basics of digital sound, part one:
digital recording

By John Strawn

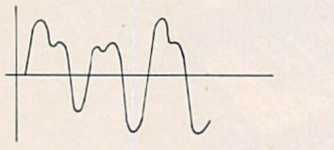
DURING THE LAST couple of years, the big news in music technology has been the application of digital electronic equipment to sound processing. Considerable work has been done in generating sound digitally, and more recently digital recording has appeared on the scene. In this two-part article we will offer a basic introduction to the theory behind digital recording and synthesis. This month's article will focus on the storage of existing sounds in digital form, and the sequel, in the June '81 issue, will explore ways in which computers and other digital equipment can be used to generate new sounds directly.

Analog Representations Of Sound

WHEN A LISTENER hears a sound, it reaches the ear after travelling through the air from some source. More specifically, the listener hears sound because the air pressure is changing within certain limits. The sound must be loud enough to be audible — that is, the fluctuations in pressure must be great enough to be felt by the sensitive bones of the inner ear — and the frequency must be neither too high nor too low.

The simplest method of summarizing such pressure variations is to draw them in the form of a graph, such as that shown in Fig. 1. At the

Figure 1.

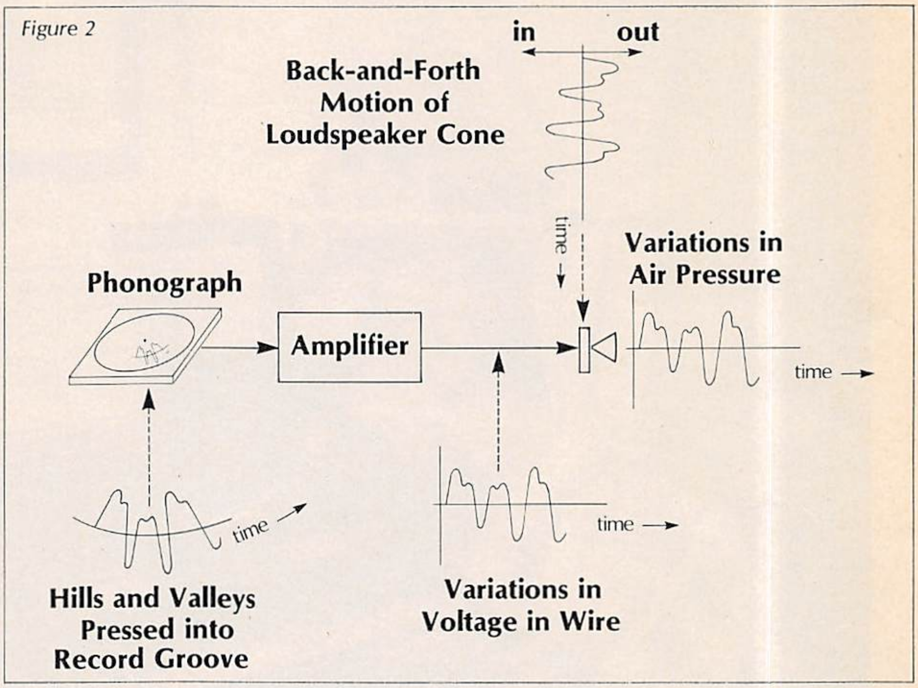


points where the curved line is near the bottom of the graph, the air pressure is lower, and when the curve is near the top of the graph the air pressure is higher. The passage of time is shown from left to right.

Sound involves changes in air pressure. But similar changes can occur in other media than air. For example, when a loudspeaker is generating sound, the loudspeaker will move in and out in a way that matches the changes in air pressure. The loudspeaker's movement inward from its rest position creates a decrease in air pressure, and its movement outward creates an increase. Still another type of change which matches the changes in air pressure would be the electrical quantity called voltage in the wire connecting the speaker with an amplifier. Note that we will not have to worry here about defining what voltage is. For the purposes of this article, we can simply assume that it is possible to modify some electrical property associated with the wire in a fashion which matches the changes in air pressure. Thus, Figure 1 could also illustrate the changes of this voltage or the changes in position of the speaker cone. This relationship is shown more fully in Figure 2.

There are other ways of representing sound

Figure 2



besides drawing pictures such as Figure 1. For example, Figure 1 could be reduced considerably until it fit along the groove of a phonograph record, as shown in Figure 2. The sides of the grooves on a phonograph record contain a continuous representation of the sound stored in the record. As the needle moves past the hills and valleys in the groove, the needle is moved back and forth. This motion is then changed into voltage in a wire, which is amplified and eventually sent to the speaker.

Within certain limits, it is possible to convert the kind of fluctuations shown in Figure 1 from one of these media (phonograph record, voltage, loudspeaker movement, air pressure) to another. If errors occur in the conversion process, then the sound will be changed in various (usually undesirable) ways. This can include the addition of clicks, pops, and other noise, or the introduction of one or more of the varieties of distortion commonly listed in the specification sheets for commercial audio equipment. But the fact that this conversion between media can occur at all is very important to understanding digital audio. We can summarize the rest of this article by pointing out that generating or reproducing digital sound will simply involve converting a string of numbers into one of the time-varying changes we have been discussing. If these numbers can be turned into voltages, for example, the voltages can be amplified and fed to a speaker.

Before entering the world of digital sound, however, we should point out an important characteristic of the time-varying quantities we have introduced. Each of them is more or less exactly *analogous* to the other. If we draw a graph of air pressure variations, for example, it will look very similar to a graph of the variations in loudspeaker position. The term "analog" serves to remind us of the analogous relationship of these quantities.

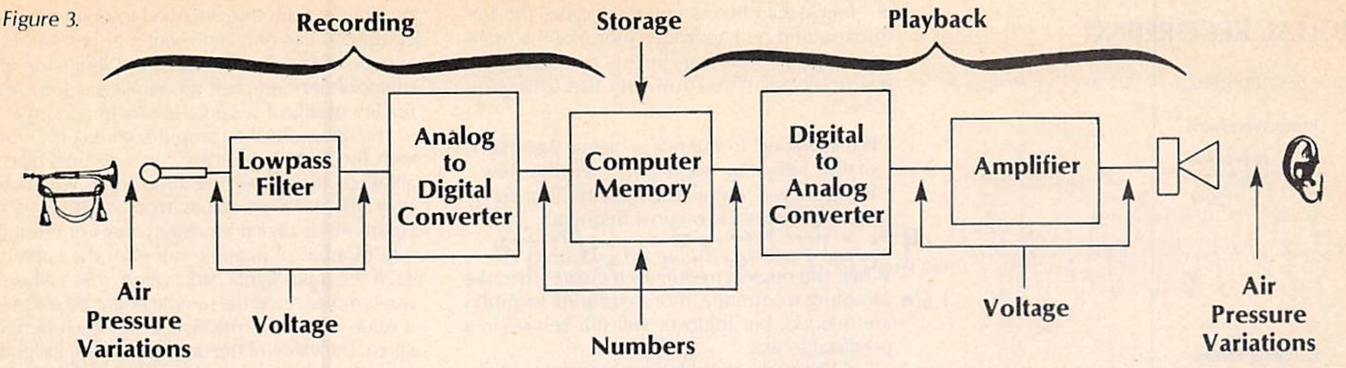
Digital Recording And Playback Of Sound

THE PROCESS BY which a computer can record a sound and then play it back is summarized in Figure 3. The sound is changed from air pressure variations to voltages by the microphone, and the voltages are passed through a wire to the *analog-to-digital converter,* commonly abbreviated ADC. This device changes the voltages in the wire into a string of numbers — a necessary step because computers work with numbers, not with variations in voltage.

The conversion to numbers is illustrated in Figure 4. Figure 4a shows an analog waveform (a sine wave), and Figure 4b shows its digital counterpart. After each number is created by the ADC, the number is stored in memory by the computer. As long as the computer's memory is not changed, the sound can be stored indefinitely without loss of fidelity.

Actually, there are several different media besides computer memory which can be used at this stage for storage. Various tape recorder companies (such as 3-M, Japan Victor, and Matsushita of Japan) have developed digital tape recorders in which one or more audio tracks are recorded on tape in digital form. Anyone interested in one view of the state-of-the-art of this kind of digital tape recorder should write to

Figure 3.



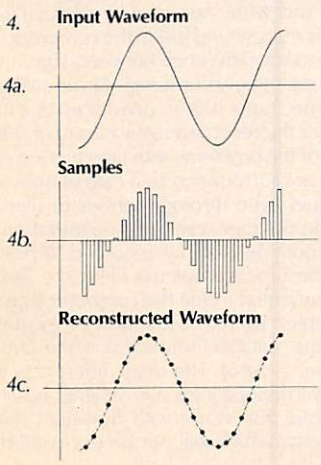
Studer Revox America [1425 Elm Hill Pike, Nashville, TN 37210] and ask for their free 30-page booklet on audio PCM, which just came out. Other companies (Pioneer, Sony, Philips, Telefunken) are working on converting video disks for storing digital audio. It is also possible to convert a video cassette recorder for digital audio. For now, we will use "computer memory" to stand for any kind of digital storage. At the end of this article, we will talk about the practical applications of digital audio.

When the listener wants to hear the sound again, the numbers are read one by one from the computer memory and passed through another deviced called a *digital-to-analog converter,* abbreviated DAC (pronounced "dack"). This piece of equipment changes the numbers into one of the analogous representations of sound shown in Figure 2, specifically voltage. From this point on, the process is the same as that shown in Figure 2; the voltage is amplified and fed to a speaker.

Sampling

THE SIGNAL shown in Figure 4b has some properties that make it significantly different from the original signal shown in Figure 4a. The first of these is the fact that it is defined only at certain points in time. This happens because the signal has been *sampled.* That is, at certain times

Figure 4.



a sample of the original signal is taken. Each vertical bar in Figure 4b represents one sample of the original signal. In the recording process, the ADC measures the level of the voltage and records its measurement as a number, and these numbers (the samples) are stored in the computer. The higher the bar in Figure 4b, the larger the number. A sample is used to represent the original signal until the next sample is taken. Sampling is often called *pulse-code modulation,* or PCM.

The rate at which samples are taken, the *sampling frequency,* is expressed in terms of samples per second. This is a very important quantity in digital sound recording and synthesis. For reasons which we will discuss shortly, sampling frequencies as high as 50,000 samples per second are not uncommon in high-quality digital audio work. Of course this represents an enormous stream of numbers. At a sampling rate of 50,000 cycles per second, the computer must be able to handle no fewer than three million samples for one minute of monaural sound. For stereo, this quantity would be doubled!

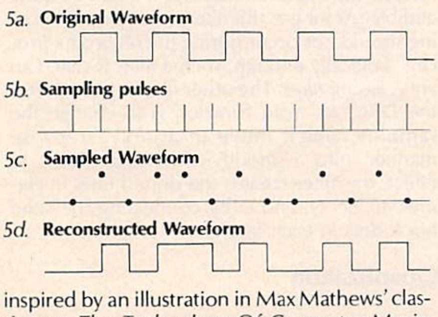
Between the bars, Figure 4b does not show what the value of the signal is supposed to be. In fact, the width of the bar is usually assumed to be extremely narrow, perhaps lasting only .00002 second (two *hundred-thousandths* of a second). This means that if the original signal changes between the bars, the change is not reflected in the height of a bar, at least until the next sample is taken. In technical terms, we say that the signal in Figure 4b is defined at "discrete times," each such time represented by one sample.

Part of the magic of digital sound is that if everything works right, the digital-to-analog converter and associated hardware can exacty reconstruct the original signal from these samples. This means that, given certain conditions, the missing part of the signal between the samples can be restored. In effect, the samples are connected together (the dotted line in Figure 4c) when the numbers are passed through the DAC, so that the signal sent to the speaker looks and sounds very much like the signal that went into the computer.

Aliasing (Foldover)

BUT THE PROCESS of sampling is not quite as straightforward as it might seem. Just as an audio amplifier or speaker can introduce distortion into sound, sampling can play tricks. An example is given in Figure 5. (This figure was

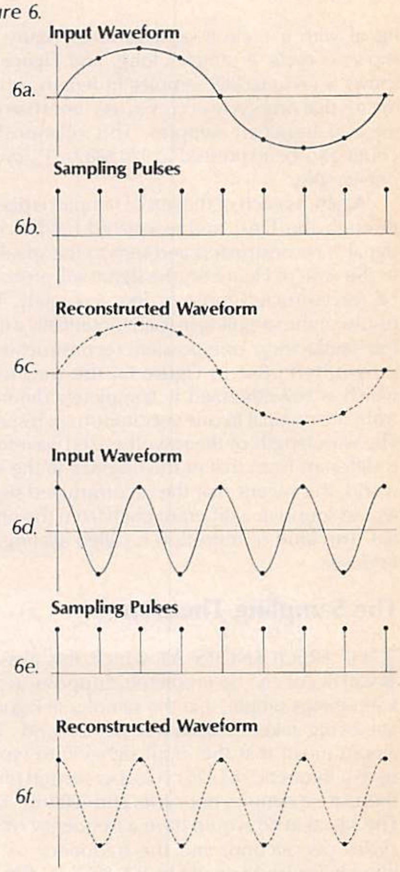
Figure 5.



inspired by an illustration in Max Mathews' classic text *The Technology Of Computer Music,* published 12 years ago by the MIT Press.) The input waveform shown in Figure 5a is sampled at the times shown in 5b. As before, the resulting samples, shown in 5c, are stored in the computer memory. But when we attempt to reconstruct the original waveform, the result pictured in 5d is completely different.

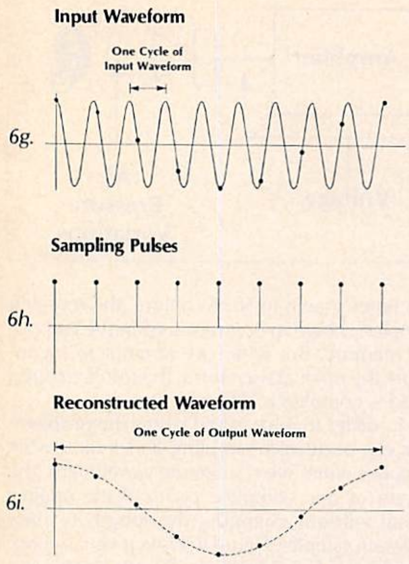
In order to understand better the problems that can occur with sampling, let's look at what happens when we change the wavelength (the length of one complete cycle) of the original signal without changing the length of time between samples. Figure 4 shows a signal whose wavelength lasts 18 samples. Figure 6a shows a

Figure 6.



John Strawn graduated from the Oberlin Conservatory in 1973, where he studied organ with Haskell Thomson and electronic music with Ron Pellegrino and Robert Moore. After two years at the Technische Universität in Berlin on a Fulbright scholarship, he spent another year abroad in various electronic music studios on a Thomas Watson Fellowship. Since 1976, he has been a Ph.D. candidate at the Computer Center for Research in Music and Acoustics (CCRMA) at Stanford University, working with John Chowning. In addition, he has served as one of the editors of Computer Music Journal for the past three years. This article is excerpted from an anthology entitled Computer Music Tutorial, scheduled to be published in 1982 by the MIT Press.

Figure 6, continued.



signal with a cycle 8 samples long, Figure 6d shows a cycle 2 samples long, and Figure 6g shows a cycle 0.8889 samples in length, which means that one cycle occurs in less time than the interval between samples. This relationship could also be expressed as 1/0.8889 = 1⅛ *cycles per sample.*

Again, as each of the sets of samples is passed through the DAC and associated hardware, a signal is reconstructed and sent to the speaker. In the case of Figure 6c, the signal will probably be reconstructed more or less accurately. The results of the sampling in 6f are potentially a little less satisfactory; one possible reconstruction is shown here. But in Figure 6i, the waveform which is resynthesized is completely different from the original in one very important respect: The wavelength of the resynthesized waveform is different from that of the original. In the real world, this means that the reconstructed signal would sound at a different pitch from the original. This kind of distortion is called *aliasing,* or *foldover.*

The Sampling Theorem

THE FREQUENCIES AT which this aliasing will occur can be predicted. Suppose, just to keep things simple, that the samples in Figure 6 are being taken 1000 times per second. This would mean that the input signal in 6a would have a frequency of 125 cycles per second (since there are 8 samples per cycle, and 1000/8 = 125). The signal in 6d would have a frequency of 500 cycles per second, and the frequency of the input in Figure 6g would be ~~888.8888.~~ . . But the frequency of the output signal in 6i is different. You can count that there are about 8 samples per cycle of the *output* waveform. In actuality, the output waveform occurs at a frequency of 111.1111. . . cycles per second. Thus the frequency of the original signal has been changed by the sampling and reconversion process. This represents an unacceptable change for a musical signal, and it must be avoided if possible.

We can generalize from Figure 6 that as long as there are at least two samples per cycle of the original waveform, the resynthesized waveform will have the same frequency. But when there are fewer than two samples per cycle, the frequency (and perhaps much more) of the original signal will be lost. In this case, the new frequency can be found by the following formula:

> If the original frequency is higher than half of the sampling frequency, then the new frequency is equal to the sampling frequency minus the original frequency.

If the original frequency > sampling frequency, then the output f is the input − sampling f.

When the original frequency is greater than ~~the~~ 1.5 x sampling frequency, more complex formulas are needed, but foldover will still behave in a predictable way.

In less mathematical terms, suppose we have chosen some fixed sampling frequency. We will then start with some signal at a very low frequency, sample it, and resynthesize the signal from the samples. As we raise the pitch of the input signal, the pitch of the resynthesized signal will still be the same as the pitch of the input signal until we reach a pitch which corresponds to half of the sampling frequency. As we raise the pitch of the input signal even higher, the pitch of the output signal will start to fall again! When the input signal reaches the sampling frequency, the entire process will repeat itself. Another way of saying this is expressed in the so-called *sampling theorem,* also called the *Nyquist theorem.* This says that in order to be able to reconstruct a signal, the sampling frequency must be at least twice the highest frequency in the signal being sampled.

This means that for a particular application, we must know the highest frequency which will be sampled by the ADC. Then the sampling frequency can be specified as being twice as much. In many real-world situations, the sampling frequency is taken to be somewhat greater than twice this highest frequency, because the converters and associated hardware cannot perfectly reconstruct a signal near one-half the sampling frequency.

As already mentioned, the sampling frequency is often set as high as 50,000 samples per second for digital audio applications. There are reasons for making the equipment work so hard. Depending on the individual, the human ear can hear frequencies as high as 15,000-20,000 cycles per second. To represent 20,000 cycles per second adequately, a sampling frequency greater than 40,000 cycles per second would have to be chosen. For a variety of reasons, sampling frequencies near 50,000 have become popular.

In order to make sure that the system shown in Figure 3 works properly, two filters are included. One is placed before the ADC, to make sure that nothing (or as little as possible) in the input signal occurs at a frequency higher than one-half of the sampling frequency. Such input by itself would probably not be audible to human ears, but after foldover it might be quite audible. As long as this filter does its work, aliasing should not occur during the recording process. Logically enough, such a filter is called an *anti-aliasing filter.* The other filter is placed after the DAC. Its main function is to change the samples, which move in a jerky, stepwise manner, into a smooth continuous signal. In effect, this filter creates the dotted lines in Figures 4c, 6c, 6f, and 6i by connecting the solid black dots in each figure.

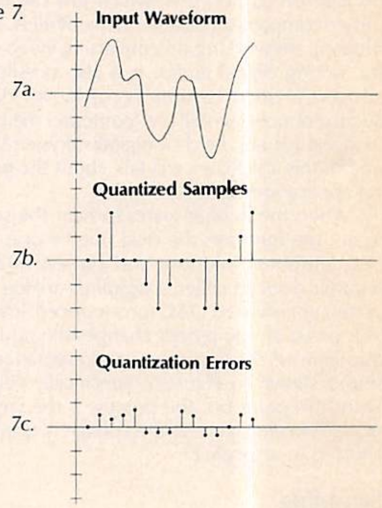
Quantization

ANOTHER DIFFERENCE between the signals of Figure 4a and 4b is that the height of the bars in 4b cannot take on any given value. In order to explain this, we need to point out that computers can only represent numbers within a certain range. The details of this limitation are unimportant here, but the implications for the fidelity of digital sound are very important.

For example, if a computer could only work with the "whole numbers" (no decimal places allowed) from 1 to 100, then there would be only 100 different values that could occur in Figure 4b as digital representations of the infinite number of possible values of the curve in 4a. If the input signal had a value of 54.3 when it was sampled, then the sample might be assigned a value of 54. This means that for each sample taken, the value of the sample will differ slightly from the value of the original signal. This aspect of digital signals is known as *quantization.*

The errors that occur in quantization are illustrated in Figure 7 (other kinds of errors are

Figure 7.



associated with the conversion process as well, but we won't discuss them here). If we were to listen to just the errors in Figure 7c, they would sound like noise. If these errors are quite large, one might notice something similar to tape hiss at the output of the system. The amount of such *quantization noise,* as it is called, varies with the signal and with the accuracy with which the signal is represented inside the computer. There is one major difference between tape hiss and quantization noise, however. On an analog tape machine, tape hiss is produced no matter whether the recorded signal is loud or soft. This is one of the problems with tape hiss — very soft signals are surrounded by a halo of noise which continues even through periods of silence on the tape. But there can be no quantization noise when nothing is being recorded. If the input signal becomes completely silent, the signal will be represented inside the computer by a series of samples, each of which will be exactly zero (or some suitable number that the DAC will interpret as zero). The small differences in Figure 7 will disappear for such a signal. (Quantization noise also varies with the input signal in other ways which will not be discussed here.)

Dynamic Range

THERE IS ANOTHER important implication of the degree of accuracy with which a sound wave can be quantized. In order to explain this, we need to refer to the number system used inside a computer. Instead of digits (from 0 to 9) computers work with *bits* (the word is a contracted form of *binary digit*). We won't discuss what a bit is, or how one adds and subtracts using bits. But it can be important to know *how many* bits are used for representing a single sample of sound inside a computer. If we want to represent 1000 different possible values of a sampled signal, then we will need to store 10 bits per sample. If we want to be able to distinguish 1,000,000 different possible values for a sample, we will need to store 20 bits per sample. (This assumes that each sample is converted and stored as a whole number.) The specifications for digital sound equipment typically include data on the accuracy of the system, which is sometimes expressed as the number of bits which the system uses to store each sample.

This number of bits per sample is important in calculating the *dynamic range* of the system. In general, the dynamic range is related to the difference between the loudest and softest sounds that the system can handle. The dynamic range is measured in *decibels* (abbreviated dB). Even without going into the technical definition of decibels, we can work with decibels if we keep the following facts in mind: The most sensitive range of human hearing extends from approximately 0 dB, which is the threshold at which the softest sound can be heard, to something near 120 dB, which is the threshold of pain. And a difference of 1 decibel between the level of two sounds corresponds to the smallest difference in level which can be perceived. (These figures vary with age, training, pitch, and the individual.)

In recording music, it is important to capture the widest possible dynamic range if we want to reproduce the full expressive power of the music. In a live orchestra concert, for example, the dynamic range can vary from silence to an instrumental solo at 40 or 50 dB to a full orchestra, which can exceed 100 dB. The upper limit of the dynamic range of most analog tape equipment is dictated by the physics of the recording process, and lies somewhere around 70-75 dB for the better consumer tape machines. The dynamic range of phonograph records is even narrower. This means that when a recording is made onto tape or records, the softest passages must be made a bit louder by the recording engineer, and the loudest passages must be made a bit softer. If this weren't done, the loudest passages would produce distortion and the softest passages would get lost in the hiss and other noise of the system.

Returning to the number of bits per sample, we can calculate the dynamic range of a digital recording/playback system by the following simple formula:

dynamic range in dB = number of bits × 6

Thus, if 12 bits are used to record sound digitally, the dynamic range will be approximately 72 dB, which is no better than the dynamic range of the better analog tape equipment. But if we store 16 bits per sample, the dynamic range is boosted to 96 dB, a considerable improvement. Of course, if the original sound signal did not have such a wide dynamic range, using this many bits would not *increase* the dynamic range. Rather, some of the capabilities of the system would be left unused. (This assumes that we are storing each sample as a whole number. Other *encoding schemes* offer other possibilities.)

Unfortunately, the wider dynamic range means a more expensive system. As one might expect, it takes almost twice as much memory to store 20 bits per sample as to store 12 bits. In addition, the cost of the DAC and ADC escalates with the number of bits they can accurately convert. Converters can cause a wide variety of distortion themselves, which we will not discuss here except to point out that a 16-bit converter is not necessarily accurate to the full dynamic range implied by 16 bits. One can build a 16-bit digital recording and playback system, but if the system only has an overall accuracy to 14 bits, then the dynamic range implied by the 16-bit figure will probably not be achieved without some distortion or noise occurring.

Digital vs. Analog Recording

WE HAVE SEEN that a digital signal can only take on a value at certain discrete times, and that the numbers used to represent the digital signal can only take on certain discrete values. Such a signal, with discrete values occurring at discrete times, is known as a *digital signal.* This is significantly different from analog signals, which, crudely expressed, can take on any value at any time.

To close, we should talk about the advantages of digital audio recording. Ever since tape recording was invented, there have been problems with tape noise, print-through, and other phenomena associated with (analog) magnetic tape. Tape hiss is especially irritating, even in a single track; and when several tracks are mixed down to one, all of the hiss in each of the original tracks is added into the final mixdown. Likewise, when you make a copy of a tape, the amount of noise increases somewhat, and making copies of copies causes further deterioration. Of course, noise reduction methods like Dolby or DBX can help, but they still don't tackle the problem at its source.

These problems more or less disappear in digital audio. When you make a digital copy of a digital recording, you are simply copying numbers. As long as the numbers are copied correctly, no new noise is introduced into the copy. It is sometimes even possible to correct a mistake in copying the numbers after the mistake has occurred! It can be impossible to tell the original from the copy simply by looking at the numbers in the recordings. Wow, flutter, and other kinds of distortion associated with analog tape records and LP records disappear as well.

There are already dozens if not hundreds of "digital" recordings on the market. It is important to keep in mind what part of the recording process was digital. In many cases, the recording will be digital only to the extent that a digital tape recorder stored the audio. But the mixers used to lay the original tracks as well as for mixing down and mastering may still be analog. The audio industry is just beginning to think about adopting a fully digital studio, in which the mixers, equalizers, reverberators, and other such equipment will all work with digital audio, that is, with numbers. And obviously, the record album that you buy stores sound in analog form.

It will be some time before the full benefits of digital audio reach the consumer. For the home unit of the future, imagine a recording inside a cartridge (like the computer game cartridges). You plug in the cartridge, and the playback unit reads the recording straight out of the cartridge's memory. Until we reach such a science-fiction stage, you can expect a drawn-out "settling down" period while the audio industry agrees on storage and reproduction formats for consumer products.

Understanding digital recording provides useful background for understanding digital synthesis. In the June issue of *Keyboard,* we will go on to look at how the computer and other digital equipment can be used actively rather than passively to generate digital signals.

All Styles — Amateur And Professional

CONTEMPORARY

KEYBOARD®

JMS

March 1981
USA $1.50
$1.75 Canadian
90p UK

POLL WINNER WITH 'YES'
GEOFF DOWNES

GEORGE SHEARING

ALEXIS WEISSENBERG

JOBS IN NASHVILLE

DIGITAL RECORDING

03
0 74470 66788

FROM THE PUBLISHER

EVERYTHING ABOUT KEYBOARDS. That's what this magazine is here for. But for six years we've been ignoring the most vital element a keyboard player works with — the human hand. Just as singers must understand and take care of their voices, and runners their legs and feet, a knowledge of how the hand functions is something no keyboardist should neglect.

This month's special feature, "The Hand," focuses on this unique appendage. What is the structure of the hand? How did it evolve? What's the best way to take care of your hands? If something happens to them, what are the chances of successful surgical repair? For expert information on these and other topics, we turned to Dr. Robert Chase, a surgeon at the Stanford Medical Center in Palo Alto, California, who specializes in reconstructive surgery of the hand.

Hands and fingers seldom get the respect they deserve, or even the amount of attention and care that we routinely give our cars. This collection of bone and muscle is a true miracle of evolutionary design. Yet how many keyboardists stop to realize the part their hands play in their music? The hand is the link between our brains and our instruments, and keeping it functioning properly is equally important. If any one of the three isn't working, we're in trouble.

Fascinating topic, equally fascinating article. I think you'll enjoy and learn from it.

* * * *

Over the past few months we've been introducing some new and important how-to-columns. January offered the beginning of a new series by the brilliant synthesist Roger Powell; in February we introduced columns by highly acclaimed classical pianist Garrick Ohlsson and veteran studio keyboardist Dick Hyman. And next month we'll be welcoming the multi-talented, multi-faceted multi-keyboardist **George Duke** to our roster of artist/columnists.

George has been the subject of two *Keyboard* cover stories (July '77 and October '79). In his column he will be exploring numerous aspects of keyboard performance, including such topics as backing singers, soloing, selecting tunes and arranging them, and much more. I'm sure you'll be as excited as we are to have George's contributions in the magazine, so be sure to join us next month for his first column.

* * * *

Over the years we've presented many different instructional columns by well-known artists, technicians, and historians. What have we omitted? What column topics do you wish we'd include each month? If we hear from readers that we should cover keyboard maintenance every month, for instance, we'll start looking for a columnist to write about it. So let me know what you'd like to see among our columns.

— JC

A GPI Publication

CONTEMPORARY KEYBOARD®

Vol. 7, No. 3 — March 1981

The Magazine For All Keyboard Players

Columns

Departments

Cover photo by Carol Friedman

Also from GPI: Guitar Player Magazine, Tom Wheeler, Mng. Editor • Frets Magazine, Roger Siminoff, Editor

CONTEMPORARY KEYBOARD (ISSN 0361-5820) is published monthly by GPI Publications, which also issues *Guitar Player* and *Frets* magazines.

Controlled circulation at Long Prairie, Minnesota. Canadian postage paid at Winnipeg, Manitoba. **POSTMASTER:** Please send form 3579 to P.O. Box 28836, San Diego, CA 92127. **EDITORIAL AND EXECUTIVE OFFICES:** 20605 Lazaneo, Cupertino, CA 95014; (408) 446-1105. Publisher assumes no responsibility for return of unsolicited manuscripts, photographs, or artwork. **SUBSCRIPTIONS:** $17.95 per year, outside U.S. and Canada $23. Address all subscription correspondence to Subscription Dept., P.O. Box 28836, San Diego, CA 92127. Allow 7-9 weeks for new subscriptions or change of address Send mailing label plus old and new addresses for address change.