Getting the Acoustic Parameters from a Live Performance

by Miller Puckette and Cort Lippe

Introduction

The computer is not like acoustical instruments. It can produce a fantastically wide range of sounds with an unheard-of degree of control over pitches, timbres, and their evolutions in time. The computer is harder to make music with than an acoustical instrument because there are many more choices to make. We do not bow a violin expecting to make the sound of a trumpet. But the relative simplicity of playing the violin does not stop serious violinists from practicing many hours a day throughout their careers. The instrument player intends not only to hit the right notes but also to play musically, with attention to phrasing, articulation, and all the rest of it. The difficulty in making music with the computer is in finding ways to navigate in the huge space of possible computer-generated sounds. When the instrument is the computer, much effort is required to approach the desired sonic effect, and getting the phrasing right is often extremely hard. Doing this ``live,'' i.e., specifying all those parameters in real time in front of an audience, seems impossible.

An alternative is to treat the computer as a studio instrument, working for hours to tune each measure of sonic output. This is a perfectly valid way to make music, albeit lacking some of the interesting character of live music. There is no concept of ensemble---no way to play two computers together if they are each being programmed out of real time. And the fact that every note requires a database to specify it, implies that it will be impractical to make all the experiments that one could with a live instrument; in effect there is much less time for practice.

Why live performance?

This is not to say that the reason live music might be superior to tape music is because it is unpredictable. Most musicians of all cultures spend a certain amount of time learning to play an instrument with an acceptable degree of technical accuracy. More importantly, musicians spend a large amount of time developing their ability to play expressively. Globally, this expressivity is what distinguishes one performer from another, and locally, distinguishes various performances of a piece by the same performer. By the time a first-rate player goes on stage to play a pre-composed piece of music, essentially all the technical decisions have been made. Minor adjustments are perhaps necessary to correct for the fact that the auditorium, having more people, is acoustically less live during the performance than during rehearsal. But two performances on consecutive nights will put forth essentially the same technical results. Despite the fact that pre-composed live music is largely predetermined, the music still gains much from being played live. One reason is perhaps that the presence of an audience puts the performer in a heightened state. Also, the on-stage decisions which are made during the career of a performer have a cumulative effect. The performance is informed by all the previous nights simultaneously. Any small or large improvement which is found one night can inform every performance thereafter. The palette of expressive variance a performer can offer listeners over multiple performances is, for the most part, based on decision-making in realtime, whether during the concert itself or during some previous rehearsal or concert. This is what makes repeated playing of a piece of music interesting for a performer. If each time we, as listeners, put on our favorite CD and heard a different interpretation by our favorite artist playing our favorite tune, we would have smaller CD collections.

In addition to the problem of the computer being a much higher-dimensional instrument than any other, there is also the lack of trained ``computerists.'' This is not surprising. Concert violinists often have started at three or four years of age. Nobody puts their three-year-old son or daughter up to playing the computer. We suppose that the computer, ten years from now, will only bear a slight resemblance to the computer of today and all those hours of practicing with a WX7 (or whatever) will be wasted. A major problem in live computer music is that of defining a standard user interface that players can become expert at manipulating.

Arriving at and standardizing one or more computer-music human interfaces might well take decades. Two temporary solutions are available. We can use the interfaces we have defined to date, renouncing the possibility of expert performance, or we can use real instruments (or faithful imitations of them) as input devices. In the latter case there is no problem finding expert players;

instead, the difficulty lies in defining interesting semantics that a computer could associate with the player's control input. We must also make a choice between building special instruments, such as Larry Beauregard's flute, which acts simultaneously as a flute and a control input device (fingerings are detected), or rather taking the instrument's audio output directly as the control stream. The former choice is more practical for keyboard instruments where we can add switches to the keys in a way which does not affect the play of the instrument; but in the case of most instruments -- strings, brass, woodwinds, non-keyboard percussion -- economics argues for the second choice. Who is going to implant an RF system in a Stradivarius violin?

Post-processing the instrumental sound

The most direct way to put the player of a live instrument in control of a live electronic sound is to derive the electronic sound directly as a transformation of the instrumental one. Many examples of this have been proposed and used in electronic music. An early example is Stockhausen's {\em Mixtur} for orchestra and live electronics \cite{ref-stock}, in which the live transformations were done exclusively using ring modulators; most of the variety in the live electronic sounds was obtained by varying the instrumental sound transformed.

Composers have since looked for techniques which offer more freedom to write live and electronic parts with more independence. Ideally, one could write separate parts for the computer and the instruments, sometimes playing together, sometimes alternately, sometimes in counterpoint, and so on. Also, composers would like the option of deriving the electronic sounds less directly from the instrumental sounds without losing the unity that the two parts derive from their common source.

A recent musical example showing progress in these directions is Philippe Manoury's {Jupiter} for flute and live electronics (1987). In this piece a majority of the electronic sounds are provided by three simple operators: a single sideband modulator, a bank of pitch shifters, and a reverberator. The three are configured so that any one can feed into any other. In the opening measures shown in Figure 1, five different notes played by the flute -- all low C sharp -- are transformed into a sustained chord of six notes.

In this example, the dynamics and timbres of the notes in the chord are all taken directly from those of the flute. At the same time, the electronic part does not have to correspond to the flute part note by note; a melodic passage played by the flute has become a chord in the accompaniment.

To operate at this level of detail it is essential to have some kind of score follower, a software object which synchronizes the electronic part with the instrumental one \cite{ref-puckette}. The synchronization must be quite accurate temporally. The search for more accurate and more robust score followers is ongoing. Using many different types of cues we can imagine score followers which are sensitive to many kinds of musical input: middle C, trills, a crescendo, sul ponticello playing, or even silence can all become useful information.

Taking the instrument's timbre as an abstract control

Real-time decision-making which a musician does while interpreting a piece of music in a performance situation is quantifiable to a certain degree. We can track pitch, amplitude, and timbre in real time using a computer with some level of accuracy. We can use this information to control and influence compositional, signal processing, and sound generation procedures in real time.

How to use this information is a question best left to the individual composer. But recognizing what a musician is doing on as many different levels as possible gives composers correspondingly more ways to answer this question. For example, pitch tracking can be used to distinguish different pitches and determine the stability of pitch on a continuous basis. On a musical level this means we can safely start to distinguish portamento, glissando, trills, tremolo, etc. As for amplitude, envelope following of the continuous dynamic envelope can be the starting point for all sorts of articulation detection: flutter-tongues, staccato, legato, sforzando, crescendo, etc. In the short term frequency domain, FFTs, pitch tracking, and filtering can be used to track continuous changes in the spectral content of sounds for identifying things like inharmonic/harmonic ratios and timbral brightness, which are useful in recognizing multiphonics, sul ponticello, etc. Thoughtful high-level event detection which combines the analyses of frequency, amplitude, and spectral information can provide rich control signals that reflect subtle changes found in the input signal.

Since 1993, a research program at IRCAM involving the authors, Trevor Wishart, and Stefan Bilbao, has sought to explore the possibilities of using instruments in a more declarative way as synthesis controllers. The first step is to extract signals corresponding to timbral parameters of the sound of the instrument. For any instrument, we can hope to extract an amplitude envelope and Wessel's timbral parameter (defined as the first moment of the instrument's power spectrum considered as a measure.) For instruments with a well-defined pitch, we can add both the value of the pitch and the ``pitch error'' of the sound, defined as the power of the difference of the signal from itself delayed one pitch period. In some instruments (such as strings) it is also meaningful to speak of the relative strength of the even partials of the sound as compared to the odd ones.

An interesting special case is the voice. The voice is the instrument offering the widest range of possible variation in timbre. In addition to Wessel's parameter, it is meaningful to extract formantic information from the voice, for instance, the center frequencies and relative strengths of the first three formants.

We consider the measurements listed above as {\em primary} timbral measurements in the sense that they are extracted in some sense from the momentary timbre of the sound. We can also construct several {\em secondary} timbral parameters by regarding the evolution in time of the primary ones. For example, one of the many possible definitions of ``roughness'' is a time variation in the amplitude envelope at between 50 and 150 Hz.

The voice or other instrumental sound can be regarded as a continuously moving n-dimensional joystick tied to any chosen parameters related to electronic synthesis or audio processing. The dimensions of the joystick can include, for example, pitch, loudness, Wessel number, pitch quality, and so on, plus any desired secondary parameters. The joystick can control spatialization, modulation indices, relative amplitudes of signals being mixed together, sequencer playback tempi, probability distributions, wavetable selection, time stretch factors, pitch or frequency shifter settings, or anything else that can vary in time.

We can now provide an instrumentalist with a high degree of timing control, and a certain level of expressive control over an electronic score. But how do we really measure musical expression? Recognizing a tremolo or a note played stacatto is not too difficult, but there is a danger in confusing signal analysis and musical analysis. Likewise, musical expression is far from musical knowledge. Musical knowledge is probably something best left to the composer. In any case, we would like to use the computer in ways that go beyond what George Lewis has called triggering \cite{ref-lewis}. If we recognize something and therefore do something else on a simplistic one-to-one level, we might as well embrace the current trend (brought on by the availability of software for reading soundfiles in real time) towards a ``tape music'' approach to computer music using live playlists. Intelligently using all this available analysis information presents a more interesting compositional situation than just synchronizing a performer and a computer or triggering a playlist.

A dynamic relationship between performer, musical material, and the computer can become an important aspect of the man/machine interface for the composer, performer, and listener, in an environment where musical expression is used to control an electronic score. Compositions can be finetuned to individual performing characteristics of different musicians, performers and computers can interact more intimately, and performers can readily sense consequences of their performance and their musical interpretation.

% \begin{figure} % \psfig{file=block} % \caption{A short extract of Jupiter by Philippe Manoury} % \label{fig-mofo} % \end{figure}

\pagebreak

\begin{thebibliography}{[Templaars 77]}

\bibitem [Lewis 92] {ref-lewis} Lewis, G., 1992. ``A Max Forum'', {\em Array} 13/1, pp. 19-20.

\bibitem [Puckette 92] {ref-puckette} Puckette, M., and Lippe, A. C., 1992. ``Score Following in Practice,'' Proceedings, International Computer Music Conference. San Francisco: Computer Music Association, pp. 182-185.

\bibitem [Stockhausen 68] {ref-stock} Stockhausen, K. 1968. {\em

Mixtur, fur Orchester, Sinusgeneratoren und Ringmodulatoren}, Universal Edition (London).