Designing for Cumulative Interactivity: The _derivations System

Benjamin Carey
Faculty of Arts and Sciences
University of Technology, Sydney
Benjamin.Carey@uts.edu.au

ABSTRACT

This paper presents the author's _derivations system, an interactive performance system for solo improvising instrumentalist. The system makes use of a combination of real-time audio analysis, live sampling and spectral re-synthesis to build a vocabulary of possible performative responses to live instrumental input throughout an improvisatory performance. A form of timbral matching is employed to form a link between the live performer and an expanding database of musical materials. In addition, the system takes into account the unique nature of the rehearsal/practice space in musical performance through the implementation of performer-configurable cumulative rehearsal databases into the final design. This paper discusses the system in detail with reference to related work in the field, making specific reference to the system's interactive potential both inside and outside of a real-time performance context.

Keywords

Interactivity, performance systems, improvisation

1. INTRODUCTION

Within the field of interactive computer music, a great diversity of approaches to the design, implementation and evaluation of systems for improvisatory performance can be found. The basis of such a diversity of approaches may be summed up succinctly in the words of George Lewis (a noted pioneer in the field), who has observed that: "Musical computer programs, like any texts, are not 'objective' or 'universal', but instead represent the particular ideas of their creators."[1] With specific reference to machine improvisers, Collins suggests that their creators often link these systems to particular musical styles or contexts [2]. As such it can be stated that the various innovations within the field represent the multiple practical, aesthetic and technological preoccupations of the system designers, all of whom bring to the field a diverse range of experience from a variety of musical and technological backgrounds. The author's musical background (therefore informing the design of the present system in question) is as classically trained contemporary saxophonist, with a strong interest in abstract and non-linear musical structures and an involvement in non-idiomatic or "free" improvisation in performance. Such an improvisational aesthetic is often characterised with respect to its subtle and detailed consideration of the timbral content within a particular musical context. This can be contrasted with improvisatory traditions that privilege musical innovation within a melodic/rhythmic/harmonic framework (such as jazz for instance).

1.1 Timbral Awareness in Interactive Performance Systems

In a recent survey of improvisatory systems by Hsu [3], the author makes particular note of design approaches dealing directly with the analysis of timbre as it relates to the design of improvisatory responses within such systems. The systems surveyed can be characterised as displaying characteristics of what Rowe has defined as "performance paradigm" systems [4], i.e. systems that are conceived as providing the human with a *virtual* or *artificial* counterpart in improvised musical performance. Hsu's own AHRS system is a performance system that uses timbral analysis from the live saxophone input to automate and control a series of virtual improvising agents. These improvising agents make use of gestural curves analysed and stored throughout a performance to automate their various synthesis and processing parameters [3]. Enabling system awareness of the current and past timbral context within an improvisational performance is often a theme with such systems. Of interest in this regard is Ciufo's system Beginner's Mind, using real-time descriptor analysis to dynamically build a measure of the "perceptual identity" of phrases analysed from the live input over the duration of a performance. The author then uses this to relate the current performance state of the human improviser with data analysed previously in the same performance [5].

1.2 Designing for Rehearsal

In discussing his model of conversational interaction, Paine [6] asserts that a true interactive relationship between interactor and system can only exist if a system is capable of changing and evolving over time. Systems of this type would be designed to reflect the "cumulative experience of interrelationship" exemplified in a human conversation model of interaction [6]. Dynamic change and evolution is inherent in a human conversational context not only through the interactions between people at a given time and place, but also over longer time scales through the navigation of shared and cumulative history of the interactors. Within Paine's conversational model of interactivity, two characteristics of human conversation are of particular note in this regard, namely that a conversation is "unique and personal to those individuals," and in addition that the conversation is "maintained within a common understood paradigm (both parties speak the same language, and address the same topic)." [6] It is suggested that these attributes are by nature cumulative throughout multiple encounters between participants, situating the interaction within a wider temporal frame beyond the present communicative context. This of

NIME'12, May 21-23, 2012, University of Michigan, Ann Arbor.

course finds direct parallels with musical rehearsal and performance, and in particular in improvisatory contexts.

The cumulative building of an interpersonally shared histories between musicians in performance and rehearsal has been discussed in the work of Bastien and Hostager. [7] In addition, the authors note that performance-time interactions between musicians also depend upon “suprapersonal” common histories of individual participants, based upon individual players’ prior interactive experiences with other musicians outside of the present interactive context. [7] Taking this cumulative view of interactivity into account, it is then suggested that the rehearsal/workshop space may be equally important to consider in the design of new interactive performance systems. This is especially true of systems designed to interact with a solo instrumentalist, as consideration for cumulative interactions between human and machine may enhance a system’s potential for enabling a “unique and personal” conversational form of interaction between the two participants. [6]

In the literature, the rehearsal/practice space is referred to in large part with respect to the initial design and the evaluative phases of interactive music system design. In the work of Johnston et al., collaborative rehearsal sessions with a professional performer/composer in the initial stages of a system design are shown to advance prototype versions of a software artefact [8]. At the other end of the design process, Hsu and Sosnick make reference to the rehearsal space as forming integral part of an evaluative framework for comparing system design in mature and instantiated system designs [9]. In addition to the design of novel methods to automate system responsiveness during performance time interactions, it is suggested that system designers may also consider how such systems may use data collected over multiple rehearsal sessions to enrich any subsequent performances with the software. Such an interactive paradigm can also be seen as privileging the role of the performer in shaping the overall shared history developed between both performer and system.

2. _DERIVATIONS

2.1 Overview and Motivations

_derivations is an interactive performance system developed in MaxMSP for use by a solo instrumentalist, and is designed to derive all of its sonic responses to improvisational input from the instrumentalist's live performance. A central aesthetic concern of the system is the coherent integration of the sound palette of both the acoustic instrumentalist and computer, achieved through a ‘hands-free’ and unmediated mode of instrumental interaction with technology (i.e. the human and machine communicate through sound only). The system is designed to display an awareness of the sonic context presented by the instrumentalist with respect to a broad definition of its current and past timbral content, achieved herein through the grouping and storage of real-time sound descriptor analyses of the incoming acoustic audio stream. A form of timbral matching is used to relate the most recent performance state of the instrumentalist to an expanding database of recorded and analysed performer phrases. The system makes direct use of these recorded phrases as sonic and gestural source material for transformation and recombination via the linked synthesis and processing modules.

A unique aspect of the _derivations system relates to its intended use both inside and outside of a performance time interaction, taking into account the conceptual significance of the rehearsal/practice space of musical performance in the system design. The design of _derivations addresses the characteristics of Paine’s conversational model of interactivity in a broader temporal frame, taking into account the unique and cumulative nature of the rehearsal/practice space in musical performance in the design of the system itself.

2.2 The Phrase Database

The system uses recorded audio and spectral information stored throughout a performance as the source material from which new sonic gestures are formed through later processing and re-synthesis. Continuous live sampling of the incoming audio stream provides the system with access to all that has been played by an instrumentalist throughout an improvised session. A single audio buffer of a user-defined length (default 15mins) is used to record this stream of audio throughout the duration of a performance. The audio is simultaneously time-stamped and assigned a phrase index for later reference. In this system, phrases are detected in the audio signal through the use of a simple envelope follower which reports the end of a performed phrase after a given silence threshold in the audio signal has been reached (a similar technique has also been described in [10]). This particular technique was chosen due to the inherent phrasing of wind instrumentalists, whereby musical phrases are often directly linked to the player’s breathing patterns during performance. To account for the use of instruments unaffected by this biological restriction (stringed instruments, piano, percussion etc), the system also provides a maximum phrase length with which it can automate the segmentation of the audio signal into phrases.

In addition to continuously sampled audio, the system also builds a database of sinusoidal models analysed from the live signal, making use of the iana~ external object for this analysis. [11] To build the database, the analysis module extracts sinusoidal models of the audio signal comprised of up to fifty frequency and amplitude pairs, sampled continuously at an interval of ten milliseconds throughout the duration of each phrase. Each model is then associated with the phrase index in which it was analysed, enabling the subsequent re-synthesis to

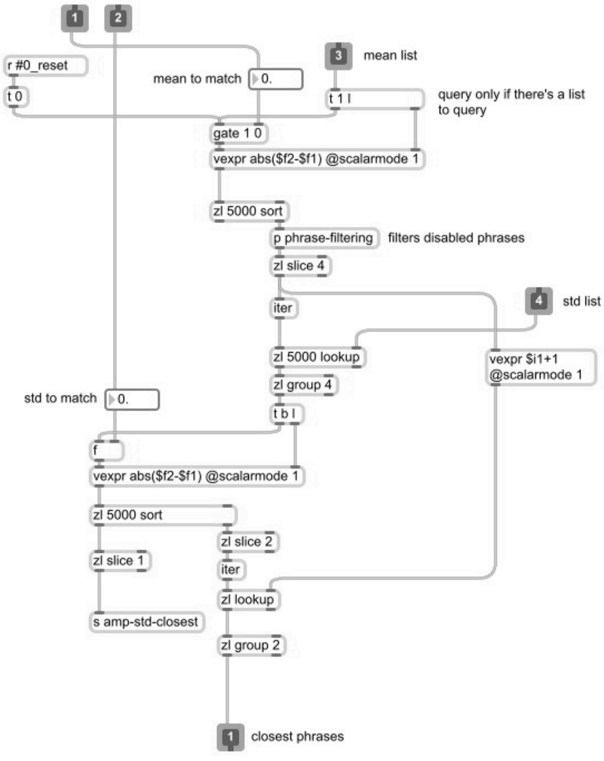


Figure 1. Phrase matching algorithm. Input queries of most recent mean and standard deviation are compared against the statistics database to find closest match

reference groups of sinusoidal models with respect to their corresponding phrase in the database.

2.3 Phrase Matching

An integral part of _derivations' automated decision making is its ability to define and access previously performed phrases with respect to their overall timbral content measured across the duration of the phrase, as compared with the most recent phrase performed by the instrumentalist. Making use of Tristan Jehan's analyzer~ object [12], the system collects the continuous output of four sound descriptors from the live signal throughout the length of each phrase. The descriptors streamed from analyzer~ include *pitch, loudness, brightness* (spectral centroid) and *noisiness* (a measure of spectral inharmonicity). Upon the completion of a phrase, the system computes both the mean and standard deviation of the accumulated data for each descriptor, storing the resultant eight statistical values in a database indexed to each individual phrase.

The system response in _derivations relies upon this database of averaged sound descriptor data to select phrases stored in the expanding database that best match the most recently analysed phrase performed by the instrumentalist. This matching process is described as follows:

For each of the four sound descriptors, the mean value of the most recently analysed phrase is compared against the database of all of the means calculated for that descriptor. The algorithm then returns the four closest matching phrase indexes to this input. Following this, a further comparison is made between the standard deviation value of the input phrase and the standard deviation values of the four phrase indexes returned previously (Figure 1 illustrates this process in the form of a MaxMSP sub-patch). This process is executed simultaneously across all four sound descriptors, returning a total of eight phrase indexes (two per descriptor) as potential matches. The phrase index with the highest number of occurrences across these eight matches is chosen for selection. Ties are broken at random in the case of two or more phrase indexes with an equal number of occurrences, whereas in cases where no descriptors return the same index, a random choice is made from the between top ranking index chosen by each descriptor.

2.4 Synthesis and Processing

The sonic response in _derivations is a combination of audio processing and re-synthesised spectral data. The three synthesis modules used in the system are a bank of four phase vocoders, a granular synthesiser and a sinusoidal re-synthesis module, all of which reference the central phrase database for their initial sonic materials. Although each module differs with respect to their individual processing and synthesis parameters, the manner in which these parameters are altered over the length of a selected phrase is consistent throughout. When accessed by a processing module, any given phrase is treated as sonic and gestural source material from which new gestures and temporal associations are formed, and not replayed directly as it appears in the recorded buffer. Each phrase chosen and sent to the modules is therefore treated as a discrete unit of time, dislocated from its original temporal context. The length of the original phrase provides the temporal boundaries within which all processing and synthesis parameters are contained. Below the three synthesis modules in the _derivations system are described in detail.

2.4.1 Phase Vocoders

The phase vocoder module is comprised of four identical yet independent phase vocoder based phrase players built around the flexible and high-quality supervp.scrub~ object from IRCAM's supervp~ library of objects.[1] In the global control module, the user is given the option of defining the maximum number of players that can be active at any one time, so as to control the potential density of material output from this module in performance. With respect to the temporal profile of a given phrase, a flexible approach to both playback speed and direction was desired, and as such the module makes use of an automated scrubbing algorithm constrained by a user-defined speed range upon output of a chosen phrase. As the phase vocoder enables the independent manipulation of both speed and pitch, the transposition of phrases is also treated in a flexible manner within this module. This is enabled through the use of a global transposition range applied across all four players, with phrases output at either a fixed or gliding transposition value constrained within a user defined range.

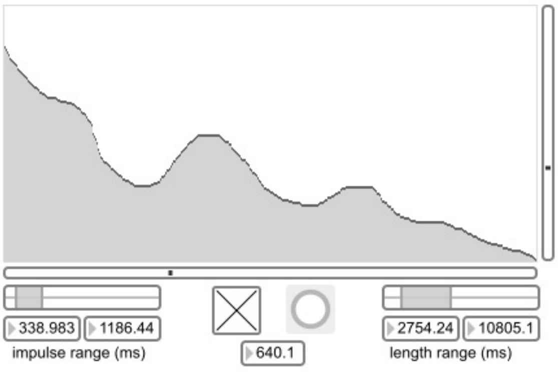
2.4.2 Granulator

The granulator module represents a further layer of abstraction of the segmented phrases stored and recalled during an improvisation with _derivations. Each discrete processing parameter is treated independently to process an input phrase, enabling a great variety of resultant sonic gestures. Processing parameters such as grain duration, panning and transposition are constrained by user-defined random ranges, whilst grain density and scrubbing position are controlled by automated scrubbing algorithms identical to those found in the phase vocoder module.

2.4.3 Pitch Models

Pitch models is a polyphonic additive synthesiser, representing the re-synthesis counterpart to the sinusoidal decomposition and storage module described in section 2.2. Upon output each sinusoidal model is assigned a unique voice in the synthesis algorithm, allowing for polyphonic spectral variation of the individual sinusoidal models. These variations take the form of pre re-synthesis interpolated partial amplitude scrambling and vocal formant filtering post re-synthesis. A unique aspect to this module is the flexible manner in which these sinusoidal models are output, both with respect to their temporal context and the synthesis variables available within the module.

Ensuring variety in the output of the stored models is a system of user specified presets for up to 23 discrete parameters available in the module. The diverse range of parameters accessible in the preset bank includes standard synthesis parameters (ADSR, transposition), minimum and maximum ranges for random processes (partial scrambling, transposition



338.983 1186.44
impulse range (ms)

640.1

2754.24 10805.1
length range (ms)

Figure 2. Example of a user-defined rhythmic function within the Pitch Models module

[1] *SuperVP* Max/MSP modules are released in the IRCAM Forum. http://forumnet.ircam.fr/708.html?L=1

range), as well as parameters affecting rhythmic output of the models themselves. In performance, these presets are automatically interpolated between using a simple random walk algorithm, with the interpolation rate mapped to the length of the phrase currently chosen for output. The rhythmic output of the stored models is controlled by user specified rhythmic functions that are executed upon the triggering of a phrase. As the functions are also included in the preset bank, the interpolation between one preset and another causes great variety in both the rhythmic and polyphonic output of the stored models.

3. REHEARSAL DATABASES/USER INTERFACE OPTIONS

As suggested previously, the design of interactive performance systems may benefit greatly from a consideration of the rehearsal/practice space as an integral part of any interactive performance paradigm. The _derivations system is currently attempting to address this aspect of musical interactivity through an extension of Paine's conversational model of interaction to encompass the cumulative interrelationship present both inside and outside of a performance time interaction. By definition this then includes a privileging of the role of the performer as an active and creative decision maker in this process. Below I will briefly describe how this is currently being applied within the _derivations system.

3.1 Cumulative Rehearsal Databases

In the current system design, following an improvised session with the software a performer has the option to save all of the stored information from the current performance to disk in a collection of rehearsal files for later re-use. The phrase, statistics and sinusoidal model databases collated throughout a performance with _derivations are saved to disk as text files, with the addition of a master reference file containing the file paths of all files saved to disk. The audio recorded throughout the duration of the current improvisation is saved from the global audio buffer to disk in its entirety as one continuous AIFF audio file. Within _derivations, the text files are designed to be cumulative so as to enable the merging of many subsequent rehearsal databases into the one file.

Upon launching the software, a performer may choose to load a saved rehearsal database containing any number of pre-analysed rehearsal sessions. This action loads the collated database files saved previously to disk into the internal data collections of the system, including the master file that points to the location of the various saved audio files. In order to optimise system memory and improve the efficiency of the system, full-length audio files containing previously analysed sessions are not read into system memory in their entirety, but segmented as individual phrases *only when selected* by the phrase matching algorithm to be sent for processing. This is achieved through referencing the original time-stamp information contained in the phrase database to selectively load individual phrases of varying lengths from these larger audio files into the system dynamically.

Once a database of any number of previous rehearsal sessions has been loaded, a performer may choose to perform with or without the default real-time phrase storage component enabled. In practice, this enables a musician to effectively define in advance the entire reservoir of source material to be used by the system before a performance begins. Further customisation of the loaded rehearsal databases includes the ability to disable individual phrases, or even entire rehearsals from being selected for output by the matching algorithm.

Each of the above mentioned user options are easily accessible to the performer in the rehearsal/practice space through the aid of an intuitive user interface, enabling the system to become an exploratory environment for the performer both inside and outside of a performance time interaction with the software. It is suggested that this can further enhance the unique creative potential of interactive performance systems of this type, through a simple consideration of the potential of the rehearsal/practice space in the design of such systems.

4. CONCLUSIONS

This paper has presented the current iteration of the author's _derivations system, a new interactive performance system for solo improvising instrumentalist. The system's design criteria and implementation have been discussed, with specific reference to the system's use of timbral matching to relate the most recent performance state of the instrumentalist with a large and expanding database of phrases analysed and stored throughout an improvised session with the software. The specific synthesis and processing modules used by the system have been discussed in detail, and the current implementation of cumulative rehearsal databases in this system has been proposed as an initial point of departure towards the expansion of Paine's conversational model of interactivity to include the rehearsal/practice space in interactive musical performance [6].

5. REFERENCES

[1] Lewis, G. Too many notes: Computers, complexity and culture in voyager. *Leonardo Music Journal*, 10 (2000), 33-39.

[2] Collins, N. and d'Escrivan, J. The Cambridge Companion to Electronic Music. Cambridge University Press, City, 2007.

[3] Hsu, W. Strategies for Managing Timbre and Interaction in Automatic Improvisation Systems. *Leonardo Music Journal*, 20 (Dec 2010), 33-39.

[4] Rowe, R. Interactive Music Systems: Machine Listening and Composing. The MIT Press, Cambridge, MA, 1992.

[5] Ciufo, T. Beginner's mind: an environment for sonic improvisation. In *Proceedings of the International Computer Music Conference* (Barcelona, 2005).

[6] Paine, G. Interactivity, where to from here? *Organised Sound*, 7, 03 (Dec 2002), 295-304.

[7] Bastien, D. T. and Hostager, T. J. Cooperation as communicative accomplishment: A symbolic interaction analysis of an improvised Jazz Concert. *Communication Studies*, 43, 2 (Jul 1992), 92-104.

[8] Johnston, A., Marks, B. and Edmonds, E. 'Spheres of Influence': an interactive musical work. In *Proceedings of the Second Australasian Conference on Interactive Entertainment* (Sydney, 2005).

[9] Hsu, W. and Sosnick, M. Evaluating Interactive Music Systems: An HCI Approach. In *Proceedings of New Interfaces for Musical Expression* (Pittsburgh, 2009).

[10] Hsu, W. Managing gesture and timbre for analysis and instrument control in an interactive environment. In *Proceedings of the 2006 conference on New interfaces for Musical Expression* (Paris, 2006).

[11] Todoroff, T., Daubresse, E. and Fineberg, J. Iana~ (A Real-Time Environment for Analysis and Extraction of Frequency Components of Complex Orchestral Sounds and Its Application within a Musical Context). In *Proceedings of the International Computer Music Conference* (Banff, 1995).

[12] Jehan, T. and Schoner, B. An Audio-Driven Perceptually Meaningful Timbre Synthesizer. In Proceedings of the International Computer Music Conference (San Francisco, 2001).