AUDIO VIDEO SYNTHESIS

IN AN AGE so dominated by television and pleas for integration, it is not surprising that "art" would mirror these social interests.

In this article, author Jerry Hunt discusses audio/video synthesis as a logical extension of sound composing and illustrates the problems and advantages of interactive audio/video synthesis.

by Jerry Hunt

For me video is a direct, immediate, and natural global extension of the compositional activities associated with sound. I have not and do not now distinguish between electronic and non-electronic means as a procedural definition. Video is an electronic environment extending from the theatrical reality of producing sound; similarly, audio is an extension from the environmental necessities of sound producing.

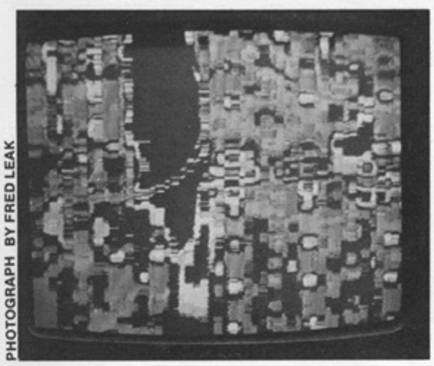
My background and training has been exclusively concerned with sound production, as a pianist and composer (1950-1960) engaging in compositional and performance activities which increasingly utilized electronic resources. David Dowe's specialization of training is visual, and the history of visual thinking in his own development is as extensive as mine has been in an aural orientation. This I think suggests something of the successfulness of our approach for us: the compositional activity can be highly specialized to take full advantage of the workers involved, and at the same time the system procedure (audio/video together) provides a situation of equivalency in variational decision-making.

The method, compositionally and procedurally, of working with interactively adaptive audio-video systems has also led to a reorganization of our thinking habits associated with the disciplines of music, film, painting, sculpture, dance, and theater. From my own independent work before association with David Dowe and the Video Research Center, I have developed a group of precedures and systems for the audio and video components of my compositions. Gradually areas of contact between audition and vision became more clearly self-evident. The early independent work and my work with David for the past five years has substantially reinforced my original intuition that the composing activity can powerfully operate and should operate in ways which are not dependent upon the patterns of feedback and feedaround through one special orientation of thought (action and process).

The perceptual, structural, and dynamic modalities of image and sound are profoundly unique and different in sometimes difficult ways. Certainly in regard to the requirements of parameter pattern extraction from signals produced by video cameras and microphones, these differences become exaggerated. The fortunate circumstance of systems for the electronic generation of image and sound is exactly that the parameters required for effective and powerful interactive operation exist in separable, definable, pattern-coherent structures, and the signal formats can easily be made compatible. (Unfortunately however, this in no way provides solutions to the still difficult synthesis problems for such systems).

Some aural/visual correlates seem to possess perceptual and/or historically reinforced constants of association—for example, the variation of average luminance in a visual display space and the dynamic variation of the intensity/spectrum relations of sounds through loudspeakers, and image orientation with reference to the viewing frame and sound distribution (regarded in both spatial and spectrum modulation aspects). Even more interesting was the situation in which image, (*still*—in a spatially defined viewing space), and sound (rhythmic/melodic *drone*) could be associated over a limited but highly variable range of characteristics. Furthermore, in this special situation provided by electronic systems, the variational histories of developments of characteristic features of dynamic pattern-sequences of sound and image can become the central compositional procedure.

Special attention to two areas of work became necessary to meet the demands of composition in this assumed interactively adaptive situation. First, the systems for signal generation had to be implemented in such a way as to allow selective parallelism of dynamic variation in image and sound and yet provide sufficient control flexibility/predictability to preserve the specific characteristics and integrity of a possibly large repertoire of image and sound development. The audio and video generators were implemented and have been continually modified and updated with the constant objective of increasing the integrity of *ensemble dynamic* parameter variation. (In electronic music systems the greatest problem in emulation has been in this area.) The objective required very early in the development of the signal generating systems a movement away from modular conception to a more integrated global structure. Second, because parameter-structure variations of image and sound develop in dissimilar ways, a system for extraction and analysis of ensemble-pattern over short-time histories was essential. Although unexpected, the results of *direct* interconnection of signals and controls are predictably limited and trivial.



PHOTOGRAPH BY FRED LEAK

A video image created with Jerry Hunt's and David Dowe's audio/video synthesis system.

The accompanying block diagrams illustrate the organization of the audio and video generating systems, the interactive-adaptive processor used with these systems, and a smaller derivative system of recent design for use with performances involving voice and voice emulation and specialized generators for image-sound.

The *audio generating unit* is a four-voice system organized as a parallel processor. All the function-subunits employ primarily analog electronic technology; all the static programming, matrixing, and routing subunits utilize digital implementations. A ROM[1] plug-in relates an array of 128 switches to major static program changes, allowing continuous performance manipulation. All parameters of all subunit functions are completely electronically presetable and full-range variable. Spectrum specification, formant, frequency, waveform, and amplitude modulation processes are all dynamically interrelated by user-defined specifications. Electronically programmed multiphase function generators operate on all levels of the system subunits and may be dynamically controlled. The input transducer ports accommodate a complement of a pressure/pressure-derivative sensing keyboard or, interchangeably, a similarly responsive band-controller, two three-dimensional foot pedals, and two three-dimensional trackers (one rotary and one similar to a joystick-array). The output structure allows re-entry and parallel organization of separate processes and signal channels. A component of the output subunit consists of a generalized delay processor which allows specialized spatial, spectrum, and frequency modulation.

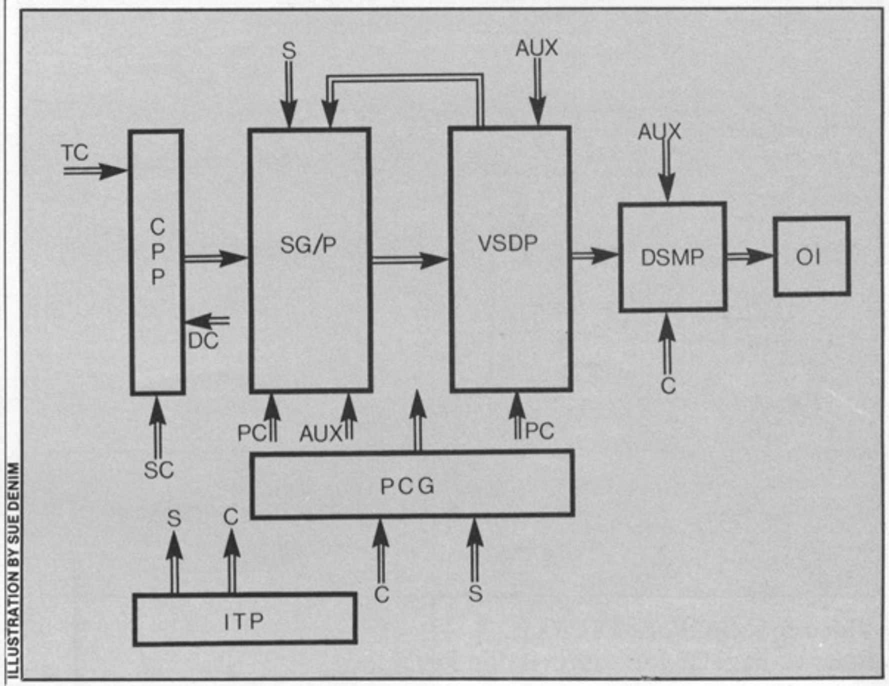
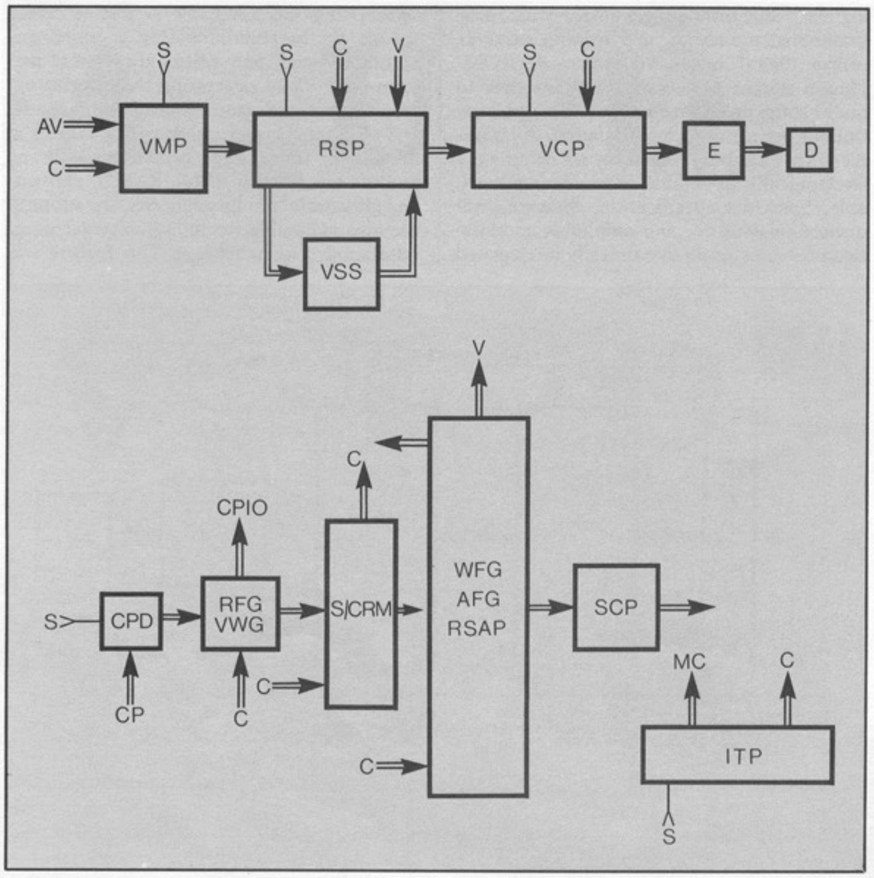


ILLUSTRATION BY SUE DENIM

Audio System Block Diagram.
Refer to page 29 for abbreviation key.

The *video generating unit* is a multiport input system also organized as a parallel processor. The system is completely synchronous: all waveforms and modulation involved in video generation of primary processing loops are related to the composite synchronizing signals[2] employed in NTSC[3] color video. The system generates up to six separate electronic images in the display space (television color monitor), and video camera signals and electronically generated video signals can be simultaneously accessed and accommodated. Subunits of the system perform basic video processing tasks: arbitrary colorization and color modulation, spatial and textural reorganization of the images in the display space, edge/line derivatives and modulation (PWM, PPM, PAM[4]), chroma[5] and gray-scale[6] modulation, etc. The subunits are also isolatable (including external patch interfacing and switching). This feature was necessary to satisfy the requirements of widely divergent applications of the system. Subunits for generating groups of electronically variable waveform sequences produce basic image structures; modulation of these subunits produces images in an arbitrary-scan format. Independent electronic control is available for spatial transformation and translation of images, movement, size and distance, apparent distance, surface and "environment" texture and shading, and viewing angle and color. A scan-conversion subunit

The perceptual, structural, and dynamic modalities of image and sound are profoundly unique and different in sometimes difficult ways.



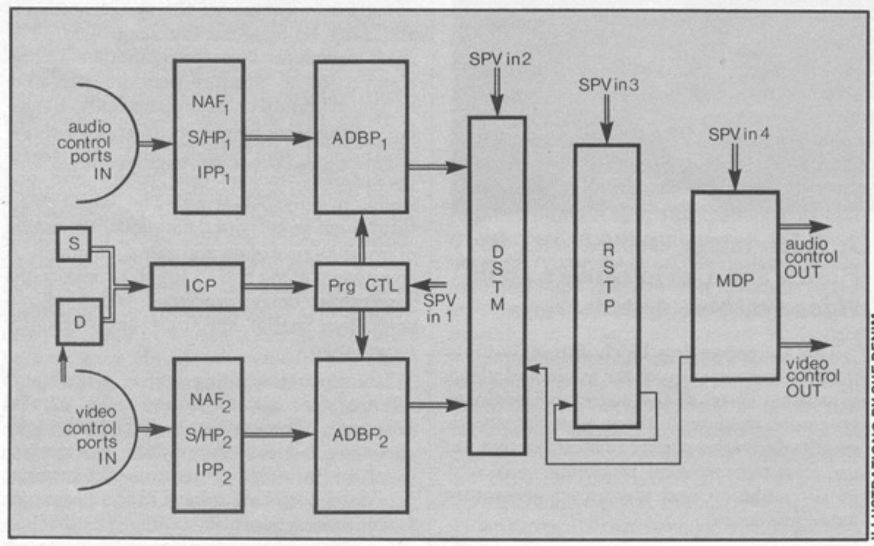
Video System Block Diagram.
Refer to page 29 for abbreviation key.

performs the format translation for NTSC and is input-output uncommitted to allow electronically generated and camera-scanned signals to be interchangeably processed, mixed, and routed. The scan-converter subunit is an optically coupled system; a selective digital framestore subunit allows the static storage, delay, and processing of limited image feature-extraction components. The final mixture of signals is encoded into the NTSC composite color video[7] signal and displayed over color television monitors, video recorded, and/or broadcast through standard television transmission systems.

The *interactive processor*, like the generating systems, has been continually altered in implementation and structure and has undergone two complete rebuildings of hardware since 1972. Control signals describing an array of parameter designations to be monitored from the two systems are scaled and modified for electrical compatibility and each routed to one of three varieties of preprocessors. Slope derivatives of dynamically varying controls and synchronizing signals control the access to these inputs. The preprocessors perform in each parameter channel a nonlinear adaptive filter[8] function; several channels allow sampling functions and/or delay processing. Programmable weighting and matrixing is then performed to produce groups of unique signal branches. Processors similar to adaptive linear variable-threshold amplifiers[9] are then utilized to generate sequences of parallel pattern codes which are stored in a recirculative memory system. Programmable dynamic 'template' models are applied according to program instructions which provide for dynamically variable goals under some arbitrary set of limits. This process is cyclic and continuous and produces, in relation to input signals, output codes which are dematrixed and arithmetically manipulated to generate for the audio and video systems two separate groups of interrelating parameter controls. Pre-program settings establish degrees of parallelism and 'goal-pointers' between image and sound parameter designations. The control signals to and from the generating systems are routed using firmware connections and may therefore be arbitrarily assigned to a given channel location.

The large system has been employed to produce components of audio/video recorded work, but primarily it has found use as a system for investigation and dynamic real-time performance operation (*Haramand Plane*). The smaller, derivative special purpose system (used for performance of my *Cantegral Segment* (*s*) 19, 20 and David Dowe's *Repons*) and its associated synthesis modification systems have been developed since 1975 as a direct outgrowth of the larger system. The signal sources involve microphone and camera transducers (and interchangeably since January 1977 a performable voice-emulator and special image-synthesizer). Image and voice parameter extractors operate on a limited selective range of features to provide strings of signals which control the variation of the processors (and/or generators) with respect to dynamic performance definitions. We do these works frequently on tour since they are more easily transported, adaptable to various circumstances, and require minimum rehearsal and set-up.

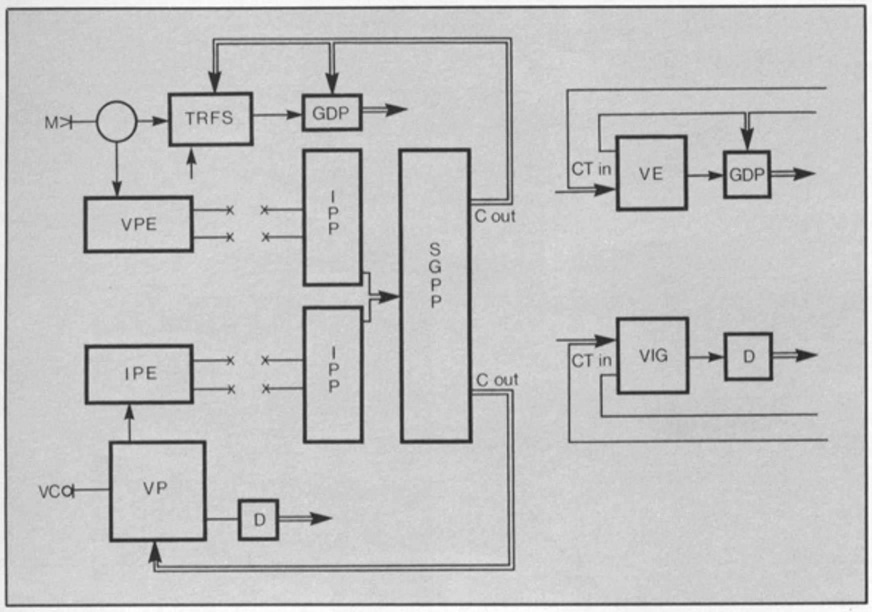
No specific philosophy of design or technical suppositions have been imposed upon any subunit or system aspect of the devices other than electrical compatibility and all-electronic



Interactive Processor Block Diagram.
Refer to page 29 for abbreviation key.

ILLUSTRATIONS BY SUE DENIM

Since the beginnings of our work independently and later together, most of the technical and developmental work has been financed by us directly (with the very important exceptions of the National Center for Experiments in Television in San Francisco at the time of establishing the Video Research Center and, currently, the Rockefeller Foundation). This fact and the important influence of constantly re-evaluated cost/performance assessments have been, I now think, frequently fortuitous forces in defining proposed implementations. A further auxiliary observation of increasing importance for me was the byproduct of work on coding and pattern recognition theory during preliminary study and auditory space—for the video particularly a 'tactile' element seems to develop as a side benefit. Third, the interactive systems are in no way "sequencers" or randomizers but rather responsive mediums of information *exchange* and *extension*. Although not intelligent in a cybernetically austere sense, the systems have begun to produce for our work some results which suggest several modes of re-orientation for us: composition is dynamically related to the responsive cycles of activity required (composition as performance); goal setting and patterning in this dynamic activity of composition dissolves some aspects of the image/sound separation; finally, and most importantly, pattern-struc-



Cantegral Segment(s) Processor with transducer pickup and direct versions.

for my work *Haramand Plane* (the first of our works interactively executed): an insight into the possibilities of virtual high-level operations using carefully limited computation and storage elements. This observation involves economics in no way related to technical or financial considerations but informs directly the eventual result for electronic art activity generally.

Finally, for our work together involving interactive processes, several delightful situations occur. First, because of the system structure of the audio and video generating units, dynamic sequences of deviation from *still* (video) and ornamental *drone* (audio) can be arbitrarily interrelated by a wide range of different variational principles. At the same time complete performance autonomy is maintained. Second, because of the dynamic provisions, both systems are essentially performance instruments (this is important even in recorded, highly editorial modes of working); this fact encourages the finding of ways of working which sometimes closely correspond to the experiences of encounter in visual turing formalities emerge which relate to a procedure which approaches one of my long-felt enthusiams for electronic means—an almost colloquial and at once multidimensional/global exchange of aspects of the perception of living.

Video System

S sync
Avin auxilliary video inputs
Cin control inputs
Vin primary video inputs
RSP raster-scan processor
VSS video selective frame storage
VCP video colorizer-processor
E encoder (NTSC)
D distribution driver
VMP video mixer/processor
CPD clock processor/distributor
CPIO camera processor-inputs/outputs
RFG raster function generators
VWG video waveform generators
S/CRM signal/control routing matrix
Cout control signals-outputs
WFG waveform function generators
AFG arithmetic function generators
RSAP random-scan arithmetic processors
SCP scan-conversion processor
Vout video outputs
ITP input transducers-processors
MCout matrix/routing controls out
CP clock program (input)

Audio System

TCin transducer controls-inputs
SCin static controls/presets-inputs
DCin dynamic controls-inputs
ITP input transducers/processor
S sync
Cout controls (all) out
PCG programable control generators
Cin control inputs
PCin program controls-inputs
Aux in auxilliary program or signal inputs
VSDP voicing, spectrum, distribution processors
DSMP distribution and spatial modulation processors
OI output interface (mixing, drivers etc.)
CPP control program/processors
SG/P signal generators/modulation processors

Interactive Processor

S sync
D slope derivative generators
ICP input control processor
Prg CTL program control
SPV statis program variables
NAF nonlinear adaptive filters (preprocessor subunit)
S/HP sample-hold processors (preprocessors subunit)
IPP input signal preliminary processing (Preprocessor subunit for scaling etc.)
ADBP arithmetic, distributive and 'blend' processors
DSTM dynamic sequential 'template mapping'
RSTP recirculative sequential template processor
MDP matrix decoding processor

Special purpose processor system

M microphone
VC video camera (B&W)
VPE voice parameter extractor
IPE image parameter extractor
VP video processor
D display (distribution drive etc.)
TRFS transversal/recursive filter structure
GDP generalized delay processor
IPP input control parameter processors
SGPP sequential 'goal-pointer' processors
Cout control out to TRFS control inputs and VP control inputs resp.

CTin control transducers-inputs
VE voice emulator
VIG video image generator

---

[1]ROM—read only memory.

[2]sync—synchronizing pulses employed to establish time relationships.

[3]NTSC—National Television Standards Committee—used generally to refer to the electrical and related standards to which American television systems must conform.

[4]PWM, PPM, PAM—Pulse width modulation, pulse position modulation, pulse amplitude modulation.

[5]chroma—the color-information carrying component of a composite color video signal.

[6]gray-scale—the graduations of brightness between black and white in a television display or signal (luminance).

[7]composite color video—a signal containing timing (sync) chroma and luminance information electronically combined.

[8]nonlinear adaptive filter—a filter whose characteristics vary according to some function of the input/output relationships.

[9]adaptive linear variable—threshold amplifier—an amplifier whose region of linear amplification varies according to a required characteristic of the input/output.