Composition and diffusion: space in sound in space

BARRY TRUAX

Schools of Communication and Contemporary Arts, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6
E-mail: truax@sfu.ca
URL: http://www.sfu.ca/~truax

1. INTRODUCTION

Composition and diffusion can be understood as two complementary and related processes: bringing sounds together, and spreading them out again in an organised fashion. In the Western tradition, these two processes are frequently carried out by different people at different times, each drawing on specialised knowledge. The electroacoustic tradition, even if much briefer, offers the possibility of the composer designing and implementing both aspects of the music, and interrelating them in highly specific ways. Computer control offers the greatest precision in dealing with the complexities of these processes, even though, at present, separate programs are usually required.

I am mainly referring to the practice of timbral composition, which may be thought of as shaping the space within the sound, that is, its perceived volume (Truax 1992). By this term I mean not merely the loudness of the sound, but rather its spectral and temporal shape, both of which contribute to its perceived magnitude and form. Diffusion, as the performance mode for these sounds, refers to the distribution of the (usually stereo) sound in a space through the use of a mixer and multiple loudspeakers. However, we can also understand the success of such a performance as a matching of the space within the sound with the space into which it is projected. This can be done even more effectively with multiple channel inputs where each soundtrack can be kept discrete and projected independently of all others.

At Simon Fraser University (SFU) we have been developing specific digital signal processing (DSP) techniques for each of these operations. The main techniques used for timbral composition are digital resonators, using variable-length delay lines with controllable feedback, and granulation of sampled sound used for time stretching, both of which allow the composer to shape the volume of the sound (Truax 1994). Recently both of these processes have been integrated into the same program (GSAMX). The diffusion project is a custom-designed multiple DSP box, the DM-8, designed by Harmonic Functions in collaboration with SFU, at the centre of which is a computer-controlled 8×8 matrix with which eight input streams may be simultaneously routed to any of eight output channels, either in fixed or dynamic trajectory patterns. A commercially available 16×16 matrix called the Audiobox has also been developed (for information, consult the links on my Web site).

These two aspects of the compositional process – timbral design and spatialisation – are usually dealt with separately, both by the composer and by researchers. However, certain common threads are beginning to emerge as we focus more closely on the micro level of sound design (Clarke 1996). The key factor appears to be the extent to which signals, or components of a signal, are uncorrelated, that is, the extent to which they have independent time behaviour (the technical measurement of this property is called autocorrelation).

Uncorrelated signals will increase the apparent volume of a sound provided there is a basis for perceptual fusion of these components into a single, possibly complex auditory image. In Truax (1992) and below I refer to these as 'unsynchronised components' based on empirical evidence with granular synthesis. Similarly, Kendall (1994) points out that when uncorrelated stereo signals are projected through loudspeakers, they tend to be perceived as separate sources, much as they would in the natural environment when two distinct, but similar sounds come from different directions. Correlated signals, on the other hand, produce an unstable 'phantom image' that is highly dependent on the listener's head position. When the listener is slightly off centre, the precedence effect determines that the closer or louder source dominates.

Thus, the presence of uncorrelated signals and their components appears to be a central determinant of the auditory system's interpretation of both the volume of a sound source and its spatial deployment. It should be noted that highly correlated signals are often the result of electroacoustic techniques and do not occur naturally, fixed waveforms and stereo panning being two common examples. Similarly, electroacoustic technology has often allowed audio parameters to be controlled independently, something that cannot occur naturally, the most common

Organised Sound 3(2): 141–6 © 1999 Cambridge University Press. Printed in the United Kingdom.

https://doi.org/10.1017/S1355771898002076 Published online by Cambridge University Press

example being a change in intensity level independent of spectrum, as practised in studio mixing. Therefore, the search for alternatives to these simplistic techniques may reflect a disillusionment with the artificiality of electroacoustic sound practice, and a search for greater ecological validity. That is, we need to test our models against the complexity of the natural environment and the perceptual strategies of the auditory system which presumably have developed to interpret it.

2. SHAPING THE SPACE INSIDE THE SOUND

The volume, or perceived magnitude, of a sound depends on its spectral richness, duration, and the presence of unsynchronised temporal components, such as those produced by the acoustic choral effect and reverberation. Electroacoustic techniques expand the range of methods by which the volume of a sound may be shaped. Granular time-stretching is perhaps the single most effective approach, as it contributes to all three of the variables just described. It prolongs the sound in time and overlays several unsynchronised streams of simultaneous grains derived from the source such that prominent spectral components are enhanced. It should be noted that delays of only a few milliseconds are sufficient to decorrelate the different grains streams and thus increase their sense of volume. Further, the grain streams are routed independently to one of the two channels available, with no use of panning. Therefore, both the grain streams and the individual channels are uncorrelated, thus linking timbral design to spatial diffusion.

In addition, my GSAMX software allows each grain stream to have its own pitch transposition, either downwards or upwards, according to a scheme where the untransposed pitch is the fourth harmonic in the scale of transpositions. That is, three downward harmonic pitches are available, plus four or more harmonics in each octave above the original pitch. However, processing the material through one or more resonators (using a waveguide or delay line) prior to granulation will also shape the spectrum of the sound quite strongly and bring out particular harmonic or formant regions.

The Karplus–Strong model of a recursive waveguide with filter has long been regarded as an efficient synthesis technique for plucked string sounds (Karplus and Strong 1983). The basic model for the waveguide uses a delay line of p samples which determines the resonant frequency of the string, a low-pass filter which simulates the energy loss caused by the reflection of the wave, and the feedback of the sample back into the delay line. The initial energy input is simulated by initialising the delay line with random values, that is, introducing a noise burst whose spectrum decays to a sine wave at a rate proportional to the length of the delay line. The model applies equally to a string fixed at both ends or a tube open at both ends, at least in terms of the resonant frequencies all being harmonics of the fundamental. If the sample is negated before being fed back into the delay line, the resulting change of phase models a tube closed at one end, which results in only the odd harmonics being resonant, and lowers the fundamental frequency by an octave, since the negation effectively doubles the length of the delay line. For the basic model, the fundamental resonance equals $\mathrm{SR}/(p + 1/2)$ where SR is the sampling rate and p is the length of the delay line.

However, since the technique models a resonating tube as well as a fixed string, it is equally suited for processing sampled sound. Because an ongoing signal activates the resonator, rather than an initial noise burst, a feedback gain factor must be used to prevent amplitude overflow and to control the amount of resonance in the resulting sound. The current realtime implementation offers a choice of delay line configurations (single, in parallel or series), plus the options of adding a comb filter (to add or subtract a delayed signal) and signal negation (which lowers the fundamental frequency by an octave and produces odd harmonics). Particularly interesting effects occur when the length of the Karplus–Strong delay and the comb filter delay are related by simple ratios. Each delay line has realtime control over its length, and hence its tuning, up to a maximum of 511 samples during direct sample playback and 1,023 in GSAMX. The user also controls the feedback level which can be finely adjusted to ride just below saturation, in combination with the input amplitude which can be lowered to facilitate higher feedback levels. The use of sample negation also makes it easier to control high feedback levels since the length of the feedback loop is essentially doubled.

The complex behaviour of these resonators, particularly when driven to their maximum feedback level (termed hyper-resonance) cannot be tracked by the ear at normal speed, compared to when such sounds are time-stretched and their internal variations become more evident. Although the process of resonating a sound may resemble equalisation, the presence of feedback results in a stronger phase-shifting of the various spectral components, thereby decorrelating them and creating an increased sense of volume. By time-stretching the sound, the resonator essentially expands to the dimension of a large room without a change in the resonant frequency. Here again we find a link between the perceived volume of the sound itself and the space in which it is heard, reverberation being perhaps the most common example of uncorrelated signals (in the form of

https://doi.org/10.1017/S1355771898002076 Published online by Cambridge University Press

reflections) being added to a sound to increase its volume.

In practice, the sound may be resonated first, using a chain of up to two or three resonators, then resampled and granulated; or else, one can introduce a single resonator directly into the processing chain during granulation, using a specific option in the GSAMX program. Such processing lengthens the decay of the resonance to an arbitrary duration, hence suggesting a very large space, while keeping the resonant frequencies intact. That is, resonant frequencies associated with relatively short tubes appear to emanate from spaces with much larger volumes, as in my work *Basilica* (1992). Vocal sounds subjected to this processing resemble 'overtone singing' in a reverberant cathedral, because the resonant frequencies are strong enough to be heard as pitches. The addition of simple harmonisation at the granulation stage, such as an octave lower, enriches the sound further and gives the impression of a choir (ensembles being another traditional method of increasing perceived volume through the presence of uncorrelated sources). The additional decorrelation of the harmonised version guarantees its being perceived as coming from a separate source, even when the interval is an octave.

The two-stage version of this processing (resonance, then time-stretching with or without harmonisation) was used in my electroacoustic music theatre work *Powers of Two*: *The Artist* (1995) (Truax 1996a), which is the second act of the opera *Powers of Two*. The sounds employed in subsequent acts have been created using the integrated approach where the resonance is added during the time-stretching process. In one particularly striking example, found in *Powers of Two: The Sibyl* (1997), natural sounds such as a recording of rain and thunder, and another of ocean waves, are hyper-resonated to the point where the original sounds are engulfed by a low resonant mass of sound pitched at 60 Hz (the North American electrical frequency). Then, as the scene progresses, the amount of feedback added to the process is gradually reduced until the original sound is once again audible. This effect underlines the tension in each scene between a character associated with the modern, technological world and one associated with traditional visionary insight.

3. SHAPING THE SOUND INSIDE A SPACE

Although conventional diffusion is remarkably effective with a stereo source, both the two-channel bottleneck, and the limitations of manual control and too little rehearsal time, are currently the weak links in the performance of electroacoustic music. Having eight discrete sources available, all independently controllable, is not only acoustically richer for tape

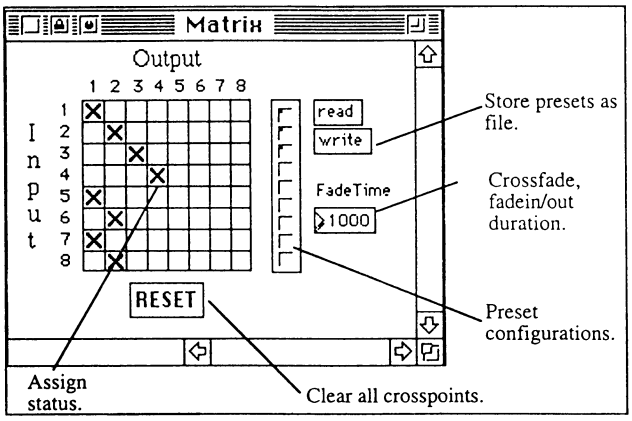


Figure 1. Matrix display/controller.

music (since detail is not lost through stereo mixing) but also challenging compositionally in order to integrate a spatial conception into the work. However, the same system can be used for live, or mixed live and tape performance, since nothing is assumed about the relation of the eight input signals.

The DM-8 system is essentially an 8×8 matrix (figure 1) which routes eight channels of input (for us, the Tascam DA-88) to eight channels of output, presumably going through a conventional amplifier and speaker configuration. The hardware is a custom-designed box, external to the host Macintosh, equipped with 4 Motorola 56001 chips and a 68000 controller, communicating via MIDI system exclusive messages to the graphic front end. The software for user control is a Max application, written by Chris Rolfe, which can be used either in a live performance mode with mouse-triggered events, or else as a pre-programmed score synced with the MIDI timecode on the tape. Presets and an editable mixing score allow each of the input tracks to have its amplitude controlled. These mixing levels can be graphically entered, or tracked from the user's control of virtual potentiometers in real time. These recorded levels are analysed and compressed by the program for an optimum data representation and can later be edited by the user.

The 8×8 matrix allows manual input/output connections to be made via mouse clicks (i.e. speakers are turned on and off), preset patterns of which can be stored and implemented with variable fade times. The cross fade from one configuration, say a stereo reduction, to another, for example a multichannel distribution over 5–10 s, is a typical operation that would be difficult to achieve manually but is aurally very attractive.

A set of 'players' extend the matrix control to either 'static' speaker lists, or to 'dynamic' trajectories (figures 2 and 3). Unlike the matrix operation, they automate both the turning off of outgoing channels as well as the turning on of new channels. The

https://doi.org/10.1017/S1355771898002076 Published online by Cambridge University Press

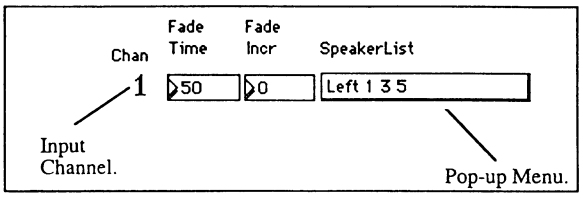


Figure 2. Static assign.

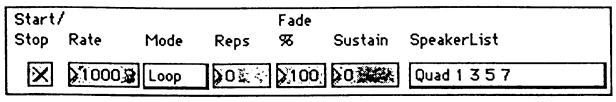


Figure 3. Dynamic assign.

dynamic assignments generate a series of cross-fades, moving an input from speaker to speaker in what we call a 'trajectory' at a specific rate with adjustable fade patterns. Predefined speaker patterns can be looped, cycled (forward and reverse), or randomly assigned. Since eight such patterns can be simultaneously running, very complex movements can be easily generated. All of the player parameters transfer directly to the score method of control, hence a particular trajectory configuration can be tested in real time, then copied into the score with its precise point of implementation.

Of interest to electroacoustic composers is the ease with which a given set of speakers can be substituted for another when a new performance configuration is encountered, or when a mixdown is needed. A speaker list is defined once and labelled (e.g. left, circle, etc.) with nothing assumed about where those speakers are located. To change to a different speaker configuration, only the list needs to be edited, not each instance of its use. The label also assists the composer in dealing with particular spatial configurations independently of the often confusing lists of speaker numbers.

The nature of cross-fades between speakers is a particularly tricky subject, and the software assists the user with both graphic displays of the levels involved and realtime aural tests of the effect (figure 4). Cross-fade percentage is a key variable, allowing a continuum of effects from jumping between channels to completely smooth transitions to be achieved. A 'sustain delay' parameter delays the fadeout of the previous speakers in a dynamic sequence to create a more 'polyphonic' effect (analogous to the vapour trail behind a jet). Finally, the 'fade increment' is a simple method to generate the cascaded entry of a speaker list, similar to the way one might bring in a set of speakers incrementally during conventional manual diffusion to create another polyphonic effect. It should be noted that the use of cross-fading between channels is the only instance in the system of any type of panning technique, as criticised earlier. In practice, the 'pans' are brief, dynamic and usually used between speakers that are in close proximity. Under these restricted conditions where no 'phantom image' is being created, the use of amplitude correlation seems justifiable.

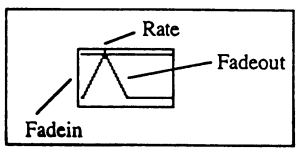


Figure 4. Fade display.

Although the system is designed for controlling eight source channels, other uses are possible. For instance, a stereo source could be duplicated up to four times at the input of the matrix, and four pairs of distinct trajectories or speaker assigns defined. The composer could then use the mixing score or manually controlled input levels to cross-fade between the different spatial treatments. Alternatively, the entire matrix could be considered to be an effects send and return system for studio work with, for instance, two 'dry' channels and six channels of processing being mixed together.

The DM-8 has been used in performance at the 1995 International Computer Music Conference in Banff and in 1996 at various Vancouver New Music electroacoustic concerts, and is currently available for use in the Sonic Research Studio at SFU. At present it is most practical to record a spatialised version of the output onto another eight-track tape for distribution to other centres. Although an extended (16×16) commercial version has been developed, the existing hardware and software configuration is already extremely useful for electroacoustic diffusion. The software can also be extended by programmers wishing to add new features or more complex lower-level control patterns.

As mentioned earlier, multiple loudspeaker diffusion using a stereo source is remarkably effective, particularly when the person controlling it makes creative use of the psychoacoustic precedence effect; that is, by increasing the level of sound in a particular speaker, that location becomes the apparent source of the sound, even though the same sound may be present at lower levels in other speakers. The illusion works best when the change in level is correlated with particular sounds, hence the necessity for the performer to know the work quite intimately. The prominence being given a particular sound in a particular speaker makes the listener believe that it is coming from that direction. The limitation, however, is that it is generally possible to 'position' only one such sound at a time, and only those sounds which do not overlap others which it may also be desirable to highlight.

https://doi.org/10.1017/S1355771898002076 Published online by Cambridge University Press

Instead of regarding stereo diffusion and discrete multiple-channel systems as opposing techniques – which has unfortunately characterised some recent discussions – I would like to suggest that the multiple-channel system can be understood as an extension of stereo practice. Eight-channel tape, for instance, can be thought of as four contrapuntal stereo layers. The key concept, though, remains the use of multiple speakers as point sources, each of which can be fed an independent (i.e. uncorrelated) signal. The model on which this technique is based, it needs to be remembered, is that of the natural environment where individual sources emanate from discrete spatial locations, even when the component sources of a sound are linked but spatially separate (e.g. a stream, waves, waterfall, tree branches in the wind, etc.). However, to extend diffusion technique to multiple channels, computer control is required since manual control with two hands is limited in the two-channel model. Smooth trajectories are also difficult to control with stereo, and those in contrary motion are physically impossible to achieve with a conventional mixer. With computer control, though, the human performer is still a necessity as the resultant sound in a given space needs to be optimised by someone listening and making adjustments for the complexities of room acoustics and speaker characteristics, neither of which can be completely anticipated.

4. RECENT COMPOSITIONAL APPLICATIONS

As mentioned above, the eight-channel tape component of my electroacoustic opera *Powers of Two* plus the tape interludes in the opera *Sequence of Earlier Heaven* (1998) and *Sequence of Later Heaven* (1993), were realised utilising the techniques described here, both for the design of the component sounds and their static and dynamic distribution in eight channels. Two other recent compositions for solo performer and stereo tape also illustrate the timbral work, namely *Wings of Fire* (1996) for female cellist and tape, based on a poem by British Columbia poet Joy Kirstin, and *Androgyne, Mon Amour* (1997) for male double bassist and tape, based on poems by Tennessee Williams. In both works, the source material is derived from a reading of the poems as well as sounds recorded from the live instrument, all of which are processed with granulation and/or the use of resonators simulating the open strings of the instrument. When the voice is processed in this way on tape, it is given some of the character of the instrument, and in each piece the love poetry appears to be addressed to the instrument as the person's lover. In other words, the spoken voice on tape appears to be resonated through the instrument being played, hence symbolising their amorous union.

Other sounds recorded from the cello and bass are also used to excite the resonators. These include bowing on the bridge, natural and artificial harmonics, *col legno* attacks, snap pizzicato and various kinds of body percussion sounds. By raising the feedback level of the resonators (tuned to the open strings), a noisy sound such as bowing on the bridge slowly changes from resembling breathing to regular bowing on the strings, once again highlighting the intimate relation between the performer and the instrument. Interestingly enough, when the length of the delay line is shortened to produce a very high pitch, the noise component once again becomes dominant, as at the end of the opening section of *Wings of Fire*. In *Androgyne, Mon Amour*, the tuning of the resonators (independently controlled on each channel) changes more frequently during the reading of the text, suggesting a kind of harmonic accompaniment performed by the instrument. The live instrument, which is frequently played in a number of unconventional postures, sometimes mimics this accompaniment, or creates a counterpoint to it.

The multiple channel approach, not surprisingly, is particularly well suited to the performance of soundscape compositions (Truax 1996b). Although spatially distributed sound sources are an important aspect of some musical compositions, they are an *inevitable* part of all soundscapes, and spatial perception is integral to making sense of them. Bregman (1990) and others have termed this 'auditory scene analysis', though unfortunately environmental examples are seldom studied by these researchers. However, it is clear from their psychoacoustic models that the auditory system is particularly adept at grouping correlated spectral components into unified images of sound sources, each of which can be distinguished from various others (at least in what R. M. Schafer calls a 'hi-fi environment') because of their spatial placement and other independent characteristics.

In my eight-channel tape piece, *Pendlerdrøm* (1997) (The Commuter's Dream), recordings made in and around the Copenhagen train station are layered in four simultaneous stereo pairs of tracks, each channel of which is fed to its own speaker during the realistic portions of the piece. Given that a train station presents a complex soundscape of multiple, somewhat unpredictable sources, the reproduced soundscape appears very plausible, if somewhat busier than at the time of the original recording. As with the 'cocktail party effect', the listener is able to focus on any particular source of momentary interest and ignore all others, or else scan the entire scene created by the surrounding speakers. In fact, the optimal stereo imaging of the original source recordings

https://doi.org/10.1017/S1355771898002076 Published online by Cambridge University Press

assists rather than detracts from this illusion, and is probably more successful than eight monophonic source sounds.

At two points in the piece, transformations of selected sounds in the environment are gradually introduced to suggest that the listener, as 'commuter', lapses into a daydream or imaginary world where the musicality of various sound objects in the environment is explored. During these sections, the eight channels are used to create circular spatial trajectories around the listener, in contrary directions, symbolising the fluidity of the dream state, and contrasting with the static placement of the untransformed material heard earlier. The first of these transformed sections grows out of the repetitive sound of a train passing the listener in a dramatic lateral motion, supported by a loop of the wheel percussion, both of which are resonated. The arrival of the local commuter train draws the listener out of this reverie, and the diffusion pattern returns to the static, discrete channelling mode. Once apparently ensconced on the commuter train which pulls away from the station, the listener enters another dream-like transformed sequence in which fragments from the previous sections (e.g. public address announcements, whistles, brakes, and door slams) randomly appear in loops that rotate spatially over the drone of the engine, all of which are granulated and resonated to create a sense of larger-than-life unreality. Triggered by the announcement of the next station, the listener is 'wakened' by a magnified door slam, the original of which was heard earlier, and the scene reverts to the apparent realism of the recordist descending from the train and leaving the station by a wooden stairway. Although originally designed for a stereo CD, the eight-channel version is much more successful in portraying both the realism and the imaginary world of this scenario.

At present, the processes of shaping the 'volume' of the sound, its internal space, and distributing the sound via multiple loudspeakers into the external performance space, occur in two different design stages, much as traditional studio composition and live diffusion have been carried out. The compositional challenge is to create significant relationships between the two processes. However, if we continue to use similar DSP technology for both, it may well become feasible in future to integrate them into a single algorithm in which the individual components that create the volume of the sound are given spatial placement and definition within the performance environment. That uncorrelated signals are a key element in both processes should facilitate this integration. Sound and space would become inextricably linked, and composition could then truly be regarded as the acoustic design of space in every sense of the term.

* * *

Note: An earlier version of this paper was presented at the 1997 meeting of the Academy of Electroacoustic Music, Bourges, France.

REFERENCES

Bregman, A. 1990. *Auditory Scene Analysis*. Cambridge, MA: MIT Press.

Clarke, M. 1996. Composing at the intersection of time and frequency. *Organised Sound* **1**(2): 107–17.

Karplus, K. and Strong, A. 1983. Digital synthesis of plucked string and drum timbres. *Computer Music Journal* **7**(2): 43–55.

Kendall, G. 1994. The effects of multi-channel signal decorrelation in audio reproduction. *Proc. Int. Computer Music Conf.*, pp. 319–26. Aarhus, Denmark.

Truax, B. 1992. Musical creativity and complexity at the threshold of the 21st century. *Interface* **21**(1): 29–42.

Truax, B. 1994. Discovering inner complexity: time-shifting and transposition with a real-time granulation technique. *Computer Music Journal* **18**(2): 38–48 (sound sheet examples in **18**(1)).

Truax, B. 1996a. Sounds and sources in *Powers of Two:* towards a contemporary myth. *Organised Sound* **1**(1): 13–21.

Truax, B. 1996b. Soundscape, acoustic communication and environmental sound composition. *Contemporary Music Review* **15**(1): 49–65.

DISCOGRAPHY

Song of Songs. 1994. Cambridge Street Records, CSR-CD 9401, 4,346 Cambridge Street, Burnaby, B.C., Canada V5C 1H4 (includes *Sequence of Later Heaven*).

Inside. 1996. Cambridge Street Records, CSR-CD 9601, 4,346 Cambridge Street, Burnaby, B.C., Canada V5C 1H4 (includes *Powers of Two: The Artist*).

Pendler. 1997. Skraep double CD, Copenhagen (includes *Pendlerdrøm*).

https://doi.org/10.1017/S1355771898002076 Published online by Cambridge University Press