Surfing the Waves: Live Audio Mosaicing of an Electric Bass Performance as a Corpus Browsing Interface

Pierre Alexandre Tremblay
CeReNeM
University of Huddersfield
Huddersfield, HD1 3DH, U.K.
+44 (0)1-484-473608
p.a.tremblay@hud.ac.uk

Diemo Schwarz
Ircam - Centre Pompidou CNRS-STMS
1 Place Igor Stravinsky
75004 Paris France
+33 (0)1-4478-4879
diemo.schwarz@ircam.fr

ABSTRACT
In this paper, the authors describe how they use an electric bass as a subtle, expressive and intuitive interface to browse the rich sample bank available to most laptop owners. This is achieved by audio mosaicing of the live bass performance audio, through corpus-based concatenative synthesis (CBCS) techniques, allowing a mapping of the multi-dimensional expressivity of the performance onto foreign audio material, thus recycling the virtuosity acquired on the electric instrument with a trivial learning curve. This design hypothesis is contextualised and assessed within the *Sandbox#n* series of bass+laptop meta-instruments, and the authors describe technical means of the implementation through the use of the open-source *CataRT* CBCS system adapted for live mosaicing. They also discuss their encouraging early results and provide a list of further explorations to be made with that rich new interface.

Keywords
laptop improvisation, corpus-based concatenative synthesis, haptic interface, multi-dimensional mapping, audio mosaic

1. CONTEXT - THE SANDBOXES

1.1 The previous instances

1.1.1 Overall Sandbox#n Ethos
The current project set out to design a third instance of custom-built DSP instruments[6] in the context of the bass+laptop solo performance series entitled *Sandbox#n* [21]. This series is an ongoing practice-based research in performance where the meta-instrument is mastered to be able to do free improvisation in solo as well as within ensembles. The authors refer to free improvisation practice where there is a sublimation of the means of producing the inner-heard idea, through virtuosic knowledge of the instrument[2]. As the instrument is also used in ensemble settings, mostly with post-free-jazz acoustic improvisers, it must be able to be responsive and multi-faced. This research is therefore performer driven, and assessed solely on aesthetical and usability grounds through its use in real-life performances as an expressive extension of the performer[7].

It is interesting to note that the hardware means of the *Sandbox#n* instruments are consistent throughout the series, and consist of very simple replaceable elements: an electric bass, a laptop with a professional audio interface, four control pedals and a fader box. The reasons for this are threefold: firstly, the performer is used to them, like many other techno-friendly guitarists[15]. It is hoped that through consistency, neuro-motor reflexes and subtlety of gestures will port. Secondly, laziness plays a part in not wanting to learn a new interface to develop the same level of expressivity in performance; moreover, the bassist does not want to give up his movement freedom by adding gesture sensors. Finally, for very practical reasons, custom sensors and interfaces are not used because from our experience, material needs to be easily replaceable on tour[22].

NIME2010, 15-18th June 2010, Sydney, Australia

1.1.2 Two Previous Instances
A short description of the first two instances follows, to put in context their different strengths and weaknesses that informed the design of the new instrument. *Sandbox#1* is a DSP-instrument that is mainly loop-based, with spectral-domain, time-domain, grain-domain loopers, and a versatile feedback matrix. Most of the time, it will yield slow results in solo settings, mainly because of the need to fill the loopers and route them while both hands are being busy with the bass playing.

Sandbox#2 was an answer to this feeling of latency, without discarding the first instrument. Its limits, highlighted by practicing with it, informed the design of a new instrument. The solution to slow individual control of parameters and loops was to implement dynamically-created presets interpolated by a single pedal. To make it more dynamic, a crude mapping of input performance parameters (i.e. pitch, noisiness, etc.) to synthesis parameter (i.e. granulator parameters, tail holding algorithm, etc) was done. This approach was powerful, but the analysis was too basic, and the mapping was too rough and too abstract to yield a subtle control over the synthesis. It is fun to play with, but limited in expressivity if we agree with transparency as a sine-qua-none condition as presented in most literature[7][9][17]. Another limit of *Sandbox#2* is that it is still a hyper-instrument as defined by the French composer Michel Pascal (quoted in [15]): an extension of the bass sound, linked to its source with a limited access to other sound banks.

1.1.3 A New Set of Tools
With all these concerns influencing the desired features of the next instance, the new technology of corpus-based concatenative synthesis (CBCS) became publically available, yielding uneven musical results regarding its application to audio mosaicing. CBCS [18] makes it possible to create music by selecting snippets of a large database of pre-recorded sound (the *corpus*) by navigating through a space where each snippet is placed according to its sonic character in terms of sound *descriptors*, which are characteristics extracted from the source sounds such as pitch, loudness, and brilliance, or higher level meta-data attributed to them. This allows one to explore a corpus of sounds interactively or by composing paths in the

space, and to create novel harmonic, melodic, and timbral structures while keeping the nuances of the original sound.

When the target is set from the analysis of live input, we talk of live audio mosaicing, which seemed to be a way of addressing the issues of mapping and responsivity into the next instance of the *Sandbox#n* series.

1.2 A New Instrument

The main goal of *Sandbox#3* was to have an intuitive, expressive access to the richness of the electroacoustic sample banks of the performer's collection.

This is not a new quest: it is anchored in the current trend of controlling playback and/or granulation parameters of sound banks in more interactive ways. The *Clatter (a.k.a.Tiles)* sampler[3] with its manual classification and tiling approach, or *CataRT*[19] with its mapping of descriptor-per-axis, are going in the right direction but still rely on the mouse to navigate a two dimensional space on the computer screen. As most literature on the subject tells us, a traditional instrumentalist is able of much more expressivity on his instrument than what is allowed by simple mouse movements, expressivity usually gained through years of practice[7][17]; *Sandbox#3* should therefore make use of this multidimensionality of the expressive control, transposing the multi-dimensional nuances of the performance of the mastered interface, the electric bass.

This approach of controlling granular synthesis by the means of the guitar could be related to others (as documented by Tuncer[23]), but is going away from the event/onset (note) paradigm most systems are based on, and proposes a one-to-one mapping of many descriptors of the audio stream. As no new interface has to be learned, the instrumentalist should be able to recognise the influence of his performative actions on the audio output—despite sounding radically different—that should allow feedback on gesture as well.

This feedback is an essential part of the instrumentalist practice, as the performance is adapted to what is heard in context. For instance, Mari Kimura has documented how she changes her vibrato according to the acoustics of the hall[12]: hearing the sound generated by her action, she adapts her performance instinctively. This intimacy is multi-dimensional, usually subtle, but of the utmost importance in the intimate relation a performer has with his instrument. Interestingly, vibrato is one of the mostly used examples of multi-dimensional parameters given in papers on expressivity[7] or on mapping[11] because it falls out of the MIDI event-based paradigm.

After earlier work on live control of CBCS ([19] between sect. 5.1 and 5.2, [8] sect. 3.1), this is the first attempt of an implementation within a performative new interface approach, with the concerns of expressivity at its heart. It also incorporates elements of the chaotic poetic navigation described by Stoll[20], but with an instrumental input which provides more intimate control on the browsing.

Another important goal of *Sandbox#3*, more sonic this time, is that it should not rely on the bass as its sound source. In other words, no electric bass is to be heard; this is simply achieved, as it is an electric instrument: its audio is solely used for analysis and will not be amplified. *Sandbox#3* is therefore not a hyper-instrument in the new accepted definition cited above, and as showcased by some daring string[13] and wind players[5][14]. It is more an electric interface, not unlike MIDI wind-controllers or MIDI-guitars, but not limited to note-on/offs events, using instead a flux of descriptors to translate the performance on a much greater level of dimensions[16]. Within this *Sandbox*, the standard attack, pitch and amplitude are given, but also their variation between onsets, which opens the doors to audio translations of all alternate techniques on the bass, as well as infinitesimal pitch and timbral variations. In other words, *Sandbox#3* is a subtle instrument in phase with contemporary practice of the electric bass.

Note that the authors are aware that this proposed method is not a panacea, or not even an extensive research on new interfaces and gesture translation (as the kind reader could find in [11] or [16] for instance). It is a report on an empirical research on new ways of surfing growing collections of audio waves, with sensual, musical, intuitive and expressive interface, from a performer/composer perspective.

2. PROPOSED TECHNICAL APPROACH

We use interactive corpus-based concatenative sound synthesis based on the modular real-time implementation *CataRT* [19] for *Max/MSP* with the extension libraries *FTM&Co.*[1], making it possible to navigate through a two- or more-dimensional projection of the descriptor space of a sound corpus in real-time, effectively extending granular synthesis by content-based direct access to specific sound characteristics.

The segmentation of the source sound files into *units* can be imported from external files or calculated internally, either by arbitrary grain segmentation or by splitting according to silence or pitch change. The descriptors that are calculated are the fundamental frequency, periodicity, loudness, and a number of spectral descriptors: spectral centroid, sharpness, flatness, high- and mid-frequency energy, high-frequency content, first-order autocorrelation coefficient (expressing spectral tilt), and energy. For each segment, the mean value of each time-varying descriptor is stored in the corpus.

The selection of the unit that best matches a given *target* is performed by evaluating a weighted Euclidean distance on the normalized per-descriptor distances. Either the unit with the minimal distance is selected, or one is randomly chosen from the units within a radius, or from the *k* closest units to the target.

2.1 Real-Time Control

In our application, instead of navigating in 2D with a pointing device, we control *CataRT's* selection with the input of the real-time audio signal from the electric bass. The target for synthesis is thus given by a descriptor analysis of the live audio signal, with subsequent mapping, as explained in the following.

2.1.1 Signal Analysis and Segmentation

To pilot *CataRT* synthesis with a live instrument, the audio signal is analyzed in real time according to the same descriptors and parameters used by *CataRT* for its batch analysis of pre-recorded corpora. We segment the input target sound into short fixed windows. Every trial of deriving triggering events by onset detection in the input signal would mean that the resulting output would be one event late, since the input segment would have to be analysed completely before it can serve as a target.

The list of calculated target audio descriptor values for every segment is sent to the selection module to output the unit closest to the target. The relative weights of each of the descriptors used in the selection can be adjusted graphically. The final rate at which units are synthesized is affected by the analysis window size, the triggering method, and the segmentation method of the corpus. Units need not be output in a regular rhythm: for example, in *CataRT*'s *fence* mode, triggering happens only when a different corpus unit becomes closest to

[1] http://imtr.ircam.fr/ and http://ftm.ircam.fr/

the target. The analysis units may be filtered by rate or descriptor value before being sent to *CataRT's* selection. In particular, loudness and periodicity descriptors may be used to gate signal frames with values below desired thresholds.

2.1.2 Descriptor Mapping

The main way to influence the response of the sound output characteristics to the input control signal is by the mapping of input descriptor values to target descriptors, and by their weights. The weights control the relative importance with which a certain descriptor target must be reached, by influencing its contribution to the total distance. If the weight is large, even a small difference in descriptor value will augment the distance between the target unit and the candidate database unit. This descriptor's target value must therefore be met precisely by the selection algorithm, which is at times desired.

Regarding the range mapping, the possibly non-overlapping descriptor spaces between the bass control signal and the various corpora was addressed by a simple linear mapping that maps the complete range of each input descriptor (as determined in a calibration phase) to the full range for a given corpus. Thanks to *FTM&Co*.'s powerful matrix operators, this mapping can be performed by a simple vector expression.

2.2 The Performance System

The performance instrument has a number of simple features common to the *Sandbox#n* series, added to enrich its sonic possibilities further more. It has three channel strips, each with an independent synthesis engine linkable to one of the three interchangeable corpora. Each channel strip has simple mixing options (pan, distortion, gate, filter, gain) and a looper. This latter option is not audio-sample based, but keeps in its buffer the sequence of target units, which allows later alteration of the weighting strategy of descriptors, or even of which corpus they are played from.

3. EARLY RESULTS

3.1 First Impressions

The first impression of playing with this instrument was an exciting-yet-natural one: the instrumentalist could be playing metal cans and bells, dirty analogue synthesisers, and other various sound banks on his hard-disk, with a very responsive musical feedback. There was an immediate sense of tangible interaction with the chosen corpus, and the instrument quickly allowed to get subtle level of sonic exploration.

Feedback from the first live performances contained comments that were raving on the transparency and the instrumentality of the electroacoustic gestures, despite no bass sounds at all being heard whilst seeing a bass being performed on stage. These results tend to confirm the double axis theory of expressivity/transparency[9] where the linear mapping of the interface helps to yield an expressive performance for both the performer and the audience.

3.2 First Improvements

Here is a list of early concerns we had to reflect and act upon in the early stages of the design.

3.2.1 Latency

There is a built-in latency to this system due to the frame-based analysis method. Reducing the analysis window size was helpful, but tends to exaggerate the granularity of the sound. It was also noted that when latency is constant, short enough, and not confronted with direct sound from the bass, the performer gets used to it quite easily. It is not unlike the latency found on most MIDI guitar systems, but since this is a little annoyance, reducing it would be something to explore further.

3.2.2 Descriptor Scaling and Weight

Using a linear scaling between the smallest and largest values of both target and corpus descriptor's vector solved the problem of non-overlapping descriptor spaces described in 2.1.2. But this simple solution could be refined; for instance, offsetting could be more intuitive on certain parameter like pitch, which would then respect pitch classes by a simple octave offsetting.

We found that the choice of how many descriptors to use to specify the target, but also their relative weight in the proximity assessment, is of utmost importance, even more than previously apprehended. Usually, pitch, amplitude and periodicity are enough to give good performance; also, the relative importance of the basic pitch and amplitude descriptors over the more timbral descriptors is quite overwhelming.

An interesting finding of deferring to the instrumentalist the choice of the descriptors used, and their relative weight, is that the performer is made aware of his hierarchy of listening skills, and of their relative importance in the feeling of reactivity and expressivity. It also helps to assess and improve the mapping coherence and intelligence. Comparative studies with different bass players could be interesting to pursue, to see if these performers listen to descriptors in a similar hierarchical way.

3.2.3 Granular Sound

The very nature of the replacement of a stream of grains gives the system a granular synthesis sound. The form of the granular synthesis envelope, and the careful and customized settings of the segmentation of given corpora in relation to their content improves significantly the sound quality and helps to reduce this artefact, as pointed out by previous users of the *CataRT* package[20], but there is room for further research.

Another element that could help in the short term is to implement concatenation cost in the *CataRT* audio engine. It should improve the smoothness of the result, according to Rasamimanana[16]. Another method to explore would be the implementation of some constraint satisfaction programming[1], and/or allowing certain parameters to be 'adjusted'. Pitch for instance could be tuned straight on the target value from the nearest element according to the weighted descriptor, and timbral targets could be interpolated, but this would be at the cost of sound quality according to Schwarz[18].

4. FURTHER DEVELOPMENTS

Here are listed potential developments to this very promising instrument, further from the obvious continuous refinement of the elements presented in section 3.

4.1.1 Attack Descriptor

It seems that segmentation in heterogeneous grain size could be helpful. For instance, a short analysis window of the bass sound is needed for low latency performance, but there could be a new binary descriptor that would flag up the presence of an attack during a given grain. Such a descriptor could rely on a sophisticated transient analysis *à la* Rasamimanana[16], or could even join up different sources of information: a faster dedicated multi-string MIDI guitar system as in [4] could flag quickly the onset, its fundamental and amplitude, while slower audio streams analyses could provide more subtle timbral information over time with a given latency. On the other side, the synthesis corpus could have much shorter grains for those with the attack flag detected true, and longer, smoother grains for the sustain parts of the notes where the flag is false.

4.1.2 Other Re-Synthesis

Other methods of re-synthesis could be explored. *CataRT* is based on granulation, but other CBCS rely on spectral re-synthesis[10][18]. Even better would be to integrate the two approaches, with a time-domain attacks synthesis engine and a spectral-domain sustain synthesis.

One further idea of using this control instrument would be to map its analysis to what Hunt and Wanderley refer as *Meaningful Sound Parameters[11]*. In the pre-performance stage, complex multi-parameter synthesisers (typical but not limited to physical modelling) could be analysed in relevant descriptor spaces. The re-synthesis would work as it does at the moment, but would send parameter values to the synthesis engine. This descriptor mapping of complex synthesizers has been suggested in [8] and preliminary tests have been carried out with graphical control in a 2D space, but to use a live instrument as its control is yet to be tested.

4.1.3 Other Mapping

Finally, explorations of other, non-linear mappings (i.e. pitch to noise) could be done, as proposed in [8]. The authors do have reserves on this, as most literature on non-transparent mapping seem to point towards a loss of control intimacy and expressivity potential (for instance, [9][23]).

5. CONCLUSION

The authors believe that real-time audio mosaicing of an electric instrument is rich in potential to allow skilled performers to explore their sample collection. It definitely allows translating the expressive potential of the instrumentalist in an intuitive musical exploration of corpus. Moreover it allows the public to relate to such a performance, despite the sound world being distant of the instrument source. Further experiments are of utmost importance to exploit this approach rich in possibilities, but the feedback from performers and the public is greatly encouraging.

6. ACKNOWLEDGMENTS

The authors would like to thank Norbert Schnell and collaborators for their support with the *FTM&Co*. package, and the Canada Council for the Arts and the University of Huddersfield for their overall support in the sabbatical leave.

7. REFERENCES

[1] Aucouturier, JJ and Pachet, *Jamming with Plunderphonics: Interactive concatenative synthesis of music.* Journal of New Music Research, 35:1,35 - 50, 2006

[2] Bailey, D. *Improvisation*. Da Capo Press. Philadelphia, 146 p. 1992

[3] Batchelor, P. *Clatter.* http://www.peterbatchelor.com/maxClatter.html accessed 3 Feb 2010

[4] Bates, E. Furlong, D. and Dennehy, D. *Adapting Polyphonic Pickup Technology For Spatial Music Performance.* in Proceeding of International Computer Music Conference (ICMC). Belfast : August 2008

[5] Burtner, M. *The Metasaxophone: concept, implementation, and mapping strategies for a new computer music instrument.* in Organised Sound 7(2): 201–213 Cambridge University Press 2002.

[6] Casserley, L. *Plus ça change: Journeys, Instruments and Networks, 1966-2000.* in Leonardo Music Journal, Vol. 11 p.43, 2001

[7] Dobrian, C. and Koppelman, D. *The E in NIME: Musical Expression with New Computer Interfaces*. In Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06), Paris, France

[8] Einbond, A. Schwarz, D. Bresson, J. *Corpus-based Transcription as an Approach to the Compositional Control of Timbre.* in Proceedings of the ICMC. Montréal : August 2009

[9] Fels, S., Gadd, A. and Mulder, A. *Mapping transparency through metaphor: towards more expressive musical instruments.* in Organised Sound 7(2): 97-108, Cambridge University Press, United Kingdom, 2002

[10] Hoffman, M. Cook, P. and Blei, D. *Bayesian Spectral Matching: Turning Young Mc Into Mc Hammer Via MCMC Sampling.* in Proceeding of International Computer Music Conference (ICMC). Montréal : August 2009

[11] Hunt, A. and Wanderley, M. *Mapping performer parameters to synthesis engines.* in Organised Sound 7(2): 97-108, Cambridge University Press, United Kingdom, 2002

[12] Kimura, M. *Creative Process and Performance Practice of Interactive Computer Music: A performer's tale.* In Organised Sound: 8-3. Cambridge: Cambridge University Press: 289-296. 2003

[13] Machover,T. *Hyperinstruments : A progress report 1987-1991.* MIT Media Laboratory, 1992

[14] Palacio-Quintin, C. *Eight years of practice on the hyper-flute : Technological and musical perspectives*. In Proceedings of the Conference on New Interfaces for Musical Expression, 2008.

[15] Pohu, S. *L'improvisation électroacoustique comme point de rencontre entre la composition électroacoustique, l'interprétation instrumentale et la lutherie numérique*. Master Thesis at Université de Montréal, unpublished, December 2008.

[16] Rasamimanana, N. and al. *Perspectives on Gesture-Sound Relationships Informed from Acoustic Instrument Studies*. in Organised Sound 14(2): 208-216, Cambridge University Press. United Kingdom, 2009

[17] Schloss, W. A., *Using Contemporary Technology in Live Performance: The Dilemma of the Performer*. in Journal of New Music Research, 32: 3, 239 – 242, 2003

[18] Schwarz, D. *Concatenative sound synthesis: The early years.* In Journal of New Music Research, 35:1,3-22 , 2006

[19] Schwarz D, Cahen R, and Britton S. *Principles and Applications of Interactive Corpus-Based Concatenative Synthesis*. in proceedings of Journées d'Informatique Musicale (JIM). Albi : March 2008

[20] Stoll, T. *Beyond Concatenation: Some Ideas for the Creative Use of Corpus-based Sonic Material*. in Proceeding of International Computer Music Conference (ICMC). Montréal : August 2009

[21] Tremblay, PA. *List of works.* on http://www.pierrealexandretremblay.com/worklist.html accessed 3 Feb 2010.

[22] Tremblay, PA, Boucher, N. Pohu, S. *Real-Time Processing on the Road: A Guided Tour of [iks]'s abstr/cncr Setup*. In Proceedings of the ICMC 2007, Copenhagen, pp. 527-530.

[23] Tuncer, J. *La synthèse granulaire et son contrôle musical en temps-re_el par le jeu de la guitare.* Master thesis at ENS Louis Lumière, unpublished, June 2006.