Sound Traffic Control: An Interactive 3-D Audio System for Live Musical Performance

Naut Humon
Sound Traffic Control
San Francisco, California, USA

Bill Thibault
California State University Hayward
Hayward, California, USA

Vance Galloway
Sound Traffic Control
San Francisco, California, USA

Garnet Willis
Composer and Instrument Designer
Toronto, Ontario, Canada

Jessica Grace Wing
Warmblooded Productions
New York, New York, USA

Abstract

Sound Traffic Control (STC) is a system for interactively controlled 3-D audio, displayed using a loudspeaker array. The intended application is live musical performance. Goals of the system include flexibility, ease of use, fault tolerance, audio quality, and synchronization with external media sources such as MIDI, audio feeds from musicians, and video. It uses a collection of both commercial and custom components. The development and design of the current system is described, embodying ideas developed during over a decade of experimentation, and is evaluated based on the experiences of users and developers.

1 Introduction

Sound Traffic Control (STC) made its debut in Japan in 1984, as collaboration between Naut Humon and Marianne Amacher. Since then, the system has evolved through experimentation with a variety of hardware and software systems for live musical performance using 8 or more speakers surrounding the audience.

The current system embodies and extends ideas developed by Naut Humon. These ideas include the interactive placement of 3-D sound sources, recording and sequencing of 3-D trajectories, live mixing and switching of sound sources informed by a "DJ" aesthetic, and use of live musicians. Current development is focused on incorporation of live and recorded video sources, and incorporation of 3-D ambisonic [1] display.

In the remainder of the paper, we enumerate the goals of the system, discuss the needs of its users, describe the implementation, and finally, evaluate the system and what we've learned.

2 Goals

The primary goals driving the development of STC have been: flexibility, ease of use, high availability, fault tolerance, audio quality, and synchronization with external media.

It being impossible to anticipate the precise needs of electroacoustic composers, the flow of both audio and MIDI signals in the system has been designed to provide a wide range of options. At the same time, the design of the system attempts to flatten the learning curve and simplify operation by organizing related functions into high-level modules for mixing, recording, effects processing and spatialization.

The goal of providing the highest quality audio output possible has led to the selection of equipment of professional quality. Fortunately, this class of equipment often meets our needs for flexibility. However, accessing and applying the more advanced capabilities of such systems has required the development of custom software. Limited resources have prevented development of custom realtime DSP code, however. When required, DSP is performed offline or using commercially available effects processors.

Fault tolerance is motivated by the unpleasant experience of equipment crashes during performance. A number of backup systems allow for gracefully handling crashes of different components; dual backup systems allow for the continuation of interactive spatialization of live material in the event of a matrix failure, although with some limitations.

Much of the intended musical material is strongly rhythmic, requiring synchronization effective use of spatial and other effects. Our approach has to been to automate as much of the synchronization task as possible, using 3rd party "bpm meters" when audio is the only data stream available, as well as accepting a variety of timecode formats. All MIDI control data and audio outputs can be recorded with synchronization information for later (synchronized) playback or editing.

As part of a more encompassing goal of improving the audience experience of electroacoustic music, a "conductor's station" consisting of a laptop computer and a variety of MIDI controllers has been designed. This relatively small setup would occupy the center of the "sweet spot" at the center of the room, replacing the usual battery of mixers and various gear. The conductor has control of all parts of the system, and provides a clear target of the audience's attention. The intended venues are not usually afflicted with "forward" facing seats attached to the floor.

3 Users

The system is intended to be useful to composers of electroacoustic music wishing to prepare compositions that would either be performed live using the system, or recorded for subsequent playback. The intended application of Sound Traffic Control has always been live performances of 3-D sound environments, using a mixture of live and pre-recorded material.

Composers who have used the system include Marianne Amacher, Ned Boulahassa, Xopher Davidson, Tim Digulla, Scot Gresham-Lancaster, Alan Holworth, Naut Humon, Gregory Lenczycki, and Garnet Willis.

4 The System

Work on the current version of the system began in the spring of 1996. Previous versions of the system, and several early decisions, have greatly influenced the design.

4.1 MAX

The previous incarnation of STC used Opcode's MAX [2] software for system control, and a number of useful MAX patches were deemed desirable in the new version, leading to the inclusion of a machine dedicated to running MAX.

At one point in STC's evolution, MAX was used for control of a number of audio patchbays, mixers and switchers. A graphical interface showed the current audio routing, and supported a direct manipulation interface. This interface, the "Dub Dashboard," made use of CNMAT's multislider object and MAX's support for embedded quicktime movies (Figure 1).

4.2 LCS and MixMaster

The decision to use the Level Control Systems LD-88 Supernova digital mixer [3] initially required the use of its BeOS[4]-based console, CueStation. CueStation was optimized for "show control," using pre-arranged, stored "cues" that were downloaded to the mixer and triggered by the console during a performance. The use of cues avoided communications latencies involved with updating hundreds or thousands of individual controls over a relatively slow serial connection, but also made improvised performances problematical. Also, CueStation used a 2D "SpaceMap" interface for positioning sound sources. We found the restriction of motion to a 2D surface somewhat difficult to use for placement of sounds in 3D.

These limitations led to the development of MixMaster: custom mixer-control software written in C++ for BeOS. MixMaster also provides graphical 3-D display of sound placements and trajectories. The software uses a protocol

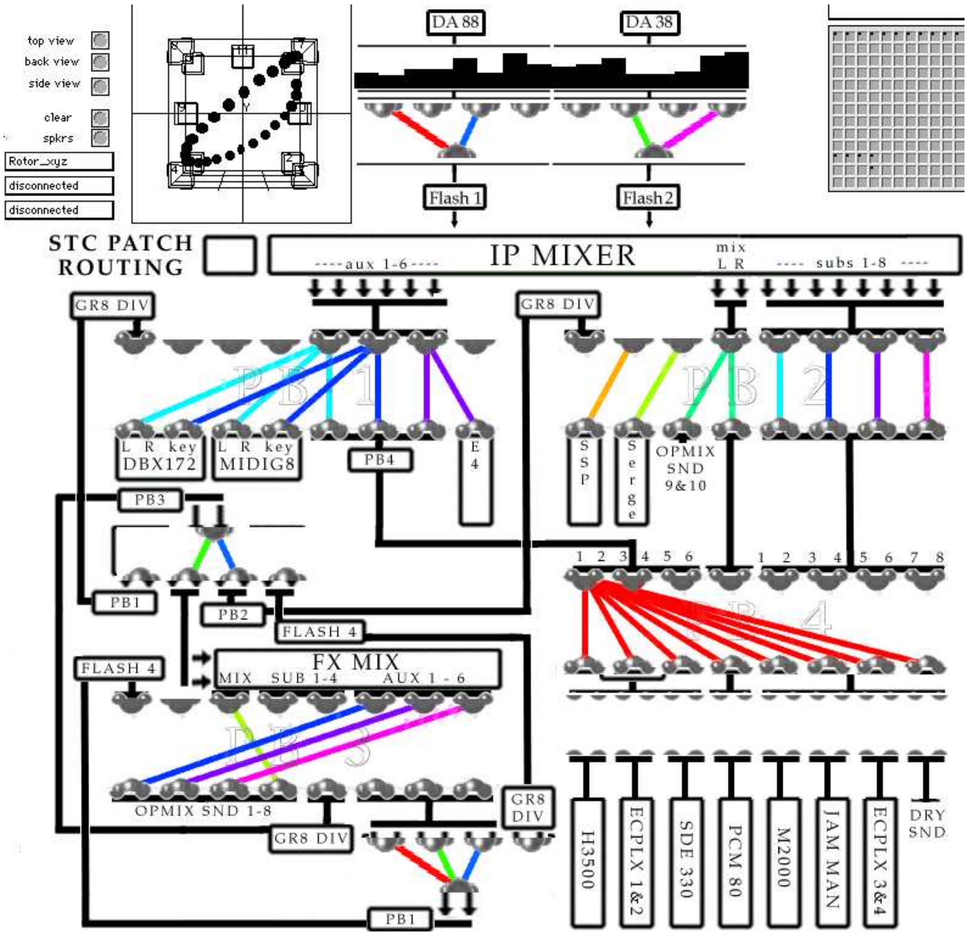

Figure 1: The Dub Dashboard: a MAX patch for audio routing.

layered over MIDI for control of matrix levels, and for a variety of higher level functions such as 3-D positions, matrix presets, programmable fader groups, proximity sensors and dynamic physics models using gravity and spring forces.

For example, the X, Y, and Z coordinates of the 3-D position for matrix input channel i are sent as key pressure (polypressure) messages for key i on channels 3, 4, and 5, respectively. (Although the use of multiple channels precludes exploitation of MIDI's running status feature, it simplifies the coding of "client" code. Besides, it isn't possible to send 3 values in a single MIDI channel message.)

Support for grouping arbitrary sets of controls into groups called fader gangs is a useful feature. Users can configure and name fader gangs, and assign them to "gang leaders:" virtual faders whose motion is followed by the gang members (also called "homies"). A commonly used fader gang consists of all of the output levels connected to speakers: control of this gang's leader provides a simple "master volume" control. (We use the term "matrix" rather loosely, to refer to the set of levels (volumes) consisting of input levels, output levels, and "matrix levels" (the set of levels that control the amount of each input that is sent to each output).)

Matrix presets allow a snapshot of all levels in the matrix to be stored for later retrieval, under MIDI control.

Physical dynamics concerns the motion of objects under the influence of forces. We include support for both spring- and gravity-like forces. This makes possible the creation of interesting trajectories that tend not to repeat themselves. 3-D control messages determine the positions of the forces, which effect "particles," associated with sound sources, in a physically plausible way. One application creates a simple "flocking" behavior by having a number of sound sources (birds) follow a gravity source (leader).

The proximity sensor feature can control a matrix level or external MIDI device based on the distance of a sound source (or 3-D cursor) to a point in space, or other sound source. One application of this feature controls a reverb send based on proximity to a particular corner of the room, simulating a reverberant space opening off of that corner. Another use of the proximity sensor controlled external samplers triggered when the 3-D cursor enters a particular region (e.g., waking a sleeping lion).

External devices communicate with the MixMaster software using simple MIDI commands such as control change and key pressure. Avoiding the use of system exlusive messages has greatly simplified the implementation of higher-level control software for sequencing and MIDI controller processing, and hides the details of the particular mixer or speaker configuration being used. The porting of sound designs between the studio and club systems is also simplified. For example, although the dimensions of the matrix and number of loudspeakers may change, commands to position sounds in 3-D do not.

A further insulation from the details of the configuration used in any particular venue is provided by support for ambisonics. An ambisonic B-format recording can be presented on any (regular) speaker array using an automatically-generated mixing matrix.

MixMaster provides a simple and compact display of matrix levels: each level is displayed with a color-coded square, whose brightness is proportional to the level, and which contains a numerical display of the level setting (0-127 in our MIDI-centric application). (Figure 2.) Groups of related controls are assigned a particular color, making it possible to quickly determine which levels are "up." (Typically, most levels in the matrix are off and displayed as black.) Clicking on the control with the mouse enables the user to change the level directly on-screen.

Live 3-D audio placement uses a variety of algorithms developed by STC: one is optimized for "cube-corner" loudspeaker placements; another supports arbitrary loudspeaker placements and control of "focus" and "size" of the sound source. A third, more specialized algorithm uses the distance between the speaker and the sound source to determine levels.

4.3 MIDI

An evaluation of available MIDI sequencing software resulted in the decision to use EMagic's Logic software[5], and has led to its central role in recording, synchronization, and playback.

The sequencer parses synchronization streams with the help of a number of Opcode Studio 5 MIDI routers: SMPTE, MIDI time code, and MIDI clock formats are all supported.

The system uses multiple computers and a wide range of MIDI input devices and MIDI-controllable effects processors. The computers serve various control functions; all computers can communicate each other, and with all attached MIDI devices, using a network of MIDI routers (Figure 3).

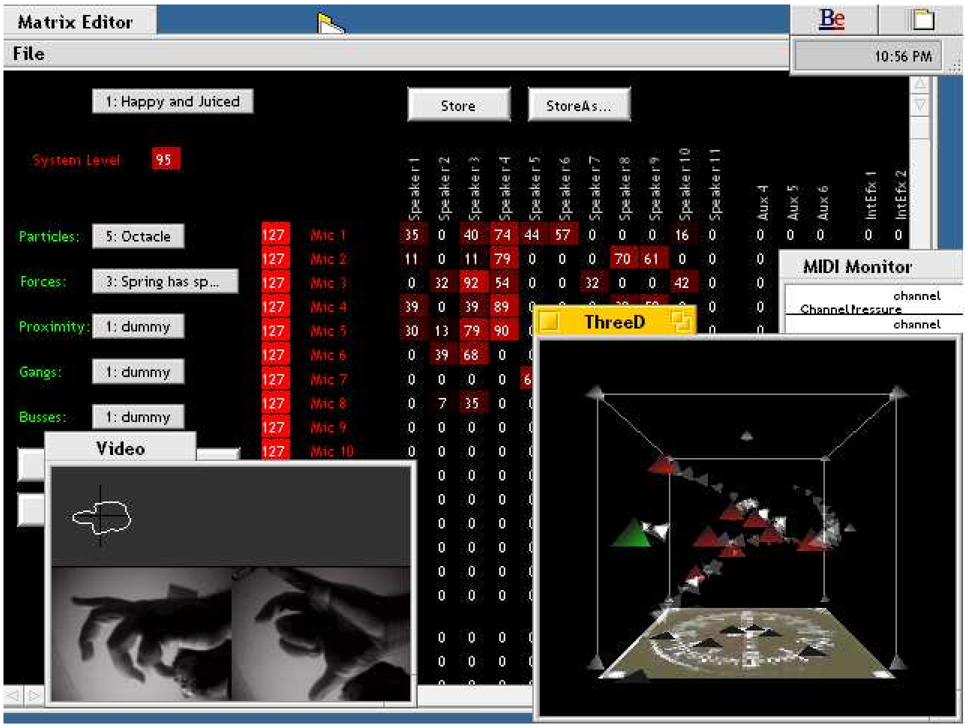

Figure 2: MixMaster's matrix and 3-D displays, being controlled with live video input.

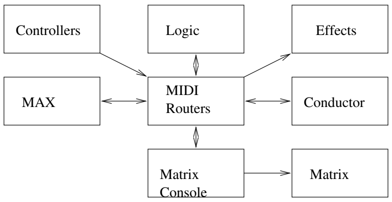

Figure 3: MIDI routing.

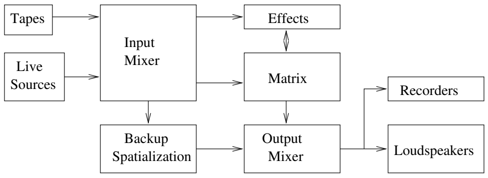

Figure 4: Audio routing.

A wide variety of MIDI controllers are used to control the system, including Buchla's Lightning[6], Steim's SensorLab[7] with a variety of custom controllers designed by or for STC, and David Rokeby's Very Nervous System[8] video motion analyser. Red Sound System's Voyager 1 Beat Xtractor[9] does an excellent job of tracking a changing tempo from an audio source and generating synchonized MIDI clock messages.

Almost all devices have a dedicated MIDI cable to the router: use of MIDI thru is kept to an absolute minimum. Routers are interconnected with multiple MIDI cables: contention for MIDI bandwidth over any physical connection is minimized. The heuristic used was to limit each MIDI cable to carry data for a single "continuous" controller, such as a fader pack or wand. Data for several, less communications-intensive devices, such as patchbays, could be multiplexed on a single wire using separate MIDI channels. Extensive use was made of the Opcode Studio Patches software, and its support for "virtual" devices and MIDI channel filtering and remapping.

4.4 Two configurations

Development has resulted in two distinct configurations: the "studio" system, capable of producing digital multi-track recordings of spatialized audio, and the more portable "club" system. Both systems can be used for live performance, and have similar audio routings (Figure 4).

The studio system is based on the LD-88 Supernova digital mixer by Level Control Systems, and is currently capable of rendering 3-D audio for up to 24 speakers at arbitrary locations. The club system uses from one to three Yamaha 02R digital mixers, which serve as both input mixer and matrix. The club system is limited to up to 14 speaker locations, due to the Yamaha 02R's support for only 8 buses and 6 auxiliary sends to external devices. The 02R was the only digital mixing console we found which supported a sufficient number of buses with external control of individual bus send levels.

The backup spatialization systems, for use in the event of a crash of the main matrix, consist of a custom Serge analog panning system and a pair of OmniSound SSP-100 Spatial Sound Processors.

4.5 Authoring

The MIDI sequencer is responsible for recording and playback of temporal control data, such as sound trajectories. There are two basic ways to create trajectory data: using MIDI controllers or using on-screen controls in the sequencer. When using MIDI controllers, the raw controller data is first converted to the 3-D position protocol (described earlier). This conversion takes place in MAX or in the sequencer itself. The 3-D position data can then be recorded as a separate "track" using the MIDI recording features of the sequencer. Using the on-screen controls built in Logic's graphical dataflow language, the 3-D position data can be generated and recorded in exactly the same way as any other MIDI data (Figure 5).

By manipulating the controller of choice while auditioning a timecoded tape containing the audio source material, the composer can time the movements of the sounds as desired. Upon playback, the motions will occur in synchronization with the audio material, as performed by the composer. Each track of the audio tape can have its own trajectory recorded by repeating the process as desired. The current system supports up to 16 individual audio tracks using a pair of synchronized Tascam DA-88 digital multitrack tape units.

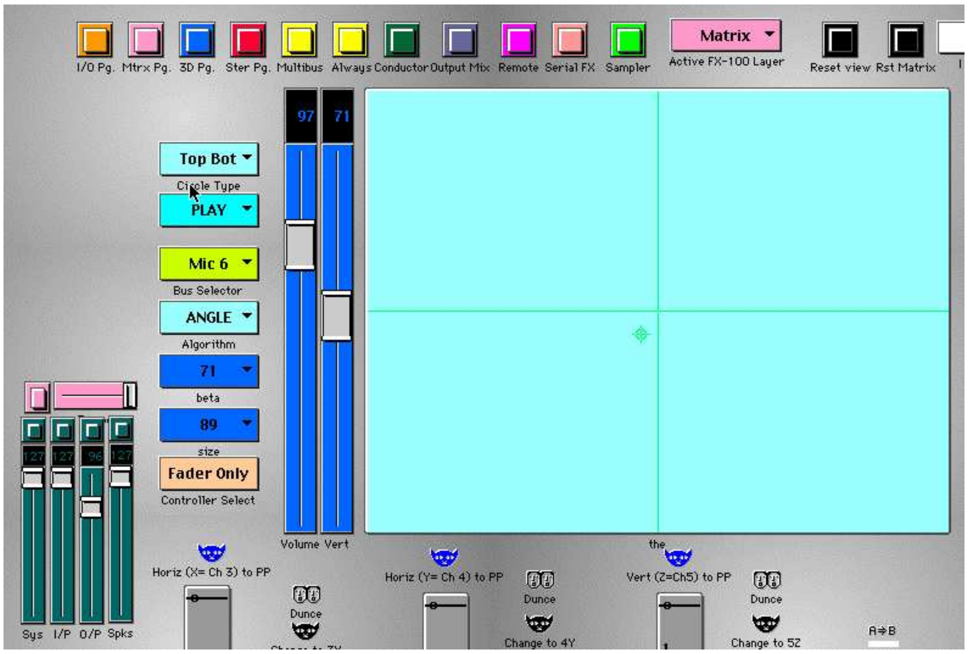

Figure 5: An example of an on-screen control surface created in Emagic Logic.

5 Evaluation

Several barriers to system reconfiguration were found during the recent implementation of the club system. Perhaps the most time consuming was the need to modify the matrix control protocol from one designed for a square matrix (with equal numbers of inputs and outputs) to accomodate a "narrow" matrix where inputs greatly outnumbered outputs. This portion of the protocol uses control change messages, which provide only $128 \times 16 = 2048$ controls. Mapping the 1600 matrix levels needed for the 40x40 matrix used in the studio system was awkward, and not extendible to the 108x14 matrix used by the club system. Specifically, the 40x40 studio system uses the following scheme: for inputs 1 through 16, channel = input, controller = output; for inputs 17 through 32, channel = (input-16), controller = (output+40); for inputs 33 through 40, channel = (input-32), controller = (output+80). The narrow matrix of the club system affords a much simpler mapping: controller = input, channel = output.

The smaller matrix width of the club system makes impractical the use of the matrix for audio patching between effects modules. The studio system provides support for user configuration of any serial routing of any 3 of the 8 main effects modules, by having the inputs and outputs of the main effects modules connected directly to the matrix. This feature had to be abandoned for the club system due to the dearth of effect sends. We are currently considering designs to allow multiple effects to flexibly share the limited number of sends.

Many difficulties encountered were due to the limited support for computation in Logic's dataflow language. Despite these limitations, Logic has proven to be extremely flexible and reliable. As development using the Logic sequencer developed, it became clear that much of the functionality involved converting MIDI controller streams into the spatialization control protocol. This reduced the reliance on the MAX software and improved the reliability of the system as a whole. MAX's role in the system has been relegated to effects control and handling of special purpose hardware, most notably Rokeby's VNS video processor. Recent development promises to eliminate the use of MAX altogether, as MIDI control stream routing is moved into Logic and as an alternative video controller (Grabbo[10]) for BeOS developed by one of the authors comes online.

Working within the limitations of the MIDI protocol is only somewhat mitigated by the use of parallel MIDI streams. Most MIDI devices will fail when a full-bandwidth MIDI stream is directed at them, although this is fortunately not the case with the mixers we've used. Most MIDI controllers will not emit a full-bandwidth MIDI stream, with the notable exception of Buchla's devices. The downsampling of MIDI controller streams is used at various places in the system: for example, MAX's "speedlim" object and Logic's "data reduction" option.

The studio system has been used in a number of live performances, in Los Angeles (1996), San Francisco (1996), and Miami (1998). The system performed flawlessly in all cases, without the need to resort to backup systems. All performances used a combination of recorded multi-track material and live sources spatialized in real-time. Reviews were uniformly positive [11,12] At the Los Angeles show, a multi-track tape arrived at the venue on the day of the performance. By assigning each track a position in 3-D, the piece, designed for a different loudspeaker configuration, was ready for playback in a few minutes. The first performance of the club system was as part of the closing ceremonies of the 1998 World's Fair in Lisbon.

6 Acknowledgements

Sam Ashley, Xopher Davidson, Scot Gresham-Lancaster, Ray Guillette, Keith Hildebrandt, Greg Lenczycki, David Sands, M. Stevens, and Richard Zvonar have all made contributions to earlier versions of the system. Many of their ideas are in the current system. This work would not be possible without the generous support of Asphodel Records and Mitzi Johnson.

References

[1] Malham D, Myatt A. 3-D Sound Spatialization using Ambisonic Techniques. Computer Music Journal 1995; 19(4):58-70

[2] Puckette, M and Zicarelli, D. An Interactive Graphic Programming Environment MAX. Opcode Systems, Inc., Palo Alto, 1991.

[3] Level Control Systems. LD-88 Supernova User's Manual, Sierra Madre, 1997.

[4] The Be Development Team. Be Developer's Guide. O'Reilly, Sebastopol, 1997.

[5] EMagic. Logic Audio User Manual, 1996.

[6] Rich R. Buchla Lightning MIDI Controller. Electronic Musician 1991; 7(10):102-108.

[7] STEIM Studio Report, ICMC 92.

[8] Cooper D. Very Nervous System. Wire Magazine 1995; 3(3):134-170.

[9] Red Sound Systems Ltd. www.redsound.com.

[10] Tebo Software. Grabbo. www.idiom.com/˜tebo/.

[11] Davis E. Global Ear. The Wire 1997; January.

[12] Ferguson S. Recombinant '96. Raygun 1997; March.