

Vidsizer -- A Visual and Musical Instrument

By

Daniel Eric Franzblau

B.S.A.D. Massachusetts Institute of Technology  
1977

Submitted in Partial Fulfillment  
of the Requirement for the  
Degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1979

© Daniel Eric Franzblau 1979

Signature of the Author \_\_\_\_\_

Department of Architecture  
June 23, 1979

Certified by \_\_\_\_\_

Nicholas Negroponte  
Associate Professor of Computer Graphics  
Thesis Supervisor

Accepted by \_\_\_\_\_

Nicholas Negroponte

**Rotch**  
MASSACHUSETTS INSTITUTE  
OF TECHNOLOGY

SEP 27 1979

LIBRARIES

## Vidsizer -- A Visual and Musical Instrument

<u>Table of Contents</u>	<u>Page</u>
Table of contents	2
Abstract	3
Acknowledgements	4
Introduction	5
History	7
Sound-light Phenomena	11
Reflections	15
Visualizing Sound	16
Theory: Hardware and Software	19
Vidsizer as a Performance Instrument	21
Conclusion	23
Appendix -- An Introduction to the Synthesiser	25
Footnotes	30
Bibliography	31

Vidsizer -- A Visual and Musical Instrument

Daniel Eric Franzblau

submitted to the Department of Architecture on June 23, 1979  
in partial fulfillment of the requirements for the degree of  
Master of Science.

Abstract

This thesis deals with the issues of creating simultaneous sound and image: how the two affect each other, what physiological basis exists for explaining sight/sound interaction, and how to approach the visualization of sound. Also discussed is how these issues apply to the Vidsizer system and the system's role in performing sound-light works now and in the future.

Thesis supervisor: \_\_\_\_\_  
Nicholas Negroponte

Title: \_\_\_\_\_  
Associate Professor of Computer Graphics

## Acknowledgements

I would like to thank Professor Negroponte for his patience, indulgence, and critical eye, Jacob White for the design and construction of the synthesiser and its digital interface, and Bob d'Ancona for providing encouragement and ideas throughout. I'd also like to thank Ben McCann, Rim Igniatis, Paul Heckbert, Walter Bender, Mike Conners, and the rest of the hacking community for a great learning environment and much learned.

## Introduction

Vidsizer is a sound and image making system. Its purpose is to make possible the creation of live sound-light events, with the goal of explaining sound through image and image through sound. Toward this end, the design of software and hardware provides for live synchronous control and synthesis of sound and image. Vidsizer, as an instrument, will hopefully express emotion and convey feeling. It is an experiment to explore the visualization of sound, and to attempt the achievement of expression and control in both realms.

The system is programmed on a 32-bit minicomputer. A digitizing tablet provides nearly all input while output comes from two sources: images from a digital video frame buffer and sound from an analog synthesiser. The synthesiser's patching and control levels are set in hardware by an 8-bit microcomputer, which receives its commands from the large computer through a single 600 baud serial line. The micro also automatically refreshes all levels in the hardware interface under interrupt control.

Having introduced the system briefly, the next sections will continue with a history of earlier sound-music

machines, and then a discussion of the ideas behind the  
vidsizer system.

## History

There have been many attempts at creating sound-with-image instruments, but even more common have been hypotheses written about them. Newton, in his 1704 publication "Optiks" equates the relative spaces of the colors of the spectrum with ratios of the notes of the scale, leading toward a vibrational explanation of harmony:

May not the harmony and dischord of colors arise from the proportions of the vibrations propagated through the fibers of the optic nerve into the brain, as the harmony and dischord of sounds arise from the proportions of the vibrations of the air?

(1)

This analogy between pitch and color was echoed by others in later years.

In 1734, Father Louis Bertrand Castel, a Jesuit mathematician, attempted a model of the first instrument for color-music performance, the "Clavessin Oculaire." (2) Castel's system assigned one color for each note of the scale -- C was blue, D green, E yellow, G red, A violet, and C again blue. In each successive octave the colors were a lighter shade. The entire instrument was never built, since Castel was more intrigued by the idea than the actual

device. He did envision at one point the colors produced by torch light shining through polished jewels, all in a room lined with mirrors! With this vision he foretold the next stages of color-music -- projected light.

Alexander Wallace Rimington was born in London, 1854. Raised to be a civil engineer, he left for Paris in his twenties to study art. Always fascinated by light and color, he finally concentrated on light itself as the basis of a new art-language. His "Colour-Organ" was finished in 1893, and installed in a studio in Norwood. It was no small device, having 14 arc lamps and drawing 150 amps. There was one lamp per note in the octave (12 in all), and 2 extra for background and contrast illumination. In front of each lamp was a system of lenses, color filters, and irises to give control over intensity, hue, and position of the image. Chroma variations were achieved through a gradation of density in each color filter. For display, Rimington experimented with flat white walls, screens made of white chord hung vertically in sheets, and thick folds of curtains, discovering that the curtains added a deep richness to the light. His overall conclusion of the project was that sound-music and light-music can be simultaneously performed, with the "mutual enhancement of their respective emotional effects." (3)

In 1919, one year after Rimington's death, Thomas Wilfred created his first "Clavilux" in Huntington, New

York. After several models, his final design had sliding keys grouped into "manuals" like a pipe organ, 4 form-throwing lamps that shone onto a 3-way split viewing screen which could be shifted to catch the image on different sections for different effects. In front of the lamps were prisms which could be tilted, twisted, moved close or far from the lamps, or shifted up or down. All four were focused on the center of the screen, so images could be combined. In addition, there were two lamps for background illumination. Wilfred believed that form, color, and motion were the visual counterparts to melody, harmony, and rhythm, so he built the "Clavilux" for precise control over these factors. His machine was played from notation like an instrument, and thus its light-music was open to interpretation by the instrumentalist.(4)

Another experimenter was Mary Hallock Greenwalt. In 1927 she applied for four separate patents to cover her process, including notation, rheostats, the illuminating means, and the means for controlling light. A pianist and a student of Leschetizky, she designed the instrument to be used in an orchestra pit; a hemispherical dome behind the action on stage was the projection surface. Her color instrument was thus designed especially for performance in a public space.(5)

One thing all these past experiments had in their favor was that they used projected light. Moving the image

was as simple as moving a lense in front of the lamp. Similarly, distortion, size, and coloration were controlled with equal ease. This "analog processing", if you will, is real-time and computation free (unlike digital frame buffers.) Projected light also solved the problem of display for an audience. These color organs performed well in large theatre-like spaces, and even outdoors if ambient light was low.

The result of these years of experimentation was an art of color-music employing projected colored light. Common analogies between music and color began to form: a single light was like a melody, two lights at a time were an interval, three or more were a chord. Fading and growing lights were the dynamics of color-music. Throughout, the notion that pitch should be represented as hue was still strong, a notion that I would like to refute later.

## Sound and Light Phenomena

To begin to understand how sound and vision relate to each other, an appreciation of just what makes up these sensing mechanisms is important. I do not intend to give a course in sensory physiology, but I will point out some interesting facts about hearing, vision, and their central processing network.

To begin with, let me emphasise the relativeness of it all. In living tissue, there is no true zero; everything works by relative means.(6) Joseph Albers states

We almost never (that is, without special devices) see a single color unconnected and unrelated to other colors. Colors present themselves in continuous flux, constantly related to changing neighbors and changing conditions.(7)

Color contrast, adaptation of the eyes to darkness, and of the ears to loud sounds are all examples of the relative way perception works.

Relative perception allows for a great range of sensitivity: the eyes can handle from one ten-thousandth to one billion candles per square meter, the ears from one ten-thousandth to ten million or more dynes per square centimeter, and for the ears the maximum can be boosted by

building up a tolerance.(8) At the very lower limits of perception, the eardrums vibrate in and out only one tenth the diameter of a hydrogen molecule, and can almost pick out the random brownian motion of the air. The eyes will respond to only a few photons of light.(9) These are amazing devices!

To talk about and understand these ranges easily, a logarithmic scale is used. A function F of neural response (or perceived stimulus strength - the two have been shown to be equivalent) versus the actual stimulus strength can be written

$$F=k*(S-S_0)^n,$$

where S-S<sub>0</sub> is the supra-threshold stimulus strength. This equation is known as the Stevens power function.(10) On a log-log plot, n would be the slope of the line given by a certain set of stimulus/response data. Some senses, like pain perception, have an n greater than 1, so that as the stimulus grows the perception grows faster. Both vision and hearing have n's well below 1, corresponding to their wide range of perceptible stimuli.

Besides these similarities of operation, there are similarities of structure between the two senses which may shed some light on sound-sight phenomena. One fact really amazed me when I came across it: ears separate and focus sound on the sound sensing parts of the ear just as light is focused into an image upon the retina. Changes in stiffness

of the basilar membrane (a membrane separating two of the three collateral canals in the inner ear) and other properties of the fluid filled canals serve to boost then fade the maximum amplitude of a sound traveling through the inner ear. The resulting place of maximum deflection is different for each frequency, thus mapping different frequencies to different points in the ear. The sensory cells are primarily excited at the site of the maximum, so each frequency excites different sensory cells. This is called the "place theory of hearing." (11)

Both eyes and ears are subject to contrast enhancement. When a neuron is stimulated, it tends to shut off the cells surrounding it. Those that surround the cell stimulated the strongest receive the strongest lateral inhibition. In the eyes, this leads to edge detection and simultaneous contrast of colors. Not much is said about the effect of lateral inhibition in the ears, but it certainly boosts contrasts between frequencies. These edge effects help the nervous system to overcome its diffuse nature by sharpening discrimination.

In the brain, senses and motor functions are mapped onto projection areas covering about 20% of the surface. The ears, as well as the eyes, have a projection area -- frequency is mapped from the ears, spatial correspondence from the eyes. The remaining 80% of the surface consists of regions called "association areas." They are connected to

the projection areas by massive fiber tracts, and usually each association area is coupled to several projection areas. Through these and additional connections at motor centers, vision and hearing are intrinsically linked.

It makes neurological sense, if somewhere in the nervous system the visual and auditory systems do interconnect, that one sense can influence the other. Helmholtz first described the summation of subliminal stimuli in muscle, and this was confirmed by Richet, Piotrowski, Locke, Basch, and others. (12) Forbes was the first to suggest that the reflex summation was not due to a second response from the same motoneurons, but due to other motoneurons affected by a central summation of the effects of two stimuli. The idea of summation is basic in the theory of neural integration, and in the auditory and visual systems there are plenty of opportunities for such summations to occur.

Experimentally, it has been shown that sound can influence vision -- if two sensory paths converge on a motor center, the reflex due to either may be altered in speed and magnitude. Simultaneity is best; other temporal patterns may inhibit rather than facilitate a response. Although there are negative findings, the weight of evidence is in favor of one organ enlarging the sensation from another. (13)

## Reflections

Joseph Albers notes in his 'Interaction of Color' that the visual memory for colors is poor compared to the audio memory for a tune. (14) Is this a valid comparison? What is the relationship between sound and image? In the environment, events that create sound can often be seen as well -- a fish jumps, a ball bounces, lightning strikes nearby. It seems only reasonable that some way of using or interpreting this concurrent information exists. I therefore propose that sound/vision concurrence is not learned, but is an inbred part of our sensory physiology. There may be no real a-b comparison of visual abilities, like color memory, to auditory abilities. All abilities may instead be part of the same audio-visual continuum. Looking for analogies between the two "separate" senses is going about things the wrong way. One should rather look for an audio or visual impetus that arouses some understanding in the related sense, and through that relation gain insight into the realm of sight-sound.



*Electric clock*



*Refrigerator*



*Dishwasher*



*Blue jay*

Visualizing Sound

A central theme of this thesis is that sound can be realized by image, and conversely, an image can be designed to induce association with sound. The visualized sounds at the beginning of this section are prime examples. Through repeated patterns and other techniques, Saul Steinberg has created a visual bridge to the audio world.

Are Steinberg's drawings that direct, though? It certainly would be hard to imagine these sound from their images alone, without the written cue and without having heard them before. But they do work. There is a lot of information in these sketches: dynamics, overall pattern, frequency, and a hard-to-describe quality. The bluejay really appears to squawk, while the soft metronome of the electric clock comes across naturally, too. The pictures are fixed in time, and therefore employ temporal cues to convey a feeling for rhythm. If they are read from left to right,

the time factor is apparent. The whole image gives an impression of the sound "integrated" over time. Besides the learned convention of left-to-right, there are other, intuited cues at work here. Dark, rough images are used for loud and harsh sounds; spatial frequency is used to represent short and long term rhythmic patterns; repeated visual patterns illustrate repeated patterns in the sound. The refrigerator and dishwasher drone continuously, but with other internal sounds that help to make up the whole.

These sketches show that visualization of sound can be achieved. In my experiments with vidsizer, I will try to avoid learned graphic conventions, concentrating rather on intuitive links between sound and image. The dynamic nature of the system will obviate the need for temporal cues. Any image is possible, limited only by imagination, programmability, and the restriction of real-time. Within these limits, for the early experiments, I will try to develop a visual language for dealing with sound. Several possible associations exist: pattern from rhythm, sequence, texture, and sound density; color from pitch changes, harmony, and texture; and form from the envelope, structure, and overall texture of the sound.

In past experiments using an abstract form under graphical control, it became evident that a simple module shape could easily overpower the total effect. The forms were static -- they remained where they were drawn, until

drawn over. Eventually the screen became cluttered and the initial effect of the module disappeared. Motion and module lifetime will need to be approached much more critically in the new attempts. Also, I'll try to get around the hard-edged rectangular form so easily produced by computer hardware. Softness is a useful visual function for illuminating sound.

Techniques for accomplishing these effects exist -- an "onion skin" brush that peels off layers of video memory will create a soft image with the proper color set. Motion can then be added by rotating colors. Another approach is to have the brush incorporate motion in its production and then clean up after itself by fading or erasing old copies. Color is important in the overall effect because contrast and therefore harshness can be amplified or diminished by the proper choices.

When creating the sound-image piece, either the audio or visual part could inspire the other, or both might be created together. Through employing known techniques or experimenting with new ones, one can be made to match the other.

## Theory: Hardware and Software

For both playing and tuning, Vidsizer uses the tablet extensively. Choosing the tablet for input was easy: they are common around the lab, and they are naturals for graphical input. Tablets work in real time, too. Systems without real-time control must be told what to do ahead of time; they rely on elaborate specification schemes with many layers of commands -- great for automatic playing, but terrible for improvisation. Real time specification allows instantaneous changes in plan, and with disk storage and playback of tablet data then automatic playing can be implemented easily enough.

The tablet can also be thought of as a raw-data generator. This idea is central to vidsizer. Tablet data such as x position, y position, z value, speed (  $f(dx,dy)$  ), and last sample's values, along with random numbers and the position of the slider make up the "system parameter list" at the heart of the system. These values are all translated into fractions between 0 and 1, so they can be used interchangeably by the video and audio programs. The tablet thus becomes like a joystick, to be used for parameter control as well as graphics. The beauty of this system is that to record a worked-out sound and image pair, only the system patchings and the (x,y)'s of the tablet need to be

saved. To play it back, putting the (x,y) pairs through the system as if from the tablet will reproduce the "recorded" effect.

Vidsizer will have the ability to play at least one live and one recorded sequence simultaneously, so duets of sound or image can be performed, or either can accompany the other. The "GROOVE" system, developed at Bell Labs in New Jersey, approached similar issues of real time control and disk storage of sequences. GROOVE uses the disk extensively, and is able to handle up to 40 concurrent disk functions of time, monitor 7 input knobs, and reference up to 40 stored periodic functions. In GROOVE, you can type algebraic expressions to define a disk function, and have the expression compiled and run by the system. One reason for the power of the system is its freedom from other duties. Its only task other than generating the functions and doing disk i/o is refreshing a crt display and getting characters from a keyboard. It is definitely a special-purpose system. Vidsizer can benefit from GROOVE's ability to edit its recorded sequences, so that feature will be included.

## Vidsizer as a Performance Instrument

Vidsizer's use as a performance instrument was anticipated since the early days of the project. Recording tablet movements allows playback, and recording internal configurations allows quick setup to a pre-arranged state. These are musts for performance ability.

A real performance instrument will be transported often, so portability and ruggedness is another requirement. In its present state, though, the system is hardly transportable. A new version, that runs only on microprocessors, is realistic. By running two or three micros in parallel, each with its own duties (input monitoring and control, video generator, and audio generation), a reasonable system could be developed.

At the moment, all video functions need to be programmed in pl/l. Through intelligent programming, one "function" can provide a wide variation of image, controlled by the tablet. So with a library of these functions the video imagery will provide interest for quite a while. New functions could be written to match new sounds, and would thus be added to the library. New programs will be modeled after the modularity and interchangeability of current functions.

To live up to its promise as a performance instrument,

some problems will need to be faced. First of all, the interface that controls the synthesiser currently will need to be made more responsive. It will need to be able to shift quickly from one sound specification to another, as fast as the performer desires. Some speed can be added by new software support, but ultimately it will be the hardware design that will prevent further speed.

Display, too, needs to be considered. At present, there is no easy way to achieve the concert-hall scope of the early projection systems. Video projectors exist, but they are bulky and expensive. One solution might be a series of Advent style monitors, which opens up the possibility of two or more different video images, maybe from two or more Vidsizers. A third solution is a series of regular tv monitors placed throughout the area. This solution is not as good, since the graphics really need to be on a large screen for viewing in public.

## Conclusion

The next steps for Vidsizer are clear. A bigger synthesiser, with more than one voice, will allow layers of sound and polyphony. Feedback from programs into the central data structure -- a computed "system parameter" for other modules to use -- would really open up possibilities. A larger more complex interface to the analog world, including many DAC's instead of the current single one, would simplify and speed operation of the synthesiser interface. In place of a digital frame buffer, a digitally controlled analog video synthesiser would help make vidsizer portable. Much less computation would be needed, so a micro could serve well as the controlling processor. The micro industry itself is undergoing revolution -- microprocessors are becoming as fast and powerful as their bigger cousins, and cheaper all the time. This means that future micro-based equipment will shrink and will keep or gain more ability. There is a trend starting among performing groups to get the clutter of equipment off stage, so remote control of off-stage equipment and compact instruments will become much more common. Microprocessors and digital control fit right in.

Home entertainment is in transition, too. Consumer videotape machines and the new videodisks are providing an alternative to commercial tv. The recording industry is

beginning to appreciate this fact, and soon live recorded concerts, complete with video, will be found next to the lp's in record stores. An instrument for sound and vision makes sense in light of these developments, as the video-music artform grabs hold and takes form.

Appendix -- An Introduction to the Analog Sythesiser

When dealing with sound, there are many terms -- volume, pitch, harmonics, fundamental frequency -- that help describe what is going on. These terms can all be related to the shape or waveform of the sound (see Figure 1.) Volume is the amplitude of the wave: a louder sound will have a larger amplitude. Pitch, on the other hand, is how fast the wave oscilates up and down. A high pitch is rapid oscilation, a low pitch is slow oscilation. Most natural sounds produce complex looking waveforms, very different in shape from a simple sinusiodal oscilation. However, all waveforms can be represented by a sum of simple sinusoid oscilations of varying amplitudes and frequencies. These summed simple waves are called "partials" of the whole waveform; when the partials are related to the lowest or "fundamental" frequency by integer values, they are called "harmonic partials."

In a synthesiser, one source for waveforms is the oscilator. Four wave shapes are commonly produced by oscilators: sine, saw-tooth, triangle, and square (see Figure 2.) A pulse-width modulated wave is simply a square wave that spends more time in one side of the cycle than the other. The ratio of up-time to down-time is called the "duty cycle" -- a square wave has a 1:1 duty cycle.

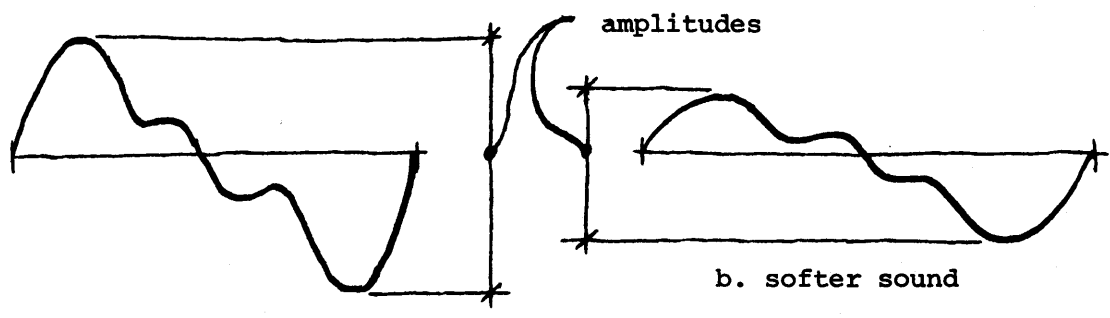
Of all these waveforms, the sine is the simplest, a pure tone. The triangle is the most like the sine, with only a few significant additional partials. A square wave has many partials, all with odd-numbered ratios (3\*fundamental, 5\*fundamental, etc.) As the partials in the square wave get higher in frequency, their amplitudes approach 0. In other words, the higher frequencies that make up a square wave are present in smaller amplitudes than the fundamental frequency and its near neighbors. The sawtooth wave has a similar fading of the higher partials, but unlike the square wave it has both even and odd numbered partials. With so many frequencies together in one wave, its unmodified sound is fairly harsh, approaching plain noise. True noise has equal amounts of all frequencies, and is a useful addition to the 4 main waveforms (it makes a great wind sound with a little filtering.)

There are many techniques available for manipulating sound. Filtering changes a waveform by allowing only certain partials through. By applying a changing voltage to a control input, modulation results. Modulating an oscillator to change the frequency of oscillation produces a vibrato effect, modulating the volume on an amplifier creates an envelope. Attenuation means simply turning down the volume (making the amplitude of the waveform smaller). Mixing two waveforms adds them together to make a third.

To manipulate sound electronically, there must be some

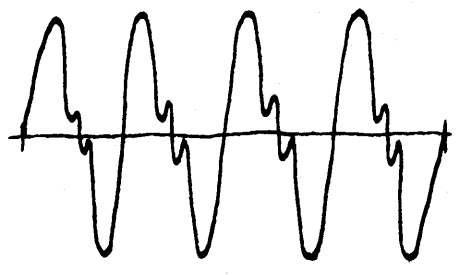
way of converting sound to electricity and back. Microphones and speakers solve this problem. As changing air pressure against the microphone moves an internal coil of wire in a magnetic field, a varying voltage is produced in the wire that corresponds to the initial sound. Electronic oscillators produce varying voltages directly, bypassing the need to start with a sound. Once in electrical form, the wave can be modulated, filtered, mixed, and otherwise transformed to the desired wave. Amplifying this result and sending it to speakers will convert the electricity back to sound.

In synthesizers, the idea of waves being voltages is important, because both the waves and the means for controlling the wave shapers are voltages. Thus an oscillator can be used to either make a sound, or to modify how another module is creating or acting on the sound. If a saw-tooth wave were used as a volume control in an amplifier, a saw-toothed envelope would result. The synthesizer in the Vidsizer system is specially designed to allow any output to be the source (to be patched to) any input, as either signal or control. It has six main modules: an oscillator with four output waveshapes, another oscillator especially for low frequencies, a mixer, an envelope generator, a filter, and an amplifier. The creation of a sound is limited thus only by the available inputs and outputs, and by experimentation with different patchings.

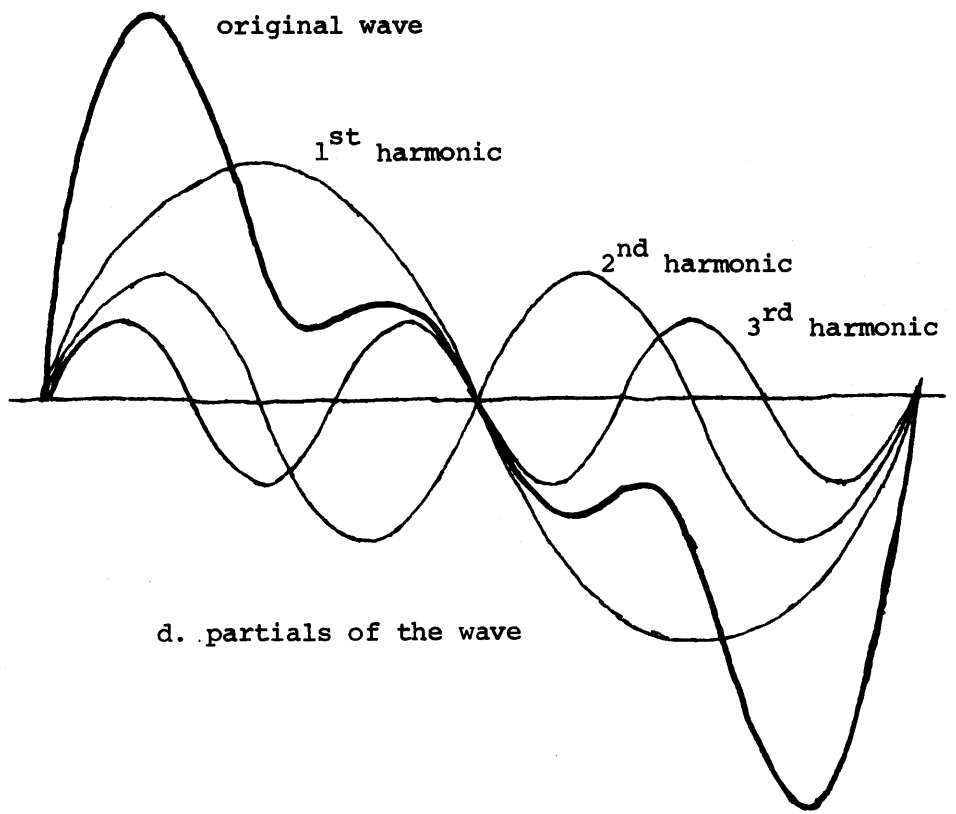


a. sound wave

b. softer sound



c. higher pitch



d. partials of the wave

Figure 1.

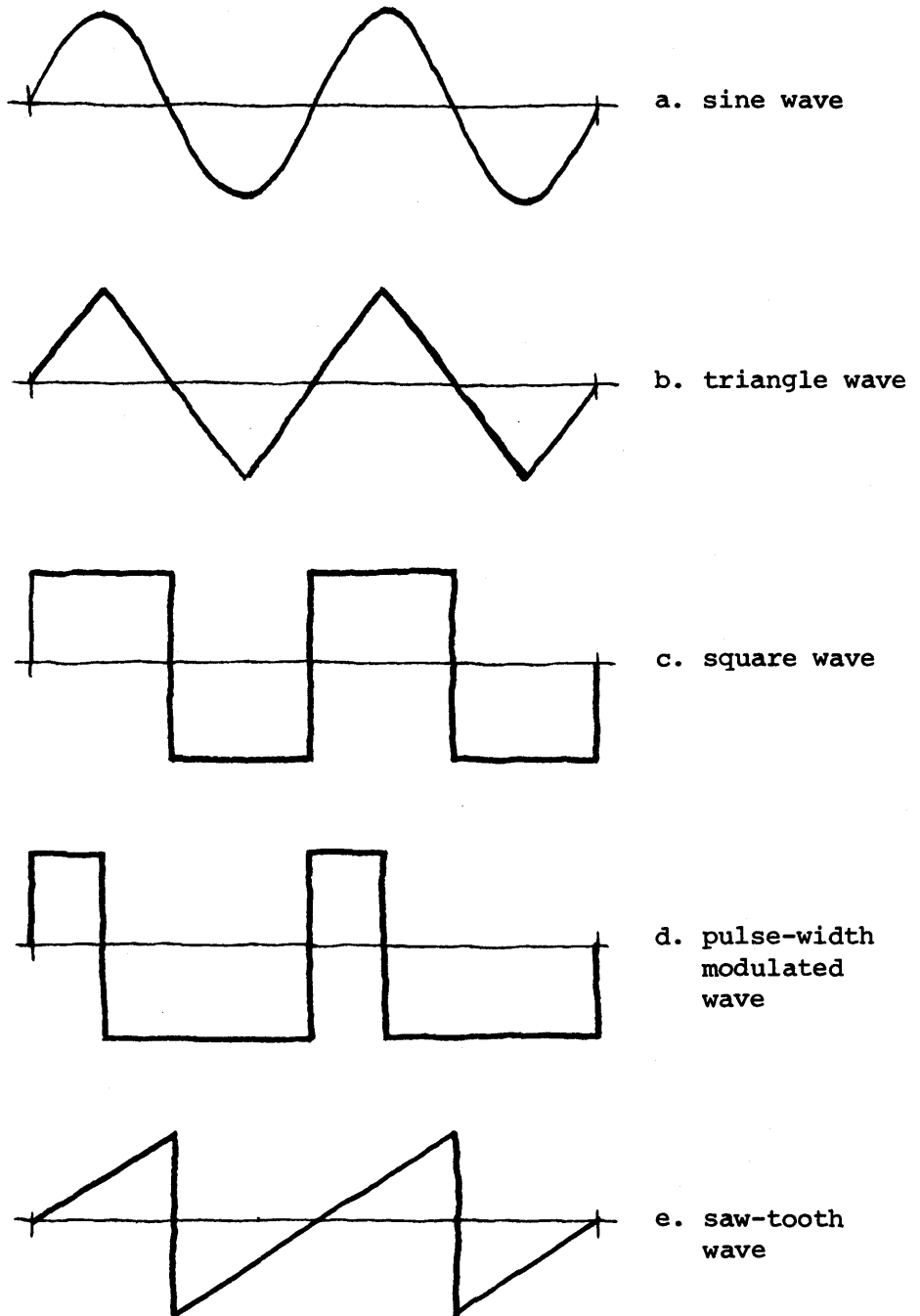


Figure 2. Typical oscillator outputs

## Footnotes

- (1) Adrian Bernard Klein, Coloured Light, an Art Medium, page 62
- (2) Klein, page 1
- (3) Klein, page 9
- (4) Klein, chapter 9
- (5) Klein, chapter 9
- (6) Donald J Harris, Ph.D., Some Relations between Vision and Audition, page 9
- (7) Joseph Albers, Interaction of Color, page 5
- (8) Harris, page 8
- (9) Harris, page 6
- (10) Robert F. Schmidt, Fundamentals of Sensory Physiology, pages 16-18
- (11) Schmidt, pages 193-194
- (12) Harris, page 41
- (13) Harris, page 47
- (14) Albers, page 3

## Bibliography

Albers, Joseph, Interaction of Color, New Haven, Yale University Press, 1971

Harris, Donald J., Ph.D., Some Relations between Vision and Audition, Springfield, Charles C. Thomas Publisher, 1950

Klein, Adrian Bernard, M.B.E., A.R.P.S., Coloured Light, an Art Medium, London, The Technical Press Ltd., 1937

Mathews, M. V. and F. R. Moore, "GROOVE -- A Program to Compose, Store, and Edit Functions of Time", Communications of the ACM 13:12-715+, December, 1970

Marx, Ellen, The Contrast of Colors, New York, Van Nostrand Reinhold Company, 1973

Risset, Jean-Claude, "Musical Acoustics", chapter 12, Handbook of Perception, Vol IV, Academic Press, Inc., 1978

Schmidt, Robert F, editor, Fundamentals of Sensory Physiology, New York, Springer-Verlag, 1978

Strange, Allen, Electronic Music, Systems, Techniques, and Controls, Wm. C. Brown Company Publishers, 1972

Von Foerster, Heinz, and James W. Beauchamp, Music By Computers, New York, James Wiley and Sons, Inc., 1969

Wells, Thomas, D.M.A., and Eric S. Vogel, The Technique of Electronic Music, Austin, Sterling Swift, 1974