméléon Kaléidoscope* and *Plis,* allowing to synchronise live instruments with environment concatenative synthesis.

The first segment-wise synthesis experimentations were focussed on a kind of “imitation”, one interesting but limited compositional approach. In a later implementation we have developed and experimented a live sequencer *_Sequencer-Research.pat*, permitting to store reference markers and access in a flexible way the segment analysis recorded in real-time. This approach allows to construct further elaborated compositional relationships between live instrument and chosen moments of the elapsed time of live performance.
The sequencing model can be described as following:

Segment analysis description:
$(Seg_1, Seg_2, Seg_3, ...Seg_n) = input$

Segment sequences select and tag (RT):

$$N_1=(Seg_a\, Seg_c)$$
$$N_2=(Seg_f\, Seg_l)\ldots N_N=(Seg_p\, Seg_x)$$

N represents a label for each sequence, determined by a start and end markers references. The list of sequences is stored on a MuBu container named *Sequencer*. Using labels the composer can play selected sequences of segments in their original order and tempo or in a relative speed.

The experimentation with sequenced segment-wise matching opens a larger panorama of possibilities concerning composition with descriptors on musical time. This approach should be still tested on a real time context, evaluating how it affects the live performance and in what degree a musician can control consciously his own influence on synthesis result. We can imagine that for some cases a prerecorded segment-wise description of all the piece, communicating by gesture follower or score follower can lead to more efficient use of this technology.

Rhythmical Structures

One idea of this work was to use the rhythmical structures of the pre-analysed sound materials as a reference for realtime algorithms developing and generating rhythmical sequences. This motivated the realisation of generators transforming the analysed rhythmical sequences to elaborate new composed structures involving some probabilistic control. In *Caméléon Kaléidoscope*, both, the written score and the control of the spatialised contatenative synthesis engaged a similar procedure for rhythmical composition.

The used algorithm has been inspired by the chameleon's skin cells mechanism and implemented on Open Music and Max/MSP. The function can be described as follows:

x_i = initial value, y = random val < y_{max}
s = output sequence, z = maximal length

$S_i=(N_1 N_2 N_3 N_4 N_5 \ldots \ldots N_n)$
S_i = sequence of durations or of classes
If $x\in S_i$, *then set* $P_{n1}=x_i \wedge Pn_2=x_i, \wedge s_1 \wedge x_1$

$s_1=N_{(P_{n1}+1)}, N_{(P_{n1}+2)}\cdots N_{(P_{n1}+y_1)}$

and $x_1=N(P_{n1}+y_1) \wedge substitute\ P_{n2}\ by\ s_1$

$S_1=(N_1 N_2 N_3 (s_1=N_{(P_{n1})}, N_{(P_{n1}+1)}, N_{(P_{N1}+2)}\ldots N_{(P_{N1}+y_1)}) N_5 \ldots N_n)$

create $(S_1, S_2, S_3, S_4 \ldots \ldots S_n)$

for each $(x_1, x_2, x_3, x_4 \ldots \ldots x_n)$

with $A=(s_1, s_2, s_3 \ldots\ldots s_n)$ *and* $N_x=z\ set\ P_{s1}=S_i$.

In order to ease the analysis and recomposition of musical phrases or individual gestures using segment-wise matching, the score proposed an radical differentiated gestural writing for each instrument associated to a formalised rhythmical writing.

Varying Real-Time Analysis Parameters

Even though, the work with 16 different instruments involved a first generation of descriptors, this experience has contributed to observe the possibilities of morphological segmentation in association with concatenative synthesis.

The analysis environment developed during this research allows to control realtime parameters defining segmentation, we have observed through their practice that is essential to adapt main parameters in accordance to acoustical context, the analysed instrument and the variability of its musical gestures. A lot of experimentation with a soundfile or a performing instrument should be engaged to accomplish a convincing segmentation, in *Caméléon Kaléidoscope* ten instruments of the ensemble were involved in realtime interaction: the *Flute, the Doublebass, 2 Violins, Viola, Cello, Harp, Bass Clarinet, Piano and Chinese Drums* , this experience has allowed to understand the needs for analysis on each instrument.

The score of *Caméléon Kaléidoscope* demands to select and change in precise moments the analysed instrumental sources, conducting to prepare individual analysis parameters and to adapt them in realtime context by an external device. The quality of the matching output is directly influenced by the fine performance of this controls.

In a concert context, the sensibility of the environment causes a conditional dependence between the musician and the matching response. Each performance is unique and the musical quality is influenced by the gestural engagement and performance precision of the instrument player.

Therefore this environment could be so understood as a composed sound-installation, conducted by the possibility to use rhythmical structures generators combined to composed instrumentations (segment-wise matching) involving in specific cases also a stochastic distributions. The sound result for each performance is unique but almost coherent and formalised.

Accessing Multiple Soundfiles

Integrating *mubu.concat* with *polybuffer~* allows for using composed combinations of instruments on synthesis in order to produce a monophonic multi-timbral response based on segment-wise matching, or a polyphonic synchronised matching with a defined number or instruments.

These two approaches were present in the realisation of the orchestration models used in *Caméléon Kaléidoscope*. They were applied either to the instrumental score or to the real-time synthesis of pre-analysed sound materials. In both cases the recomposition and acoustical augmentation of musical gestures is realised by timbres within a chosen instrumental group that responds to a soloist or another group of instruments. The structure of this process follows the metaphor of a kaleidoscope.

Matching Multiple Soundfiles Analysis with k-NN

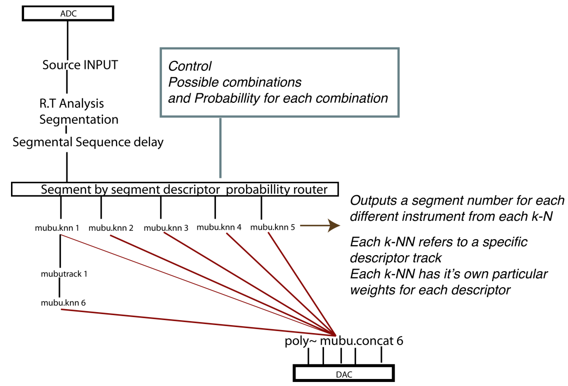
The module *mubu.knn* allows to drive dynamically the group of target descriptors within a MuBu container, a choice of heterogenous sounds produce hybrid instrumental musical phrases. This procedure is advantaged by normalisation and the *advance, attack and release* parameters of *mubu.concat* object.

Synchronised k-NN for Instrumental Combinations

We have had implemented to a patch enabling to route real-time analysis description to chosen k-NN objects pointing to different pre-analysed instruments. This multiple segment-wise matching allows to set individually the weights for combined descriptors, further more, to work and transform descriptor values in real-time by means of high-level controls.

For each segment a probabilistic choice of instrumental combinations is engaged before synthesis, this method permits to get rich results concerning timbre mixtures and instrumental density. A further experimentation on this path is to create segmental matching chains complementing direct matching, conducting a more consistent texture for spatialised concatenative synthesis. We have confirmed on this practice the importance of articulation as a definitive factor for segment-wise matching efficiency.

The described process can be represented by this schema:



Varying Descriptor Weights

During our research a lot of experimentation has been done with different combinations of descriptors weights applied to a k-NN.
It has been noticed in the actual version of morphological segment-wise description of the tested musical materials that the segment duration, the segment slope and the segment maximal energy are the most influential descriptors on perception.
Other high-level descriptors such as mode, centroid, standard deviation, energy weighted pitchness have an important influence when associated to other parameters, nevertheless, their response is less stable, thus, depending on the complexity of the analysed source and its variance on time.
MFCC coefficients, are to relevant for perception and output recognisable results on matching techniques.

Using MFCC for Mosaicking and Tilling

As defined above, segment-wise matching illustrates efficient results when both the analysed sound and its synthesis sound are clearly articulated. Continuous, statical or slow varying non pitched sounds seem inadequate for morphological segmentation techniques because they are not defined by a recognisable variation of energy, timbre or pitch. This observation motivated the experimentation and realisation of patches allowing to work with less articulated sounds for sound material of the project *Plis*.

In an earlier work *Poetry for // dark- / dolls*, MFCC frame-wise matching has demonstrated an efficient description of spectra, nevertheless, musically limited: basic MFCC does not take in account a temporal evolution of sound. When a frame is matched, the granular synthesis responds with no consideration of how the description changes on the next frame.
This means, that frame-wise matching imposes a extreme short period for analysis and re-synthesis correspondences coherence resulting on an homogenous synthesis result and

the diminution of gestural quality in synthesis. One implementation on *Poetry for // dark - / dolls* and re-experimented during this research is to impose a high-level rhythmical control to granular synthesis segments affecting duration and period.

This technique helps to reduce the frame-wise synthesis "sound-effect", and contributes in the rhythmical coherence between live performance and the granular synthesis. Meanwhile, as frame-wise k-NN takes in account only one frame for matching there is a strong probability of loosing timbre correspondences so normalisation effectiveness because of grains duration.

Periodic Segment-wise Analysis And Re-synthesis

MFCC and Mel band real-time analysis provides detailed information about spectral energy distribution for each frame, this description is temporally restricted and provides limited information about timbre.

Timbre needs a minimum time to be perceived, on slow evolving sounds a better description of timbre is achieved when MFCC or Mel coefficients variance and evolution is represented in time. For linear, continuous or not strongly articulated sounds the duration for each description should be periodically imposed and synthesis coherence is achieved by means of overlapping resulting on a continuous sound texture. When a "singularity" or a strong variance is detected, analysed sound can be articulated, the resulting segment synthetised using segment-wise matching.

Accessing Descriptors

The real-time and the off-line patches developed during this research allow to access the descriptors from three temporal perspectives: the frame, the quantised segment (tilling), and the morphological segment-wise description.

As described above, all classified descriptors values are contained in a MuBu object, and can be accessed at any time using a *mubu.track* object. Real-time description is also recorded in real-time, thus, allowing to work directly with any described parameter or group of parameters from any temporal perspective in a quite flexible way.

One interest during this project was to experiment with possible uses of information provided by individual or selected descriptors for real-time applications and transformations concerning complementary real-time treatments, in the same frame to explore ways to control descriptor values transformation in order to construct a richest palette of relationships between source and synthesis sound.

Spatialisation

During the experimental integration with ambisonics system "*Installation sonore sur des fragments de Caméléon Kaléidoscope*", we have worked on strategies to use MuBu container data in order to control spatialisation of sound.

The environment used segment-wise matching with multiple-sound MuBu containers using some methods described above. The environment used fifteen individually spatialised *mubu.concat* objects controlled by the same number of k-NN objects. When a segment is been selected, the following informations are requested from descriptors: *Soundfile name, segment duration, segment slope, segment maximal loudness.*

For each soundfile or track a particular set of rules concerning spatialisation has been constructed, the duration, the slope, and the maximal energy of each segment cause associated behaviours on the trajectory, the depart, arrival position and the speed of a determined segment to be spatialised.
These rules were arranged in collections in order to generate probabilistic spatial developments, for each step the actual position of source on the concerned spat was taken in account. Relating descriptors for each segment allowed a coherent relationship between musical gesture and spatialisation.

Transforming Individual Descriptor Values

By accessing segmental descriptors list each coefficient can be modified by determined functions and reintegrated in the list with no consequences in the temporal flux of segment-wise matching. This possibility motivated the experimentation with individual descriptors in a compositional context during our research. The following lines describe the operations explored for each analysis parameter:

Duration

We have mainly worked with proportional duration factors which can be defined as Tempo relations using *mubu.play*, or accessing directly the input analysis container.
For MuBu based sequencer modifying duration for each segment affects the musical phrase speed allowing interesting relationships for rhythm and time. When a transformation is applied to the real-time flux, the rhythmical structure of the source is preserved affecting exclusively articulation of sound.

Slope

This parameters describes in a general way the loudness evolution of a segment, and can give an idea of its envelope.

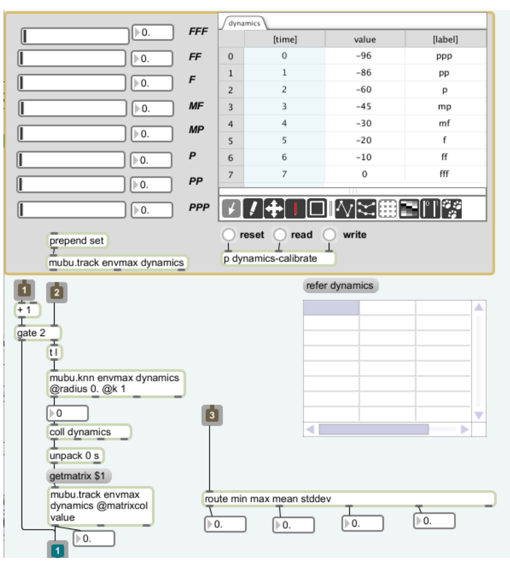
By means of mubu.track in order to get the minimum and maximal slope values in a container we have implemented efficiently some composed operations permitting to inverse slope or to construct other source – synthesis envelope transformations, for example:

$$\textit{If } slope > 0.0001 \text{ then}$$
$$scale(Slope_{max}, 0.0001)\, into(0, Slope_{min}) \text{ else}$$
$$scale(Slope_{min}, -0.0001)\, into(Slope_{max}, 0)$$

Using this simple mathematical principle we have been able to observe that even if slope descriptor simplifies the idea of envelope to a schematic expression, the controlled transformation of this descriptor favors the acoustical differentiation, see complementarity between a sound and its synthesis when they happen to be synchronised.

Maximal Loudness

As morphological segmentation criteria has to be calibrated in each performance according to the instrument and the acoustical context, we have noticed that loudness needed to be classified in musical terms and calibrated to construct a more elaborated control between dynamics.



For this sake it has been implemented and tested one patch object assigning variable values to musical indications ***(fff, ff, f, mf, mp, p, pp, ppp...)***

By simple mapping of calibrated dynamics we have been able to create flexible musical presets, one example:

fff** stays **fff**, **ff** becomes **ppp**, **f** becomes **p**, **mf** stays **mf**, **mp** becomes **f**, **p** becomes **ff**, **pp** stay **pp**, **ppp** becomes **mp

To construct an appropriate mapping for this individual descriptor classification other solution could be the utilisation of fuzzy-logic patches implemented by A. Bonardi. Further on to extend its use with a similar calibration and classification for other described parameters.

Pitch main and Pitch centroid

The pitch description as a value is not reliable for all the cases in morphological segment-wise analysis. The precision of the information is strongly perturbed for some kind of instruments, or in particular musical writing contexts. This means, that Pitch based descriptors are less easy to transform in a precise or objective manner, nevertheless, during this research we have experimented with some techniques that may be useful in particular cases.

1- Generating probabilistic chord classes for specific pitches when a soft pitch deviation and a high pitch quality is detected.
2- Using sequence chains of transposition intervals between the live instrument and it's segment-wise matching based synthesis.
3)- Learning in real-time the interval sequential logic for recognisable pitches and using this information in future appearances of recognisable referenced pitches as a sequenced substitute for pitch descriptor.

Pitch deviation and pitch quality

It has been imagined in future researches to work precisely with this descriptors applying for pitch deviation a similar implementation as the one described for slope. Pitch quality parameter by this moment has been used mainly as a reference parameter, to choice whether pitchmain or MFCC is adequate for matching.

MFCC

As MFCC coefficients are conceived as a non-linear representation of spectra, it is less easy to create a function which transforms its values in a conscious and meaningful manner. We have observed in segment-wise matching that MFCC descriptors usage has a strong relation with pitch based descriptors, this condition needs to be taken in account when operating values.

Ambisonics Spatialisation in Concatenative Synthesis

The integration of spatialisation coefficients into the *ZsaZsa* engine allows for the weighted distribution of single channel audio segments to an arbitrary number of audio channels. This way, spatialisation is available as a straightforward extension of concatenative and overlap-add synthesis.

Using ambisonics coefficients, the synthesis engine can output a multi-channel sound stream (one output signal per channel in Max/MSP) in ambisonics format. The ambisonics output steam is independent of the loudspeaker constellation and can be decoded by the *Spat* ambisonics decoder just as other ambisonics encoded signals. The number of output channels depends on ambisonics configuration (2D or 3D) and desired maximum ambisonics order. A set of parameters can be associated independently to each synthesised segment (azimuth, elevation, distance, and ambisonics order). The overlap-add functioning of the engine allows for realising an arbitrarily number of overlapping segments with an individual spatialisation of each segment.

The implementation has been integrated as an experimental extension of the *mubu.concat~* module (based on the *ZsaZsa* engine) among other possibilities to distribute audio segments to the output channels of the module.

FUTURE PERSPECTIVES

After experimenting with different types of data mapping between a real-time source analysis and a selected database, we have been able to identify some problems concerning the interpretation of parameter's values when they're used in a different approach from normalised matching.
The description we have stablished reveals a specific relationship with a particular musical context, this means that some analysis parameters are reduced in order to describe in time the development of one or combined descriptors.
Some defined descriptors, such as *slope*, give a general idea of energy tendency for a particular segment. When a morphological segment has a clear and unique behaviour the relation between described value and the envelope evolution is coherent, meanwhile when a segment has an irregular energy evolution the "Slope" descriptor reveals to be less efficient, therefore, not easy to transform in a precise musical way.
This example confirm the limited efficiency of individual descriptors to conduct signifying information about segments and the difficulty for the composer to give resulting "abstract" values a musical meaning. Other problem revealed by individual descriptors manipulation is the fact that analyse streams out "relative" values and not "perceptual values", the use of normalisation allows to construct an adapted relation between the analysed source and the synthesis, this means, that contextual matching can be some times be inconsistent in musical terms.
Segment-wise matching based upon our descriptors describe a coherent gestural evolution, but not a precise representation between a given source and a synthesis sound. In a musical context this condition has some advantages, we are not expecting to imitate but to give a new interpretation of analysis values by respecting each soundfile or instrument its own contextual characteristics.
Therefore, the implementation of methods allowing to compose and transform individual descriptors seems rather ambiguous about the musical purpose. In this context, to enlarge compositional possibilities engaging a meaningful approach to musical material we have being experimenting with a classification based on multidimensional *mahalanobis* and *k-NN* distances.
We expect, with formalised classifications to allow the composer to control musically the relations within real-time analysis synthesis contexts, and two conduct segment-wise matching in a further precise way, for example, referring k-NN object to a selected number of classes within a totally classified sound.
Other perspective is to combine classification with automated generative algorithms as a tool to analyse in real-time and to learn the evolution of musical gestures type for a specific instrument. During our research we have experimented this application with an off-line classified soundfile using labels.
The developed patch allows to create a new musical flux using the class-sequences from the original model on a similar technique as *Caméléon Kaélidoscope* rhythmical generator function described above. Material is recomposed in a probabilistic way preserving the gestural logic provided by the original soundfile.

ACKNOWLEDGEMENTS

We would like to thank for their fruitful inspiration and contributions to this project Alain Bonardi, Riccardo Borghesi, Thibaut Carpentier, Helene Colombotti, Jean-Phillipe Lambert, and Markus Noisternig, as well as Pascale Bondu, Arshia Cont, and Anne Guyonnet for their friendly accompaniment.

REFERENCES

[1]N. Schnell, M. A. Suárez Cifuentes, J.-P. Lambert, "First Steps in Relaxed Real-Time Typo-Morphological Audio Analysis/Synthesis", Sound and Music Computing (SMC), Barcelone, 2010.
Available Online:
http://articles.ircam.fr/textes/Schnell10a/index.pdf