Focus

Barry Vercoe is Associate Professor of Music and Technology at M.I.T. and Director of the M.I.T. Experimental Music Studio. Trained as a conductor and composer, he developed the audio signal processing languages Music 360 and Music 11, now used worldwide in computer music studios. His composition Synapse for Viola and Computer *is recorded on the CRI label (SD393). He recently received a Guggenheim award to pursue further work in computer music composition.*

New Dimensions in Computer Music

Barry Vercoe

Electro-pop—the synthesized musical product of the newest British performing groups—is dominating European popular music and is expected soon to make a big impact in the United States. Major exposure to the music and style of charismatic performers (one of whom, like Texas Instruments, has a best-selling hit called "Speak & Spell") is going to focus a new wave of public attention on the interaction of technology and musical expression.

Although the output of the popular recording groups may be described as a new music, these performers are in fact exploiting methods that are five to ten years old. The ability of current digital technology to produce aesthetically expressive, musically interesting sound goes far beyond the range presently used by these artists and their equipment. Research musicians working with digital methods are producing a rich variety of expression that will be increasingly available to composers and performers as a viable means of responding to creative, expressive ideas.

Earlier views that all electronically generated music should sound like traditional acoustic instruments have now given way to the belief that the electronic medium has a much larger potential. The invitation to move a speaker cone in any manner that is effective opens a vast, largely untapped territory of sound production in which digital signal processing becomes the primary vehicle of exploration. Composers who see this as an expressive opportunity have formed major centers for acoustic research and experiment with the sole purpose of exploring this new resource. The challenge is attracting many with long-established roots in traditional music, most notably Pierre Boulez, who leads an unusual tripartite life conducting premier symphony orchestras, directing the French government's research program in computer music, and composing concert music for performance by real-time digital audio processors.

The commerce in synthesized musical instruments is growing faster than the rate at which research musicians can make their findings known. Given the technology curve, high-technology musical instruments will likely become commonplace. Although this proliferation will be unstoppable, the artistic merit of the products will depend on the level of audio processing and interactive control that can be attained before the falling cost of production makes the devices irresistible. In the consumer market, for example, Casio has sold more than 300,000 of its $60 VL-Tone keyboard instruments since they were introduced in March 1981. With the emphasis on technology and with experience in packaging and merchandising other consumer electronic products, the consumer electronics giants will inevitably displace many original manufacturers of electronic musical instruments. The exception will be the market for high-end musical instruments, which will offer opportunities to smaller, specialized companies that can stay competitive by responding to the wishes of the professional or sophisticated performer.

The Loudspeaker as Instrument

Today we spend possibly half of our waking hours bathed in music, most of it coming from loudspeakers. The music derives from two sources: the sound of acoustic instruments, recorded and stored as electrical signals, and entirely synthetic sound, sometimes acoustically recorded but more often passed directly to the storage device in electrical form. In both cases, storage affords an opportunity for additional enhancement by way of editing, reverberation, and other signal processing.

The processing of acoustic instrument sounds has greatly affected the expectations and standards of the listening audience. In classical recordings we now take for granted the numerous takes and the splicing of right notes for wrong ones. More insidiously, however, the art of recording and microphone placement puts listeners in a seat they cannot buy at a live concert, successfully mixing the close-up sound of resin on the violin string with reverberant effects found elsewhere in the concert hall. Canadian pianist Glenn Gould does not give live concerts anymore but relies solely on the loudspeaker to represent his artistry. What the speaker produces becomes the new standard against which the live Gould piano would rather not be compared.

The same loudspeaker coupled with a signal processor can be the emitter of entirely synthetic sound. As pure process, it can have an independent standard. An important issue in the design and manufacture of performance systems is whether digital music processing should imitate the sounds of traditional acoustic instruments or make new sound appropriate to a new medium. The audio objectives can be quite different, although the familiar attributes of presence, expressiveness, and spatial effect will remain important. Many systems designers have chosen to pursue only traditional sound. Just as the Hammond organ vainly tried to drive a speaker in imitation of acoustic instruments, the latest results continue to prove that, even with the advent of modern signal processing, the pursuit of acoustic timbres remains difficult.

The easiest sounds for a fully controllable technology to duplicate are those that are basically linear. The sounds of a large pipe organ, for instance, have a near-perfect steady state, and the only reason early electronic substitutes were generally unconvincing was that they had no memory. This situation took a dramatic turn around 1970 when the Allen Organ Company stored one cycle of real organ sound in a digital shift register, then simply pumped it out at the desired frequency. A more difficult problem was posed by the nonlinearities normally heard at the onset or chiff attack. The observation that an organ pipe took roughly seven cycles to settle (irrespective of frequency) suggested a simple solution: the stored waveform of 8-bit samples could be slid in gradually, using a binary left shift after each of the first seven cycles. This poor man's exponential—cranking up by a factor of 2 at zero-crossings—did not duplicate the physical phenomenon, but it did duplicate rather well the psychoacoustical effect. Unfortunately, almost every other nonlinear characteristic of our musical instrument heritage has proved to be much harder to emulate.

Nonlinearities are indeed what make the sounds of music interesting, but it may be a tactical error to believe that the nonlinearities found in natural acoustic instruments are the only ones to pursue. It is a common laboratory experience, for instance, to find, after spending much time coaxing familiar-sounding timbres out of some simulated instruments, that in performance of actual music a different set of acoustic priorities mysteriously seems to pertain. The instruments then need to be modified to better delineate the more important aspects of the music. Apparently the time-varying nonlinearities that are not captured in such modeling are often the more important ones, so the designer or composer has to be creative within the medium itself. This suggests that it is the medium that has the untapped communicative power and that the industry should exploit the idioms of the new, rather than misrepresent the subtleties of the old.

A more creative direction has been taken in the so-called serious concerts of new music and in the organizations that have been developed for their support. In Paris, the Institute for Research and Coordination of Acoustics and Music (IRCAM) has been established as part of the new Georges Pompidou Center for Contemporary Art to spearhead the assimilation of scientific method into music research and concert production. Directed by Pierre Boulez, formerly music director of the New York Philharmonic, IRCAM has quickly assumed a leading role in the development of new methods of audio synthesis and in the design of audio processors for real-time modification of instrumental sounds onstage during concerts. In its collaborative approach to research, IRCAM does not expect the scientist to compose but rather to understand what contemporary music is doing and to orient his or her imagination along those lines. There is a lead here for the music industry to follow.

Audio Processing Strategies

There are three general approaches in applying digital processing to music: algorithmic methods, semi-automatic analysis/synthesis, and perceptually based techniques.

Algorithmic methods cover mainly nonlinear processes that can produce spectrally complex results from little input. These include the use of trigonometric summation formulas (to get the buzzy effect of pulse trains with variable band-limiting) and frequency modulation (FM), a technique for inducing harmonically related sidebands whose relative strengths can be dynamically varied. John Chowning of Stanford University, who pioneered the use of FM for instrument simulation around 1970 (many commercial synthesizers today draw on his work), has recently extended that application to synthesis of the singing voice. Chowning's new composition *Phone*, which premiered in Paris in 1981, demonstrates that the new method can provide realism even for notes of extreme pitch. The procedure, although computationally efficient, is difficult to control and extremely painstaking.

Another algorithmic method that produces rich timbres is digital waveshaping, in which a waveform is redefined as it is passed through a lookup table. This technique of nonlinear distortion with artistic intent has been explored independently by Marc LeBrun of Stanford University and Daniel Arfib of the University of Marseille. Arfib has completed a new work, *Le Souffle du Doux*, that demonstrates how effective it can be. Although the technique is simple, digital waveshaping produces colorful effects and is well suited to random-access-memory (RAM) technology; it should appear in a commercial instrument quite soon.

Semi-automatic analysis/synthesis methods extract information from natural sounds to enrich synthetic ones. The simplest example is the additive resynthesis of complex tones. This method, which is generally preceded by phase-vocoder analysis, is currently handicapped by the weight of its control file of amplitude and frequency data on the component partials. (John Strawn and Gérard Charbonneau of Stanford are trying to reduce data levels without losing fidelity by determining the perceptually relevant elements of each component.) Although additive synthesis reproduces the original sounds with disquieting accuracy, it affords little direct insight into the synthesis of previously unanalyzed nuances. One virtue of the technique is its computational structure, which would lend itself well to the parallelism available in VLSI.

Linear predictive coding, a semi-automatic technique that derives from research on speech synthesis, separates spectral data from pitch data. With it, a composer can stretch or squeeze the timing of spectral changes without changing the pitch contour; alternatively, a sequence of spectral changes taken from a recorded voice or instrument can be given a whole new pitch line. The problem has been to maintain the audio fidelity and realism expected from a real instrumental performance. Kenneth Steiglitz and Paul Lansky at Princeton University have recently developed a method that can produce sounds belonging to distinct families of instruments (strings, brass) and can also differentiate between individual members of a family (violin, viola, cello, bass). The procedure involves frequency warping of the synthesis filters to change the effective timbre of the instrument (and thereby its psychoacoustic size). Steiglitz's recomputation of filter coefficients is notably fast. Lansky has recently premiered *Pine Ridge*, a new work for a simulated string ensemble that attests to the utility of this method.

Perceptually based strategies attempt to establish low-dimensional control over perceptually salient aspects of sound that cannot be found with routine extrapolations of simple physics. The phenomenon of "spectral fusion," for instance (in which two distinct notes, given correlated aspects such as a common vibrato, will merge into a single percept), suggests that complex music is subject to even more complex perceptual laws. David Wessel of IRCAM is doing qualitative experiments with synthesis in an attempt to understand and control such broad perceptual aspects as spectral energy distribution, roughness, and gain. In a related step, New England Digital's Synclavier II synthesizer allows the performer to "imagine" a target spectrum and then approach its synthesis by successive approximation (or partial timbres, in their unfortunate terminology). Reportedly, an audio processing add-on is forthcoming in which the technique has been automated, allowing the machine to pursue by itself any timbre that it has heard.

Tracy Peterson of the New York Institute of Technology is pursuing perceptually based synthesis using a transform based on the ear's critical band. By viewing target sounds through physical models such as the cochlea he hopes to find a representation of musical timbres that incorporates the auditory system's natural sensitivity. Campbell Searle of M.I.T. also uses a critical band model to represent natural aural filtering, then rotates the data so as to concentrate the variance in a small number of components. This simplified sound model should lead to simpler timbre control.

Another active area of applied signal processing is digital reverberation. In computer music this is not just a recording enhancement; composers like to move sounds around in space, and the search for a realistic simulation of the entire reverberant effect has gone on for some time. The first attempts followed the work of Manfred Schroeder at Bell Laboratories and were essentially algorithmic comb filtering processes. In subsequent refinements, Chowning adapted the Schroeder model to four channels of uncorrelated sound and has recently managed to control perceptual distance independently of loudness. J. A. Moorer of IRCAM has incorporated the lowpass-filtering effect of air absorption into Schroeder-like networks. John Stautner of M.I.T. has generalized these networks into matrix-like arrays and shown that both early reflection patterns and running reverberation can be modeled with a generalized n-channel sampling of a simulated reverberant space.

Systems for Research and Experiment

Software

Growing interest in computer music composition and performance has led to the development of both hardware and software systems for efficient investigation of their potential. The first program that could support serious research into the digital processing of music was Music 4, developed in 1963–64 by Max Mathews at Bell Laboratories. Music 4 established a modular principle for the description of audio processing networks that has allowed composers and researchers to communicate in common terms. The many descendants of Music 4 provide local system users with an intuitive audio descriptor language (at once both a signal processing language and a programming language) while staying very close to the hardware of a particular machine for the sake of speed.

The major software systems currently in use are Music 360, a powerful variant of Music 4 that runs on large IBM systems; Music 10, designed for the DEC PDP-10; Music 11, which runs on any DEC PDP-11; Music 4BF, a FORTRAN version designed for portability; and Mathews's Music 5, which includes block processing of data among its many extensions. Music 11 introduced *control signal* processing (slow signals such as vibratos are calculated at less expensive data rates). A C-language Music 11 is being developed at M.I.T. for the VAX-11 computer (and any microprocessor that runs the programming language C), and a C-language variant of Music 5 is already operating at the University of California at San Diego. Music 5, Music 360, and Music 11 are the most widely distributed, the first two being appropriate for research centers with centralized computing. Music 11, distributed by Digital Music Systems, suits a different pocketbook and has induced several music departments such as those at Eastman-Rochester and Brooklyn College to buy their own PDP-11 computers exclusively for music composition and research.

These software systems are a by-product of collaborative research at a small number of centers. Both Princeton University and Stanford University have been active since the mid-1960s. The Princeton activity, largely IBM processing with off-line conversion, saw the early and fruitful collaboration of composer Godfrey Winham and signal processing expert Kenneth Steiglitz, who immediately imbued the Music 4 program with comprehensive digital filtering capabilities. At Stanford, John Chowning and the artificial intelligence community, using DEC PDP-10 machines, pioneered the use of frequency modulation in the synthesis of complex spectra. Newer facilities at M.I.T. (PDP-11s) and IRCAM (PDP-10 and PDP-11s) have brought engineering and signal processing expertise into environments created exclusively for music production. The research conducted at such facilities is made available at annual conferences on computer music and through the pages of the *Computer Music Journal*, a quarterly published by the M.I.T. Press. All of these facilities, including a recent center at San Diego (DEC VAX-11 and PDP-11), have developed system configurations that accommodate the computational demands of high-fidelity signal processing along with high-quality on-line digital-to-analog (D/A) conversion. Each has also felt the limitation of non-real-time research and has been designing and building devices for audio processing at real-time rates.

Hardware

A hardware audio processing system that is sufficiently general for all audio processing research while fast enough to operate in real time is still largely a dream. Nonetheless, the major research centers have developed interesting real-time devices that have generated most of the ideas embodied in the simpler street models now being marketed. The first large research synthesizer was built in the mid-1970s by Systems Concepts for the Stanford Center for Computer Research in Music and Acoustics. It is a 20-bit machine with time-multiplexed digital audio generators and modifiers, all controlled by command words from its host processor, a PDP-10. The system is now fully operational, but getting to that point was difficult and expensive. The development cost to Systems Concepts exceeded the single-unit price of $80,000 by a large factor. Only one machine has ever been built.

By contrast, the audio processors constructed in Paris at IRCAM were planned as a developmental series. There have been four machines to date (the 4A, 4B, 4C, and 4X), each basically a 16-bit processor controlled by a PDP-11 host. The earliest

Computer Compositions Reviewed

Nineteen different items can turn a concert into a spare-parts inventory, and a rather jumbled one at that, if you're not taking notes. . . . It being the Boston Musica Viva, though, expert, confident performances were the norm, and in Peter Child's Ensemblance *a pleasant find turned up.*

Child has put his mastery of the hardware (the M.I.T. Experimental Music Studio's Music 11) to resourceful, artistic purposes. Sly similarities between "normal" instrumental sounds and computer-processed ones tease the ears, in the manner of acoustical puns; you also get the sense of a spacious mental canvas being excitedly filled up with discoveries and insights and speculations about mixtures of sonority, but with a difference—a keen sense of harmonic color and rhetorical device. One was always fascinated and surprised with what the piano-wind-string-percussion ensemble was producing, and with what might issue from the loudspeakers—in parallel, amplification, or contradiction. Ensemblance *ended with a chime-and-piano bell-tolling aura that evoked (and not a bad thing) the ending of Stravinsky's* Les Noces.

—Richard Buell
Boston Globe, March 23, 1982

If there is one international musical figure who can truly be described as protean, it is Pierre Boulez. The former enfant terrible *of French composers, who combines a brilliant mathematical mind with an expert musical ear, Boulez has been the chief theoretician of the postwar serialist movement. . . . Boulez for the first time in his career has turned to modern computer technology to produce his newest work* Répons. . . .

It is the most impressive piece yet to emerge from the hitherto uneasy marriage of music and technology. A formidable technical achievement, it is also a work that makes a direct appeal to the emotions, the sign of a masterpiece in any era.

Répons *is scored for three groups, arranged in a large rectangle. An instrumental ensemble of 24 musicians sits on a raised platform in the center, facing the conductor. Stationed symmetrically around the room are six soloists, also on platforms, playing two pianos, electric organ, harp, cimbalom, vibraphone and xylophone, with each instrument wired for sound. A half-dozen technicians operate a bank of machines on ground level behind the conductor. The most important is the advanced 4X computer developed at IRCAM that can alter and transform live musical sounds with a speed that allows it to function as effectively as a new instrument itself.*

The 18-minute work falls into easily understandable sections, each based on the classic principle of tension and release. The first section is for the instrumental ensemble only, unaided by electronics. The tension is created by rapid, repeated-note figurations and massed sonorities. The release, such as it is, comes from a series of eerie tremolos and trills reminiscent of the doom-laden flute flutterings in Strauss's opera Salome. *The soloists enter with a computer-assisted arpeggio, vibrating and echoing over the six large loudspeakers that are stationed around the hall. Then the soloists and the ensemble interact, responding to each other in the manner of Renaissance polyphony.*

As the piece develops, a furioso *section for the ensemble is followed by electronic responses from the soloists until the entire orchestra begins to fragment, a violin jutting out here, a trombone blasting there.* Répons *gradually increases in rhythmic complexity as held notes in the brass arch over the busily insistent sound beneath. The impression is of the turning of a gigantic wheel in space. The piece ends quietly on a stationary but disquieted chord; rest is achieved at last, but not peace. . . .*

It is already clearly a major work by a composer who is still boldly extending music's horizons.

—Michael Walsh
Time, December 28, 1981

processor, a bank of time-multiplexed oscillators relying on the host for all amplitude-envelope and pitch data, kept the host seriously overloaded. It was clear that future devices had to do more thinking for themselves. The 4B model calculated its own envelopes, but at a lower resolution in time than it used for audio signals. This resembled the Music 11 control signal concept, but it was hard-wired in such a way that the user could not modify the resolution when it was inadequate. This proved to be a problem, which was corrected with the 4C model. The success of the 4C has prompted the software staff to develop higher-level patching and control languages. (The 4CED program of Curtis Abbott is described in the *Computer Music Journal,* vol. 5, no. 1.)

The 4X machine was designed for greater flexibility and increased processing power. It has a parallel architecture with eight micro-controlled audio signal processing cards. Each card can simulate any of the earlier machines by the addition of a special read-only memory (ROM). One ROM turns any card into a reverberation processor, while another can make it into a high-speed outboard multiplier. A separate MC68000-based system handles gestural inputs such as keyboards and potentiometers, while another manages the 16 channels of audio, both in and out. The system can perform either real-time synthesis or performance-oriented analysis of sound. It can do real-time linear predictive analysis and resynthesis, automatic event detection and pitch detection, plus digital recording, editing, recomposition, and playback. The system is now musically productive and the hardware sufficiently stable to allow committed software work. In its first public appearance, the 4X did real-time sound analysis and modification in the premier performance of Pierre Boulez's *Respons*. The IRCAM audio processing machines have been largely the design work of Giuseppe diGiugno. All have been proprietary inhouse developments, strongly controlled by the terms of IRCAM's French government support. It has recently been reported, however, that IRCAM may be interested in a commercial manufacturer.

Alternative Processors

Many issues in computer music research can be explored with off-the-shelf audio signal processors designed specifically for real-time research and sound production. These have the virtue of being more generally available than the systems discussed above. The DMX-1000 from Digital Music Systems, for example, is a 16-bit programmable processor with twice the speed of a PDP-11/70 at a fraction of the cost. Its 24-oscillator real-time capacity is built from low-level primitives that can easily be reconfigured for different needs; field applications have ranged from seismic data processing to psychoacoustic research. Sales have been disappointing, however, perhaps because such broad marketing is difficult for a small company.

Some of the more pressing issues arise in the context of manufacturing a performance instrument. Hal Alles of Bell Laboratories, who was a collaborator on the IRCAM 4B, designed a 32-voice oscillator card with independent amplitude and frequency controls for each voice. The design called for only about 100 inexpensive ICs and did not need a multiplier chip. When one company decided to produce a medium-priced synthesizer (the Synergy) based on the Alles design, it soon found itself developing a comprehensive audio research device before it could begin to make design decisions on the intended product. This research device took such an investment that it was marketed as a separate product (the Crumar General Development System) with a considerably higher price tag.

Since current methods of music generation lend themselves to processing in arrays (as is done in software in Music 5 and Music 11), it is possible to implement some useful real-time synthesis on standard array processor hardware by restricting input/output to control parameters in, audio samples out. Researchers at the M.I.T. Experimental Music Studio have programmed an Analogic AP-400 for real-time work. The 24-bit fixed-point device can manage several voices of average complexity, along with artificial reverberation. Array processors can also be a potent source of floating-point computation, an important aspect of audio processing involving wide dynamic range. A medium processor from Floating Point Systems, for example, can be programmed to sustain about 2 million multiply-adds per second—not a huge audio data rate, but one that allows more than adequate audio quality. In contrast, the faster special-purpose processors that have dominated real-time audio processing achievements to date have all been fixed-point machines, a situation typical of programmable signal processors generally (see *Trends & Perspectives in Signal Processing*, vol. 1, no. 3, p. 4). Recent advances in connecting several microprocessors suggest that a tightly coupled microprocessor system might offer an alternative audio processor structure. Concepts such as the butterfly switch arrangement, recently developed by Bolt, Beranek, and Newman, Cambridge, Mass., to manipulate voice data channels, have reduced processor node interconnection problems to manageable proportions. If packet communication of audio signals can provide the generality required of a good music processor, such an arrangement would be useful in real-time audio research.

Performance Systems

There are now more than two dozen digital audio performance instruments on the market. Older brand names such as Synclavier, Fairlight, and Con Brio—usually the more expensive products aimed at the high-end user—have been joined by a fleet of others spanning the price range of $60 to $70,000. The high-end models, although without the generality of full research processors, offer a wide array of controllable and expressive synthesis methods, showing evidence of aggressive and well-chosen research on the part of a few manufacturers. The low-end models, built to reach larger commercial markets, include an add-on synthesizer for an Apple computer and even a battery-powered portable keyboard.

The spectrum of potential users for high-end systems runs from stage performers to composers working in a studio. New England Digital Corp., the first company to venture into the digital synthesizer field with its Synclavier, has remained responsive to both groups. Its hardware system is geared to the performer, while an additional software language permits composers to write their own composing programs. The company has just announced a music printing attachment that produces scores and parts of any work played into the system.

Middle-level synthesizers have a more homogeneous market—primarily stage performers interested in nothing more elaborate than good sounds and a little memory. The recently announced Synergy (from Digital Keyboards, Inc.) is the first major entry in this market. Its high sound quality reflects both an extremely accurate control of frequency and innovative methods for exploiting this control. It should be noted that performing musicians will come to depend on the limitations of any instrument they are accustomed to using. Analog synthesizer players, for instance, will insist on a "fuzz tone" from overloaded amplifiers, requiring that a new digital instrument provide an imitation of this inferior technology. A product called Tubes even extols the superior ability of vacuum-tube amplifiers to distort in just the right way! The designers of Synergy have gone to some lengths to accommodate the idiosyncratic performer, providing the fat sound of detuned oscillators, among other options. When the product was unveiled at this year's national music fair in Chicago with a price tag of $5000, it reportedly prompted 2000 orders before production began.

Activity at the low end of the price spectrum is strong. Casio has introduced five portable keyboards this year and expects music to become the major part of its business by the next fiscal year. Kawai will announce a polyphonic synthesizer next summer that features reduced computation and word size. All signals will be carried as binary floating-point exponents. (Multiplies are thus adds, while adds incur format conversion.) The synthesis method is reportedly additive with sliding Fourier formants. Accurate sinusoids are claimed to be unnecessary. Marantz has introduced a new piano equipped with digital processor, control and driver circuitry, and a digital cassette for storing and retrieving music. The Pianocorder will listen to a performance on its keyboard and then reproduce it with optional changes of tempo or loudness. Cassettes of old piano-roll favorites are also available.

American home organ manufacturers generally show little interest in improving basic sound quality, preferring instead to dress up an instrument with as much add-on digital gadgetry as it will hold. Japanese and European makers, on the other hand, are interested in tonal effect. Yamaha, for instance, has secured the patent on Chowning's FM synthesis technique.

Digital Editing

Because computer music is often synthesized in sections or layers owing to the physical limits of local memory or of real-time processor speeds, the most active music centers have developed facile methods for manipulating sound files. Composition often relies heavily on editing techniques that can apply special envelopes, direction and distance cues, reverberant effects, and various kinds of equalization and filtering (as in weaving recorded instrumental or vocal sounds into the texture of a musical composition). Work on these techniques has been most active at Stanford University, IRCAM, and Queens University, Ontario.

Now the film industry, previously confined to archaic and largely mechanical processes for audio mixdowns and for synchronized accessing of prerecorded sound effects, is moving to digital methods. George Lucas has funded the development of custom digital sound and image processing devices for a sequel to the movie *Star Wars*. The audio editing stations will consist of some number of large audio discs, a reconfigurable control console that remembers the gestural preferences of each user, and a real-time audio processor capable of 16 million 24-bit integer multiply-adds per second. The audio processor, designed by J. A. Moorer and stemming from his IRCAM and Stanford experiences, can be microcoded to do signal analysis and synthesis, including fast Fourier transforms (FFTs), linear prediction, and adaptive filtering. The stations are expected to be doing production work within a year. The processor design is not presently available, but it is believed that Lucas will eventually release the design to an interested manufacturer.

Sound editing on random-access discs has been standard practice for some time at Soundstream (now a subsidiary of Digital Recording Corporation). Field-recorded digital tapes are dumped onto a disc and spliced digitally prior to production of a digital master for a client's record-cutting lathe. The company is also heavily involved in editing soundtracks for film, but with 8 channels of 50-kHz audio as standard it is running into the limits of magnetic disc technology in both capacity and speed. Soundstream is betting heavily on its laser-optic development as an alternative. This technology records on a 3×5-inch plastic card with a maximum capacity that exceeds twenty magnetic disc packs. Soundstream currently has audio playback of 2 channels at 50 kHz, with 8-track playback expected soon and 32 channels anticipated. Read/write capability suitable for audio editing is awaiting a computer interface, expected within the year. The company plans to market a commercial playback device about the size of a cigar box within two years. The card capacity will depend on the application targeted and will simply be a marketing decision.

The trend in audio storage and handling is clearly toward laser technology and fast random access and processing. The speed of this development is being hastened by the film industry, whose interests are clearly not limited to audio. Companies such as Sony and 3M are still aggressively marketing their sequential tape-editing audio products, but with Sony already marketing a read-only laser disc we can expect some major changes in its editing concepts soon.

Issues and Options

Whether in the design of faster algorithms for software synthesis or of audio processors for commercial distribution, questions invariably arise concerning audio fidelity and dynamic range. Software systems can avoid most of the problems by carrying all signals in 32-bit floating point. Real-time hardware processors designed with just 16-bit integer data paths can sometimes use double precision at the cost of some processing speed. The trouble spots for 16-bit audio processing include frequency resolution, slow ramping, and recursive filtering, all of which have been the object of many creative solutions that usually fail to go far enough. The 24-bit audio processor being developed by Lucasfilm will set a new standard for this kind of device but will still not have the generality of a floating-point system. An operating variant of Music 11 at M.I.T. has shown that 24 bits of floating point (16-bit fraction) is about right, being perceptually indistinguishable from the normal 32 bits for all standard audio processing.

Standards

The Audio Engineering Society (AES) has made a decision on digital audio standards for distribution media, which include 16-bit integer data sampled at 48 kHz (44.1 kHz if video-related). The 16 bits of linear data may not be enough. There was some lobbying for an 18-bit standard to handle higher-fidelity situations (such as the dynamic range typically found in a live symphonic performance). The proper reference point is the high-performance characteristic of the human ear, which has not only a best-case sensitivity to distortion that exceeds 70 dB but also an adaptability to local loudness levels that moves this window around, largely intact, over a wide dynamic range. The most compact representation of stored audio signals to meet those specifications would be a 16-bit floating-point format comprising 13 bits of signed fraction (for 78 dB of signal-to-noise ratio) and 3 bits of base-2 exponent (for an additional 42 dB of dynamic range). The M.I.T. studio has used a D/A device with somewhat higher specifications and found it exceedingly practical. Stanford has experimented with various other encoding schemes. Consumer-level audio products, which cannot afford to keep a 16-bit linear converter finely tuned and will have to settle for lower fidelity, would clearly benefit from a floating-point scheme. Analogic Corporation has tried to help things along by announcing a 16-bit converter that switches the data internally to sign magnitude format so that the signal-to-noise ratio at low dynamic range is improved. Digital Sound Corp., a leader in quality digital audio conversion products, employs these converters as part of its complete A/D–D/A systems.

Shortage of machine cycles in most real-time devices has kept sampling rates down and aliasing problems up. The new AES standard of 48 kHz is admirable, but it will not hasten compatibility between the anticipated distribution media and overloaded research signal processors. Up to 50% of processor overloading could be avoided by distinguishing between audio signals and control signals, but that imposes architectural demands that may be hard to meet. Real-time processors will probably continue to compromise on sampling rates for some time. At least one synthesizer manufacturer, New England Digital, has skirted the aliasing problem by using pitch-dependent sampling rates. Its Synclavier runs each voice at a sampling rate coincident with a power-of-2 harmonic above its fundamental. All aliased harmonics therefore lie on harmonic partials themselves, and the pitch-dependent sampling rate changes automatically even during glissandos. The concept is practical only for modular polyphonic voice fabrication.

The demand for flexibility plus speed in research-level audio processors has resulted variously in parallelism, pipelining, or asynchronous operation—and sometimes all three. Although music is a highly parallel acoustic phenomenon, the events within these streams are rarely spread out and tend to happen in clumps (or chords). Communicating the control information to a real-time music processor is quite problematic. In the first place, since updates of things like filter coefficients should happen together and "on time," schemes that simply share memory between a coefficient-calculating host computer and a real-time processor will be in severe trouble. Alternatively, a scheme that creates lengthy queues of precalculated parameter values ready to be burst across at memory speeds does not easily accommodate the queue jumping required by any real-time inputs such as keyboards, which need immediate servicing. The problem of providing a suitable control scheme has been consistently underestimated in all research and production processors to date. Gareth Loy at the University of California at San Diego has reported on the difficulties encountered with the Systems Concepts synthesizer; J. A. Moorer has documented the difficulties of the IRCAM 4C.

New Instruments

Future music processors will need to be responsive to human control in a wide variety of gestural modalities; improvements in real-time audio synthesis power will spawn new forms of sensory input, from touch-sensitive devices to neural connections, and the perceptual effect of each input channel should ideally be programmable. Max Mathews has devised an experimental instrument, the "sequential drum," that explores variable mapping of gesture onto effect. A melodic pitch sequence is pulled from memory one note at a time by successive drum strokes on a two-dimensional sensor. The timing, *x-y* position, and force of the stroke all control perceptual aspects of the audio synthesis. The Synergy instrument has a keyboard that allows timbre to be controlled by keystroke velocity. Experimentation and development of new performing instruments require that the command streams that control the synthesis be quite general and modifiable. For the larger research systems this is not a simple condition.

In the future there will be whole classes of new instruments, some for studio use, some for stage use, and others for the home market. We could even imagine a new amateur music-making activity, not dependent on the acquired dexterity of playing a two-keyboard, 64-button home organ, yet more artistically demanding than merely selecting an LP disc from a shelf. The amateur could "orchestrate" an existing score by selecting and assembling the timbral and spatial attributes for a personalized home-audio performance. Participating in a musical performance with such high-level decision making would be quite desirable to amateur music lovers. For this activity to result in sounds that are as aesthetically effective as those of commercial productions would require that the large computational loads be managed at prices that appeal to the consumer market. Representation of current signal processing methods in VLSI will of course hasten that goal. Equally effective will be the discovery of new signal processing methods that concentrate on the perceptually important aspects of music synthesis. That territory is receiving much attention right now, and the skills are moving almost as rapidly as the supporting technology.