Objectives

The purpose of this paper is to describe a new area of computer music research — one which is certain to become a major territory for future work. The research objective can be clearly stated:

to understand the dynamics of live ensemble performance well enough to replace any member of the group by a synthetic performer (i.e., a computer model) so that the remaining live members cannot tell the difference.

The import of this advance is to move computer music clearly into the arena of live music performance. I mean here to break clear of the "music-minus-one" syndrome that characterizes tape and instrument pieces of the present; I also mean to recognize the computer's potential not as a simple amplifier of low-level switching or acoustic information (keyboards and live audio distortion), but as an intelligent and musically informed collaborator in live performance as human enquiry.

The circumstance prompting this research has been an IRCAM commission to compose a work for flutist Larry Beauregard, along with access to the 4X real-time audio processor. My immediate instinct was to return to concepts and real-time software principles I had developed in 1972, while working on the design of a large digital synthesizer with real-time performer control. Although the present implementation is written in C and runs under RTE-11 on the PER-11/55 controlling the 4X, the principles remain the same as then, and could similarly be applied to any other target system.

The control structure for modelling a synthetic performer has three main parts:

I LISTEN:
1. Catch and parse incoming live events.
2. Extract tempo, score position, loudness, etc.

II PERFORM:
1. Set new performance tempo, loudness, phrasing.
2. Organize the computer performance.

III LEARN:
1. Observe unexpected Listen behavior or Perform difficulty.
2. Keep as adjunct to the full score for future preparedness.

The three areas will now be described in more detail.

2. The Synthetic Listener

Catching and parsing a sequence of events played by a professional flutist implies pitch detection at a speed almost impossible for audio methods alone. The strategy here was to employ two additional sources of data: fingering information, and a musical score. Through a series of optical sensors installed on the keys by flutist Larry Beauregard, the list of possible sounding pitches was reduced to three. By piping the audio signal into three appropriate filters, the 4X could then resolve the ambiguity in about 35 milliseconds.

Both pitch and time information were next jointly mapped onto elements of the score, in such a way as to permit reasonable variance by the live performer. Music recognition turns out to require a significant degree of rhythmic elasticity, combined with smaller amounts of rhythmic and pitch fault tolerance. Recorded errors of two or more notes should automatically induce more rigorous pattern matching, sometimes extensive relocation. In the event of complete distress, the best strategy is to hold the current course until something recognizable occurs.

Extracting a sense of tempo from such a matched sequence requires further absorption of effects like agogic time shift. Time shifts observed within a beat can be weighted by position; modifications in the early part of a beat generally have less tempo significance than those occurring towards the end. Once a new Listen tempo has been determined and found reasonable, it is then posted for Synthetic Performer consideration.
3. The Nature of Performance

Determining the correct Synthetic Performer tempo involves two levels of action. First, at
12 milliseconds intervals the posted Listen tempo (in the form of a beat tape) is sampled and accu-
mulated in beat-bins. This serves to integrate tempo over time with fairly high resolution.
Then, about once every 200 milliseconds the beat-bins are used to determine the apparent Live Per-
former score position. This is compared with the Synthetic Performer score position, and an
appropriately graceful catch-up action determined. Just five or so such determinations per
second seems to represent adequately the manner in which performers do this kind of thing.

Modelling the physiology of performance is a
shade more tricky. My view is that the events of
an intended performance remain in strict metrical
terms until just moments before action, when they
are suddenly converted to physiological control
objects that are impossible to retract. The
period during which the human is involved in the
physiological gesture of performance will depend
on the person and on the device, but is somewhere
about one-tenth of a second. In my performance
model, once a scored event has crossed this
threshold it virtually explodes into a cluster of
active object modules, each representing some
aspect of the event (rise time, transient
effects, frequency and loudness curves) and each
needing CPU service. The processor honors these
requests on a priority queue basis, sending con-
trol data to time-tagged buffers conceptually
residing in the synthesizer itself.

Learning to Improve

The network of Control Processes represent-
ing the above can be regarded as modelling a
neural organism that is strictly instinctive,
without learned response. The most demanding
test of a Synthetic Performer is how well it
behaves in the absence of previously gathered
information -- by sight-reading, as it were, on
the concert stage. Although one cannot avoid
imbedding some stylistic bias in real-time pro-
grams during the course of their development,
there has been a painstaking effort here to limit
its effect. For example, despite the test pieces
being primarily from the late Baroque (Handel and
W.F. Bach flute sonatas), the system was able to
respond equally well to contemporary literature
(Boulez Sonatine).

I initially included a learning strategy in
the Synthetic Performer, but that code generally
remains disabled because it complicates the
development and testing of the instinctive model.
Once those aspects manage to exhibit a suitably
high level of robustness, I plan to further
develop the adaptive and learned reponse mechan-
isms. I expect that the addition of learned
responses to particular scores and to recognized
live input will vastly improve the measurable
musicanship of the Synthetic Performer. I hope
to report on such developments in the future.

Meanwhile, this paper concludes with a
demonstration of grossly different performances
of a Handel flute sonata -- the flute part played
live by Larry Beauregard, and the harpsichord
accompaniment offered in reponse by a synthetic
performer.

4. Acknowledgements

I would like to thank the Guggenheim Foundation for support of this work. Indebtedness also
to Larry Beauregard for his tireless patience and encouragement, and to Miller Puckette for hours
of discussion.