

# Vocoders

*The secrets of the mysterious vocoder revealed.*

**I**F THERE WILL ever be a list of electronic audio instruments which have caused confusion in the professional audio and music industry, the vocoder will definitely be on it. Recently, the mysterious vocoder has become a desirable, though costly, recording studio instrument. However, the mystery element persists, often because the actual purpose and musical versatility of the vocoder are somewhat—or often, totally—misunderstood.

## THE BASIC PRINCIPLE

Though it is hardly possible to list all efforts made in history to synthesize human speech, the name of Homer Dudley is inseparable from today's vocoder. In 1936, he patented his apparatus to analyze and remake speech, which he called a Vocoder, because it was based on the principle of coding the voice and then reconstructing the voice in accordance with that code. (So please, no more "Vocorder"—it has nothing to do with recording the voice!)

The principle of the Dudley vocoder, where speech analysis is performed by a set of parallel-band filters, is commonly called the channel vocoder, as opposed to the formant vocoder, where the resonances produced by the oral and nasal cavities are simulated by several tunable-formant or resonance filters. Being merely a speech synthesizer, instead of an analyzer/synthesizer combination, the formant vocoder is less interesting for music applications and therefore will not be discussed further in this article.

Basically, the channel vocoder can be divided into two main sections: the analyzer and the synthesizer, as shown in FIGURE 1. Each consists of identical filters covering a specific frequency range. The analyzer filters divide the speech spectrum into narrow bands, from each of which a voltage is derived. The voltage is proportional to the energy in the band.

In the synthesizer, a second group of filters divides the spectrum into the same narrow bands, and is followed by an amplitude-controlling device, such as a VCA (voltage-controlled amplifier or attenuator).

By connecting all analyzer control-voltage outputs to the respective VCA control-voltage inputs in the synthesizer section, the speech spectrum can be imposed upon a carrier

signal (a musical instrument, for instance) which is fed into the paralleled audio inputs of the synthesizer's filter bank. This creates the almost-classic "talking music" effect.

A similar effect can be achieved by the so-called "mouth tube" or "mouth bag." Here, the carrier sound is acoustically injected into the speaker's mouth by a flexible tube, and thus can be more-or-less articulated. Quite apart from unpalatable side effects, such as jaw-muscle spasms, this is a crude way of imposing speech upon another sound.

It should be made clear that the vocoder deals only with harmonic structure, and in no way is concerned with fundamentals, such as the pitch of the voice to be analyzed. Therefore, a circuit which can be very useful when using the vocoder as a voice synthesizer/processor is the pitch-to-voltage converter.

The only pitch-determining component is the fundamental of the carrier sound (the artificial vocal cords). When this pitch is changed, the pitch of the voice will change accordingly. When the pitch of the carrier is kept constant, a change in pitch of the real voice will only result in a different timbre of the synthesized voice. So, in order to generate complete, melodic speech, a device will be needed to change the pitch of the carrier in accordance with the pitch of the real voice. Such a device could be a pitch follower/extractor, as it is generally called. By changing the ratio between real-voice pitch and conversion factor, either linearly or non-linearly, many unrealistic but interesting artificial voice effects can be obtained.

## THE ANALYZER

In the Syntovox 221 vocoder, which consists of 20 analyzer and 20 synthesizer channels, each analyzer channel can be split into three subsections: a band filter, a full-wave rectifier, and a low-pass filter with an LED readout. The combination rectifier/low-pass filter is also known as an amplitude demodulator or envelope follower. The analyzer section is shown in FIGURE 2A. The number 20 is arbitrary, and not a scientifically-determined minimum or maximum. (In the case of the Syntovox 221, it was partly inspired by the available standard matrix format chosen to interconnect the analyzer and synthesizer.)

If the vocoder is to be used exclusively to reconstruct intelligible speech, 15 to 16 filters are sufficient to cover the necessary frequency range from about 100 Hz to 3 kHz, with a resolution of 25 percent, which gives us a one-third octave filter spacing. This will provide a fairly-accurate picture of the speech spectrum. However, extending and subdividing the total frequency range beyond the 3 kHz limit will add more definition

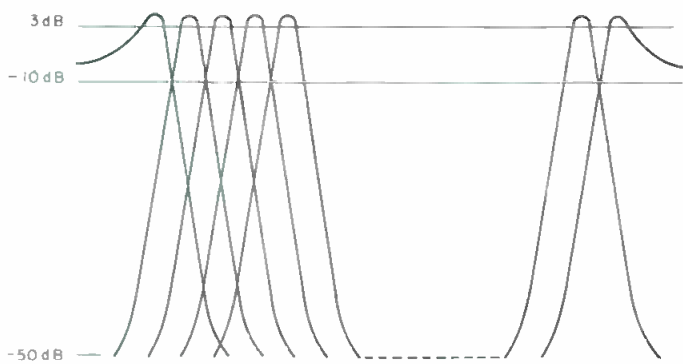


Figure 3. The response of the filters in the Analyzer and Synthesizer sections.

### VOICED/UNVOICED DETECTOR

The analyzer/synthesizer sections provide a straight-forward vocoder system, with which fairly-intelligible speech and good musical results can be obtained. However, the addition of a voiced, unvoiced detector will add even more intelligibility and clarity to the synthesized vocal effects. FIGURE 4 shows a block diagram of such a detection system, to which many variations and additions can be made. The detector can discriminate voiced and unvoiced sounds by continuously comparing the energy in two different frequency bands. One band is 30 to 800 Hz, and the other is 2 kHz and up. The decision of the detector is based on the assumption that voiced sounds have less high-frequency energy than unvoiced sounds, which fortunately is true for almost all speech sounds. Almost, because composite sounds such as "V" and "Z" can create instability of the detector when pronounced over a long period of time. Under normal speech circumstances, the detector, responding very rapidly to changes of the energy in these two bands, will switch alternately between voiced and unvoiced, and neither the indecision nor the transitions will be audible.

In order to extend versatility in the electronic music studio, and with computer applications in mind, the detector should be equipped with control inputs and outputs, as well as an inhibit input when the vocoder is to be used with other triggering devices.

Another possibility to add intelligibility to the vocoder effect has been provided by Harold Bode, who patented the clever solution of adding the high-frequency end of the real speech spectrum to the vocoder output signal. FIGURE 5 shows the basic setup (U.S. patent 4,158,751).

To provide the necessary high-frequency spectrum for fricatives and sibilants without using an expensive detection system, noise may be added to those synthesizer filter channels that are above 2 or 3 kHz. The noise is constantly present, and it will feed through to the output when high-frequency components in the voiced spectrum open the high-frequency

Figure 5. The dashed-line path improves intelligibility by adding high-frequency audio components to the vocoder output.

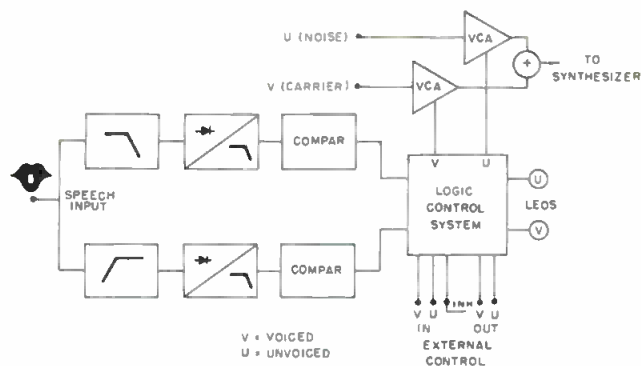
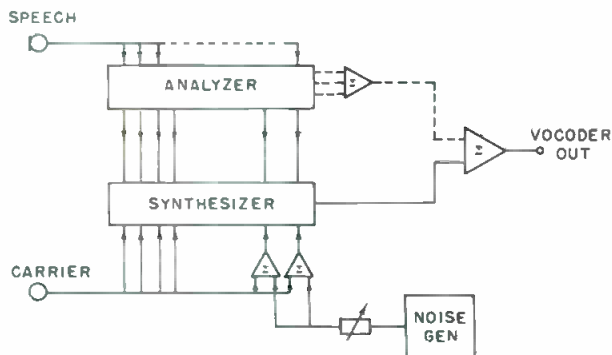


Figure 4. The Voiced/Unvoiced Detector.

VCAs in the synthesizer. In general, this solution offers better intelligibility of the synthesized voice effects.

### CARRIER SOUNDS—THE ARTIFICIAL VOCAL CORDS

The carrier sound, which will replace the signal normally produced by the vocal cords, is subject to certain conditions which will be discussed later on. For speech synthesis, it is necessary to generate a carrier signal similar to that of the vocal cords, as shown by the functions in FIGURE 6.

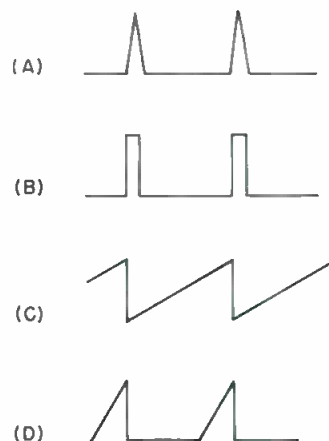
These waveshapes each give a different speech result because of their respective spectra. A very narrow pulse signal will provide a synthetic voice of piercing, rattling quality, because of its very strong high-frequency spectrum. A sawtooth and a spaced-sawtooth will give a more pleasant, mellow-sounding voice quality. The pitch generator should be controlled by external sources, such as pitch followers, envelope followers, low-frequency generators and random generators. All these control sources can take away much of the static, machine-like impression that vocoder speech usually makes. It is an interesting experiment to note that a voice will sound less artificial and even more intelligible when the pitch is modulated, instead of being kept constant.

When using external carrier sources, such as musical instruments, an automatic bypass circuit can be of great help to bypass or cross fade the carrier to the output when no analyzer input signal is available. Such a circuit has been incorporated in the Syntovox 221, and is called a "fill-in." Other manufacturers supply similar facilities, labelled "silence bridging" or "pause stuffing."

### CONTROL VOLTAGE PATCHING

One aspect of the vocoder still to be explained is the interface between the analyzer and the synthesizer. As explained previously, and shown in FIGURE 1, all control voltages generated by the analyzer section have to be fed to the control voltage inputs of the gain cells in the synthesizer section.

Figure 6. Various carrier wave shapes which may be used to simulate vocal-cord sounds.



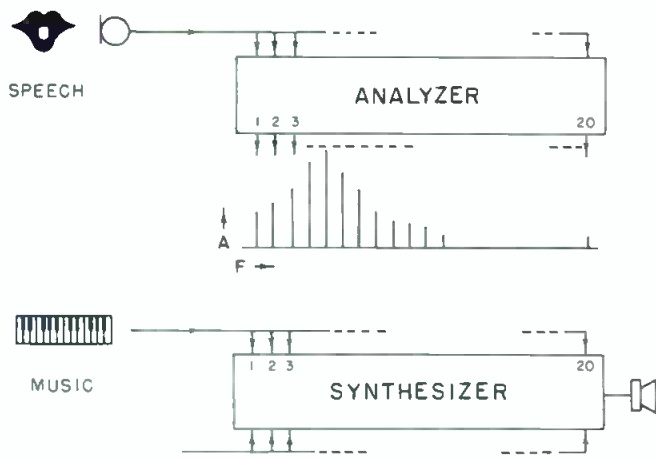


Figure 1. The basic building blocks of the vocoder.

and clarity, specifically where fricatives and sibilants are involved.

An ideal filter would exhibit a flat response within, and infinite attenuation outside, the pass band. In the Syntovox 221, a compromise has been found by giving the filters a relative bandwidth which is narrower than their one-third octave spacing, in order to achieve as little overlap as possible. Moreover, the filters—which are eighth-order—have a very rapid roll-off of 48 dB-per-octave, which increases to 54 dB-per-octave over the first octave.

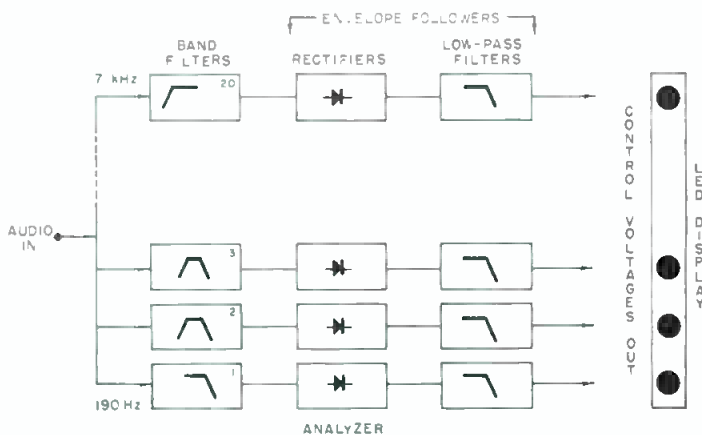
Although the dips in-between the narrow-band filters will affect the true response of the input signals to both analyzer and synthesizer, they greatly improve intelligibility and effect.

For designers of vocoders and speech synthesizers, it is unfortunate that the human ear is very sensitive to phase relations between spectral components. Frequencies, approximating the resonance frequency of a filter, will be subject to substantial phase shift, causing a different timbral perception, even though amplitude relations within the spectrum have been hardly affected. This phase shift side-effect will be more dramatic the higher the order of the filters, thus creating the unnatural speech effect that is a characteristic of the vocoder.

As with most man-made contraptions, the vocoder is full of compromises. But reducing the filter slopes, and thus reducing phase shift effects, would not result in a better vocoder. Large, overlapping areas create a blurred speech response, along with a poor effect response when the vocoder is used in music.

One more compromise in the analyzer section is the low-pass filter following the full-wave rectifier. In order to obtain rapid response of high frequencies and transients, this filter must be as fast as possible. On the other hand, a large ripple margin will definitely create intermodulation effects, which can be very disturbing.

Figure 2A. The Analyzer section.



The problem can be more-or-less resolved by adapting the cut-off frequency of the low-pass smoothing filters to the audio pass-band of the preceding analyzer filters. By choosing a roll-off point about ten times higher than the center frequency of the audio filter, an optimum can be found with respect to the phenomena of transient response and intermodulation as explained previously.

Finally, the LED control voltage read-out array was not only applied for its anticipated appeal—an expectation which has been met—but because it definitely has the function of displaying the spread, and to a certain extent the amplitudes, of the spectral components in the analyzed signal.

To cover the whole audio frequency range with a set of 20 filters, spaced at one-third octave intervals, would not be possible. Since the vocoder is only concerned with the harmonic structure of speech in the first place, it is not necessary to maintain this resolution either, so by designing the lowest and highest filters as low-pass and high-pass, the total audio range can be covered from about 20 Hz to 18 kHz.

## THE SYNTHESIZER

A description of the synthesizer filter bank will be easier, now that it is clear that the audio filter section is a replica of the analyzer filter bank. The only practical problem within a certain budget is the fact that both sections should be exactly the same, which means that only a small tolerance in component spread can be allowed. Any deviation from the analyzer center frequencies will cause unwanted formant shift.

In FIGURE 2B, the basic layout of the synthesizer section is shown. FIGURE 3 illustrates the performance of the filter bank in both the analyzer and synthesizer sections. The carrier sound is split up by the synthesizer filter bank and these signals, after being processed by the VCAs, are summed, and become the vocoder output signal. All control voltage inputs of the VCA bank are supplied with an attenuator followed by an LED readout. These attenuators can be used to “equalize” the vocoder effect.

The trickiest part of the synthesizer section is the voltage-controlled gain device following the filter. In order to copy the spectral image in accordance with the control voltage delivered by the analyzer, each synthesizer filter is followed (or preceded, as sometimes is done) by a voltage-controlled gain device allowing spectral components of the artificial vocal cords (the carrier) to pass through to the output of the vocoder. Such a gain cell could be a VCA, an OTA (operational transductance amplifier) or an electronic switch controlled by a pulse-width modulation (PWM) system.

The PWM's amplitude control is achieved by simulating a variable resistor in the audio path, by means of an electronic switch controlled by a high-frequency generator. This switch is alternately opened and closed, either by a narrow pulse whose repetition rate can be voltage controlled, or by a voltage-controlled duty cycle rate.

Figure 2B. The Synthesizer section.

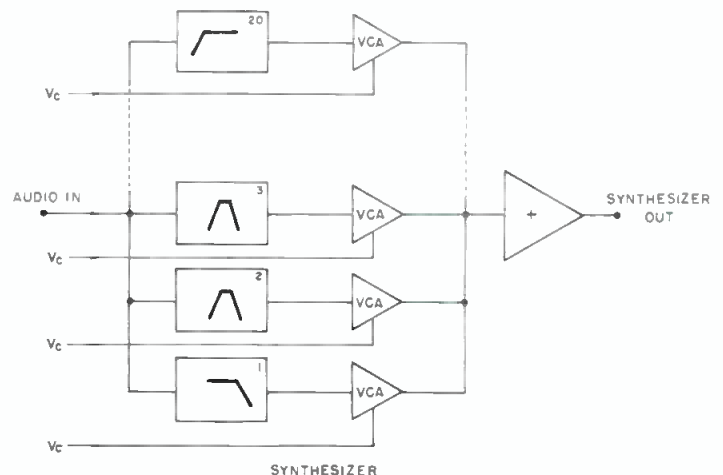




Figure 7. A 20-x-20 matrix connects the analyzer and synthesizer sections of the vocoder.

One of the advantages of not connecting them permanently is that both sections can be used to control, or be controlled by, external equipment. Also, this creates the possibility of connecting the analyzer outputs to other than their respective synthesizer channels. The problem of choosing a way to provide this facility is purely practical, and few manufacturers of larger vocoder systems have gone to this trouble. One way of providing a compact and versatile patching system is the matrix, which allows optimum freedom of routing control voltages. In addition, all outputs and inputs of the analyzer and synthesizer

should be available for use in electronic music studio setups.

FIGURE 7 shows the matrix which connects the Syntovox's analyzer and synthesizer sections. In FIGURE 8, all sections discussed so far are shown in a complete block diagram.

One of the most obvious applications of the matrix is formant shifting, which means that formants (typical resonance peaks) are transposed to frequency areas other than where they originate. This feature makes the vocoder an interesting instrument with which to generate different types of voices. Though it is often suggested that the sex of the voice can be changed by shifting formants and raising pitch, this is not true. However, the character of the voice can be changed dramatically, and a certain touch of different age can definitely be achieved.

### APPLICATIONS

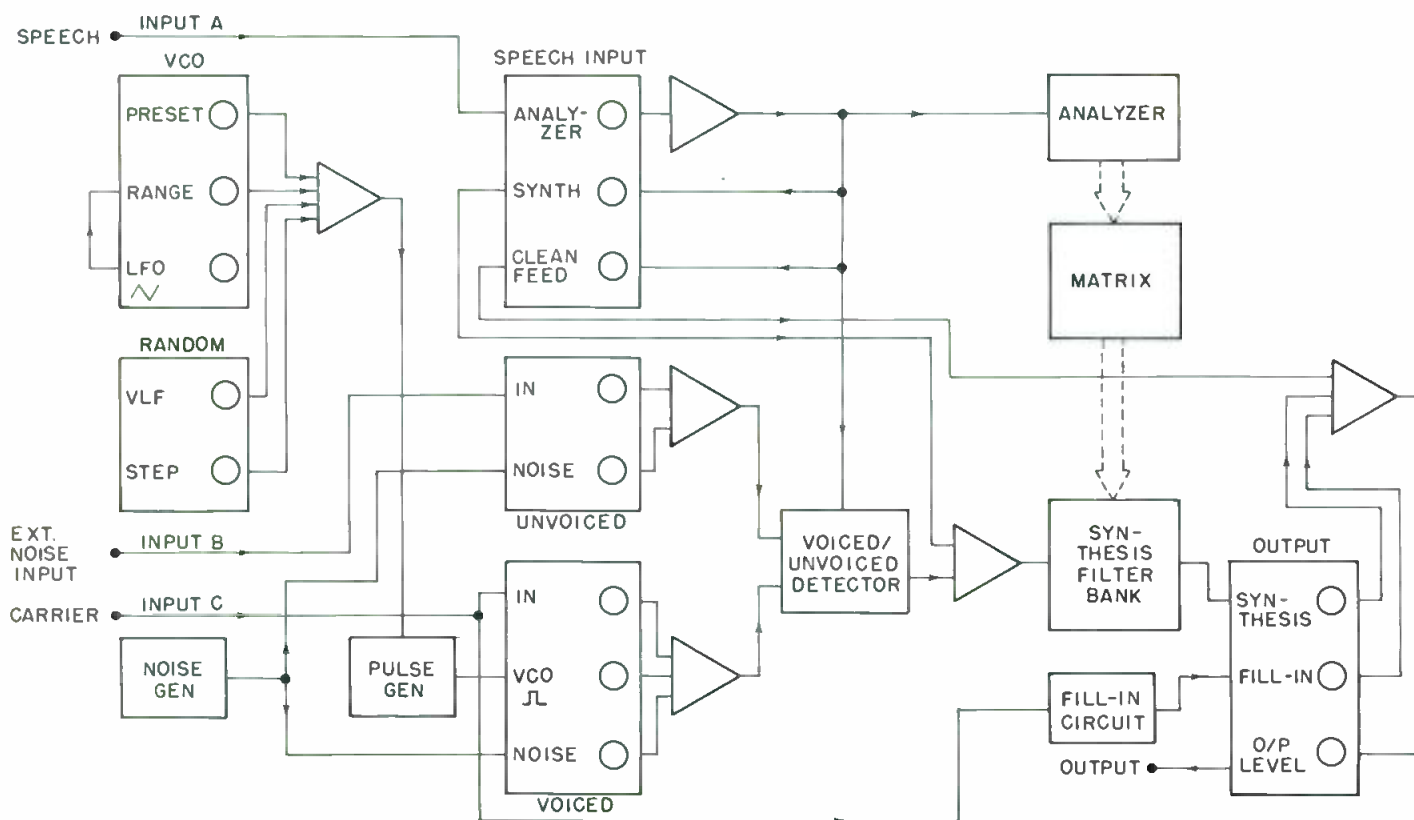
When marketing vocoders, one of the main problems encountered is that it is a typical two-input device. The significance of this is not always fully appreciated. In any demonstration, it is almost impossible to show the versatility of the vocoder, since every carrier creates its own effect, and every modulator creates a different effect with the same carrier.

The only thing which can really be made clear to interested people is some basic information on intelligibility, response, and general sound quality.

Japanese manufacturers have appreciated the two-input dilemma, and equipped most of their vocoders with some kind of keyboard instrument, producing organ-like or string-like sounds. They too must have realized that keyboard players belong to the happy few who can play and speak or sing at the same time. However, this conception can easily lead us to the conclusion that a vocoder is a typical keyboard instrument extension, which it isn't. Though it is true that in many cases the voice/keyboard combination gives the most recognizable response within a relatively short setup time, the vocoder makes it possible to use almost any sound source to modulate any other sound source.

There are restrictions, especially concerning frequency spectrum and synchronism between sound sources. In order to

Figure 8. A simplified block diagram of the Syntovox 221 Vocoder.



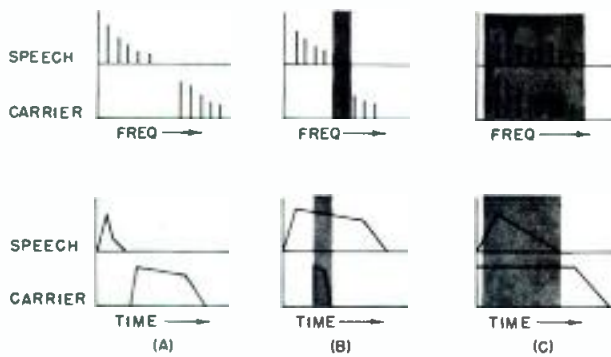


Figure 9. Optimum vocoder performance requires an overlap of spectral and temporal content. In other words, the carrier signal must be harmonically rich, and occur concurrently with the speech input signal. (A) The conditions shown will produce no effect. (B) The two narrow shaded areas indicate a slight effect, which will occur during a brief time interval. (C) Maximum effect is achieved, due to a complete spectral overlap, which occurs continuously.

get the best effect, it is necessary that both speech and carrier overlap spectrally and in time, which is graphically illustrated in FIGURE 9. A practical situation may make these examples more clear: a poor effect—or no effect at all—will be your reward for trying to modulate the sound of a bass drum with that of a flute, or vice versa. The vocoder user will also get into trouble when trying to impose speech upon a sine wave, or other pure sound.

Problems are bound to come up when modulator and carrier are of short duration and not in perfect sync. A recorded guitar track, producing a sequence of rhythmically limping chords, is not the right carrier to start with, unless the owner of the voice who is going to try to modulate the guitar sounds is alert enough to limp along in sync. Therefore, when a vocoder is used with previously recorded tracks the user may be disappointed with the results. The best way to create interesting vocoder effects is in a live situation, when there is artistic feedback between carrier and modulator. Truly, this is one of the most attractive sides of the vocoder. Being a real-time instrument, it can be used to instantaneously voice-control the timbral quality of sounds produced by electronic music instruments.

Here, the nature of the vocoder touches one of the marketing problems. Due to its complexity, the vocoder is often classified as a piece of high-technology hardware, to be sold to recording studios. On the other hand, the vocoder demands a thorough—almost musical—training of its user, which would make marketing aimed at the music industry seem the most logical way to go.

A sensible way to outline the possibilities of the vocoder may be by placing them in different fields of applications. In the first place, there is the scientific sector, where the vocoder can be a suitable instrument in speech or phonetic research. The effect of transposing formants can be simulated easily by connecting analyzer outputs to different synthesizer control inputs. Vowels can even be inverted by making partial cross patches on the matrix. A very interesting project at the Utrecht State University in Holland tested to what extent speech intonation is important to the intelligibility of the message. This was done by sampling real speech at a 10 kHz rate, and by changing the pitch of the synthesized voice under computer control, which allowed the intonation patterns to be varied every 2 milliseconds. Another purpose for which the vocoder can be used is speech training by producing properly-pronounced vowels and comparing the trainee's pronunciation of them. This can be achieved under computer control by digitizing the analyzer output information and comparing it to digitally-stored information.

A very appropriate application of the vocoder is in the field of animated film and sci-fi productions, making different alien voices by shifting formants and modulating the pitch generator in several ways. Since it is relatively simple to create numerous voice characters with the vocoder, it is strange to note what



Figure 10. The Syntovox 221 20-channel electronic effects Vocoder.

troubles sound engineers and producers go through, making their movie character voices with the help of ring modulators, filters, envelope followers and shapers, almost desperately trying to avoid the use of the vocoder.

It can be useful to process a guitar through spectrum-enriching effect devices such as boosters, fuzz boxes, doublers, etc. The same applies to other instruments generating more-or-less pure sounds. Once the carrier is rich in harmonics, it is possible to mold it into the timbral shape of vocalized sounds, sung or spoken into a microphone connected to the analyzer input. Now, dynamics and color are under complete vocal control and can be modified instantaneously due to the real-time nature of the vocoder.

On-stage use will require certain precautions, depending on acoustic conditions. Due to the phase shift introduced by the sharp filters, the vocoder can be very sensitive to acoustic feedback. Also, background noise can trigger horrifying effects. The classic trick with two reverse-phased microphones can be helpful, as well as directional, close miking. When close miking, attention should be paid with respect to inherent low-frequency boost, and when the microphone does not have an internal roll-off filter, conditioning with an external equalizer or by means of the attenuators at the synthesizer control inputs will be necessary.

Applications in the electronic music studio are almost unlimited, as long as the vocoder is an "open" system, meaning that the user will have access to everything controllable inside the unit. With the help of the analyzer section, frequency-dependent triggering or modulating effects can be obtained. A frequency-controlled band-pass filter is easily realized by sweeping a sine wave through the analyzer band. Interesting percussion effects can be generated by applying random sequences of sine bursts to the analyzer and by feeding pink noise into the synthesizer filter section. Depending on the positions of the patch pins in the matrix, intricate rhythmic structures by a multitude of percussive instruments can be obtained. In all, the vocoder can be used for a long list of effects which do not at all resemble the classic vocoder effect.

At least for the vocoder, the expression, "Your imagination is the limit," is not hype. ■

#### References

- "Analog Speech Encoder and Decoder"; Harald E. W. Bode; U.S. Patent No. 4,158,751; filed June 19, 1979.
- "Fundamentals of Speech Synthesis"; Homer Dudley. *Journal of the Audio Engineering Society*, Volume 3, No. 4, 1955.
- Phonetics* (Fonetiek); Dr. L. Kaiser; 1964.
- The Use of the Elektor Vocoder*; Felix Visser; Elektor, September, 1980.
- Vocoders Today*; Felix Visser; Elektor, December 1979.