

USING TIMBRE IN A COMPUTER-BASED IMPROVISATION SYSTEM

William Hsu

Department of Computer Science
San Francisco State University
San Francisco CA 94132
USA
hsu@tlaloc.sfsu.edu

ABSTRACT

Timbre is an important structural element in non-idiomatic free improvisation [1]. An interactive software system that improvises in such a context should be able to analyze and respond to gestures with rich timbral information. We have been working on a system that improvises with British saxophonist John Butcher, whose complex vocabulary incorporates extended saxophone playing techniques, amplification and feedback [16]. We classify Butcher's saxophone sounds into broad perceptual categories that might be useful for improvisers. The general behavior of our system is responsive to real-time timbral variations. The emphasis is on working with abstract sound, gesture, and texture, rather than more traditional parameters such as pitch and harmony.

1. INTRODUCTION

Rowe's classification scheme [5] identifies some improvisation-oriented interactive music systems as following a *player paradigm*; such a system behaves like an improvising partner in a performance with a human musician. The system should be able to analyze and "understand" aspects of an improviser's gestural language that might be perceived as significant by other human improvisers.

In non-idiomatic group improvisation, musicians clearly take timbre into account when making performance decisions. This is critical when working with saxophonists or other instrumentalists who have made extended techniques important components of their approaches, from pioneers such as Roscoe Mitchell and Evan Parker, to today's virtuosos such as John Butcher and Peter van Bergen.

We have been working on a software improvising system in close collaboration with British saxophonist John Butcher, commonly regarded as an innovative improviser who has greatly expanded the timbral palate of the saxophone (see, for example, [16]). We felt that many improvisation-oriented computer music systems are limited in their interactivity, because they do not make sufficient use of timbral information.

The design goals for our system were these:

- 1) The system will be used in the context of free improvisation.
- 2) There will be minimal use of looping or sequencing, i.e., the system will behave in unpredictable ways, like an improviser.
- 3) The system will be responsive to timbral variations in the saxophone sound.
- 4) It should work with the range of Butcher's saxophone vocabulary, from extended techniques, to small close-miked sounds, to saxophone-controlled feedback through a sound system.
- 5) The system will not be a purely player paradigm system in the sense of Rowe. That is, there will be options for a human to intervene and influence the larger shape of the system's behavior.
- 6) Overly obvious mappings of saxophone gesture to computer-generated gestures should be minimized.

Butcher and I worked closely together through residencies at the STEIM studios and a concert at Kraakgeluiden Werkplaats in Amsterdam. A few excerpts from our STEIM sessions are at <http://userwww.sfsu.edu/~whsu/Timbre>.

This paper will describe the overall organization of the system. We will present a selected survey of related work, overview how we extract timbral categories from the real-time audio stream, and describe how this information shapes the behavior of a virtual improvising ensemble. Finally, we will evaluate the current state of the system, and discuss future directions.

2. RELATED WORK

George Lewis' *Voyager* [3] is driven by a human's real-time performance plus its own internal processes. A pitch-to-MIDI converter transforms the real-time audio stream to MIDI data; *Voyager* works with the converted MIDI input, rather than directly with audio. Matt Ingalls' *Claire* has mostly been used with his clarinet and bass clarinet. *Claire* also uses a pitch-to-MIDI converter; MIDI output controls tone modules or a Yamaha Disklavier [4]. Both *Voyager* and *Claire* are player paradigm systems.

Lawrence Casserley has performed and recorded extensively with saxophonist Evan Parker, using his ISPW-based system [13]. Phil Durrant has also worked with John Butcher, using effects units to process Butcher's saxophone sound. Each system behaves like an "extended instrument" controlled directly by a computer musician. Timbral information is not extracted for configuration and decision-making.

In [5], Rowe overviews several interactive systems/pieces that work primarily with MIDI data. He also discusses aspects of Zach Settel’s piece *Punjar*, in which timbral characteristics, such as sibilance in the delivery of a vocalist, are used to influence synthesis. Lippe describes in [6] his *Music for Clarinet and ISPW*, and discusses how timbre might be used to control material generation. Puckette and Lippe [7] also discuss general approaches to using timbre from a live audio stream to influence control.

Ciufo describes in [8] a system for guitar that combines sensors and real-time audio to control processing, using brightness, noisiness, and other parameters measured through Jehan’s MSP external analyzer~ [9]. Our system uses a larger set of timbral categories to coordinate an ensemble of virtual improvisers.

3. SYSTEM ORGANIZATION

Our goal is to construct an interactive computer music system that is able to monitor real-time input from improvisers, extract timbral characteristics, and use timbral information in its decisions for generating response material. Figure 1 shows the high-level organization of our system.

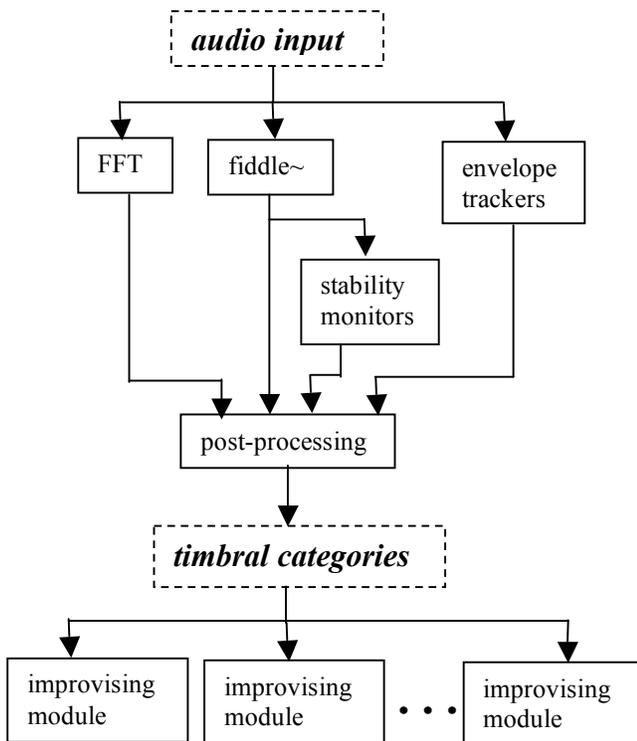


Figure 1: High level organization of system

The audio input stream (Butcher’s saxophone sound) is fed into analysis modules. The raw measurements are post-processed to yield broad descriptive categories for timbre, and other performance characteristics. This information is monitored by an ensemble of improvising

modules. Each module “performs”, based on its internal processes and extracted information from audio input.

4. TIMBRE ANALYSIS AND CLASSIFICATION

We first developed a framework for analyzing the timbre of an instrument, and forming broad timbral classifications. In improvisation, decisions need to be made promptly; a human improviser is more interested in whether a tone is rough versus smooth, rather than how a rough tone is produced. Our emphasis is on broad perceptual categories; we approach timbre largely from a listener’s perspective. We also monitor higher level parameters such as pitch range, gesture length, etc. This paper will focus primarily on timbre-related processing.

4.1. Timbral gestures and categories

Timbral variation can be an integral component of musical gestures in improvisation. For example, a long tone might be held, with stable pitch and loudness, but the intensity of multiphonics is slowly increased through embouchure control. An experienced human improviser would perceive and respond to this gestural variation.

We propose the following timbre categories as a starting point for our descriptive framework. A saxophone tone might be described as 1) *noisy* (vs. not noisy); 2) containing *harmonic partials* (vs. inharmonic partials); 3) containing a *sharp attack* (vs. no sharp attack); 4) containing *multiphonics* (vs. no multiphonics); 5) *rough* (vs. smooth).

We will describe the measurements made by our system, and how they are used to identify timbral characteristics.

4.2. Measurements and post-processing

Our system was constructed in the Max/MSP environment. Extensive post-processing of the raw measurements was often necessary to extract usable information and produce the timbral categories that we needed.

At the initiation of our project, Jehan’s analyzer~ [9] was not publicly available. We worked mostly with the standard MSP FFT objects, and Puckette’s fiddle~ for pitch estimation and partial tracking [10]. While fiddle~ (and similar pitch trackers) work reasonably well for clean, sustained tones, its results can be unreliable near the attacks and decays of a tone, or when the tone itself is noisy or has a complex and unstable spectrum. We monitor the stability of the pitch estimation from fiddle~, over several analysis windows. Since the behavior of strong partials tend to mask the behavior of weak ones, we configured fiddle~ to produce the twelve partials that are lowest in frequency, and sorted them to select the six strongest ones for further analysis.

Additional raw measurements include: relative spectral centroid (absolute spectral centroid divided by the

estimated pitch), zero-crossings, peak energy distribution (strength of top ten FFT peaks relative to overall energy in signal), and the presence of very sharp onsets. We also monitored the mean and variance of most measurements, over several analysis windows. The measured data is only considered “meaningful” to the analyzer modules when the energy in a frame is above a tunable threshold.

4.3. Identifying timbre categories

Noisiness refers to the prominence of breath noise in a saxophone tone. Noise usually results in an extremely unstable pitch estimate from a pitch tracker like fiddle~. In addition, its energy tends to be widely distributed across its spectrum, rather than concentrated at a few spectral peaks. If our measurement of peak energy distribution is below a threshold, and the pitch estimate is extremely unstable, we classify the signal as noisy.

Prominence of inharmonic partials is detected by comparing the frequencies of the six strongest partials obtained from fiddle~. The pitch estimate should be relatively stable.

Presence of sharp attacks corresponds to techniques such as slap tongue or amplified key clicks. These are identified by steep rises in the amplitude envelope.

We found the *presence of multiphonics* to be strongly correlated with a stable pitch estimate, a concentration of energy in relatively few spectral peaks, and a high normalized centroid. We do encounter “false positives” with rich tenor tones, and have trouble identifying closely-spaced multiphonics. We are currently working on improving our identification of this category.

We based our roughness estimate on measurements of the fluctuation of waveform amplitude envelopes, after the work of Pantelis Vassilakis [11]. A “rough” tone can be produced by techniques such as flutter-tongue and throat tremolo. We place in this category tones whose amplitude envelopes fluctuate periodically, with a deviation of more than 10% about its average value, at frequencies of about 10 to 50 Hz.

5. MATERIAL GENERATION

5.1. Language and materials in free improvisation

Free improvisation has emerged as a cohesive movement since the 60s [1]. While the choice of material is very open, the general practice is to avoid references to established idioms.

In free improvisation, the role of pitch tends to be downplayed or obscured; greater weight is placed on loudness, duration, and timbre. Pitch choice is likewise of secondary importance in our system. Greater effort is placed on managing duration and, especially, timbre.

The use of large gestures that may attract undue attention is always carefully managed by improvisers. Similarly, our system works more with smaller gestures that incorporate nuanced timbral changes. Drones and thicker textures can also be generated, with parameters that are influenced by the audio input. With the decreased importance of pitch, timbre and the rate of change of timbral parameters become more important structural elements. Gestures whose features evolve slowly are perceived very differently from gestures whose characteristics undergo abrupt and rapid modifications. Gesture generation in our system involves the pseudo-random selection of a number of parameters, within tunable ranges, their rates of change, and how they might be influenced by audio input.

5.2. Choice of improvising agents

We organized the generative components of our system as a small virtual ensemble. Each agent controls a module that transforms the input audio stream, or “plays” a virtual instrument. Each module receives a stream of messages describing the timbre and other characteristics of the input sound.

A module may join or leave the ensemble at any time. It may generate material solely according to its internal processes, or may “perform” when it detects specific combinations of timbral or performance characteristics in the audio input. The gestural parameters of the generated material might be influenced by the current or past timbre of the saxophone sound.

Our sound transformation modules include a classic effects chain with a comb filter, flanger/chorus, and pitchshifter. Another is a granular synthesis module from Nathan Wolek’s granular toolkit [15], which was useful for working with the saxophone’s clean sound.

We chose synthesis modules that evoked wind instrument sounds, to blend with the saxophone. To work effectively with Butcher’s performance range, these should be capable of a variety of gestural and timbral nuances. One module is a noise generator with a resonance filter, with variable brightness, roughness, envelope shapes etc. We also implemented a waveguide-based bass clarinet module, with a variety of envelope and embouchure effects.

For contrast, we implemented a modal synthesis module that simulates the sounds of resonating metallic objects, such as bells and woks. Clusters of the resonance peaks can be controlled independently, to provide natural-sounding timbral variations. We also coded a simple acoustic guitar player based on the mandolin~ object from the Percolate library of physical modeling objects [14]. A finite state machine provides high-level control of guitar gestures. Transition between states is influenced by the module’s internal algorithm, along with timbre and performance information.

7. REFERENCES

5.3. Interaction design and coordination

We designed our modules such that they may act independently of each other, or form coordinated subunits within the ensemble. For example, several agents may coordinate to form clouds of short gestures; the density, frequency range and other parameters of these clouds may be influenced by the timbre of the audio input.

While each improvising module is able to swiftly respond to the changes it perceives in the audio input, a user should be able to make some organizational and structural choices during a performance. In this respect, our system has similarities in concept to Butch Morris' *conductions*, in which improvisers are coordinated through real-time gestures [12]. In our system, a user can choose the combination of modules that will participate in a performance, the timbral categories and other performance characteristics each agent may respond to, types of gestures and parameter ranges that might be used, and the manner of coordination between some groups of modules. These parameters and strategies can be changed in the middle of a performance. In general, a user does not directly control the gestures of each improvising module, only the larger overall shapes of the improvisation.

6. OBSERVATIONS AND FUTURE WORK

We found that our system is evocative of a responsive improvising ensemble. Pre-configuration is not necessary; the system is capable of monitoring the audio input stream, and activating a selection of modules to begin playing. The behavior of each agent is responsive to changes in timbre, at a speed that is clearly impossible if we rely solely on direct human control.

Early versions of the system had an undesirable characteristic: when the saxophonist paused, the system eventually became silent, because timbral categories were ignored at low signal levels. To alleviate this problem, we gave more autonomy to some agents. We also added "memory" to the broadcast stream of timbral information; during a pause, the timbral description for the past few seconds is looped and still available to the agents. Hence, each agent continues to "hear" similar sounds, and respond in a consistent manner.

This is very much a work in progress. We are working on more transformation and synthesis modules, and more sophisticated module coordination strategies.

Finally, we would like to thank STEIM (Amsterdam), Chris Burns at CCRMA, and Anne LaBerge at Kraakgeluiden Werkplaats (Amsterdam), for their support through the development of our project.

- [1] Bailey, D. *Improvisation: Its nature and practice in music*. Da Capo Press, 1993.
- [2] Rowe, R. *Interactive Music Systems*. The MIT Press, Cambridge, Massachusetts, 1993.
- [3] Lewis, G. "Too Many Notes: Computers, Complexity and Culture in *Voyager*", *Leonardo Music Journal*, Vol. 10, 2000.
- [4] Ingalls, M. *Personal communication*, 2001.
- [5] Rowe, R. *Machine Musicianhip*. The MIT Press, Cambridge, Massachusetts, 2001.
- [6] Lippe, C. "A Composition for Clarinet and Real-Time Signal Processing: Using Max on the IRCAM Signal Processing Workstation", *Proceedings of the 10th Italian Colloquium on Computer Music*, Milan, Italy, 1993.
- [7] Puckette, M. and Lippe, C. "Getting the Acoustic Parameters from a Live Performance", *Proceedings of the 3rd International Conference for Music Perception and Cognition*, Liege, 1994.
- [8] Ciuffo, T. "Design Concepts and Control Strategies for Interactive Improvisational Music Systems", *Proceedings of MAXIS International Festival/Symposium of Sound and Experimental Music*, Leeds, UK, 2003
- [9] Jehan, T., Schoner, B. "An Audio-Driven Perceptually Meaningful Timbre Synthesizer", *Proceedings of the ICMC*, 2001.
- [10] Puckette, M., Apel, T., Zicarelli, D. "Real-time audio analysis tools for Pd and MSP", *Proceedings of the ICMC*, San Francisco, 1998.
- [11] Vassilakis, P. *Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance*, Doctoral dissertation, University of California, Los Angeles, 2001.
- [12] Morris, B. *Current Trends in Racism in America*, program note, Sound Aspects LC 8883, 1985.
- [13] Casserley, L. "A Digital Signal Processing Instrument for Improvised Music", *Journal of Electroacoustic Music*, Vol. 11. See also http://www.chiltern.demon.co.uk/DSP_Inst.html
- [14] Trueman, D., Dubois, L. *PerColate* website, <http://www.music.columbia.edu/PerColate>
- [15] Wolek, N. "A Granular Toolkit for Cycling74's Max/MSP", *Proceedings of SEAMUS 2002*.
- [16] Keenan, D. "Mining echoes", *The Wire*, November 2004. (See also <http://www.johnbutcher.org.uk/>)

