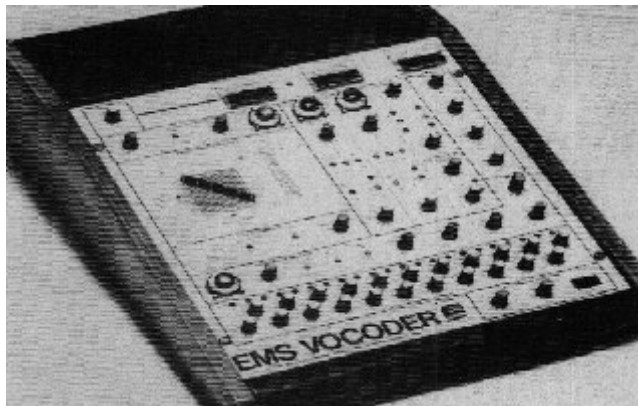


EMS vocoder

EMS Vocoder

Dropbox

Published in ? date ? by Nik Condron and Hugh Ford
REPRINTED FROM Studio Sound, JULY 1977



AN OPERATIONAL ASSESSMENT

THERE'S nothing new about vocoders-in fact they have been around since before the last war. Their function is to analyse the human voice and recreate it electronically. The voice is basically a complex sound generating device, and consists of a frequency and amplitude-controlled oscillator known as the larynx, and set of tone filters, ie nasal cavity, mouth and throat. The first thing to do when designing a similar system is to take these individually simple devices and translate them into block schematic form. Thus a chain may be visualised whose components can be separately converted into discrete circuits-so the larynx becomes a vco coupled with a noise generator, the controlling source being (in synthesiser terms) a dc signal derived from a voicing detector. The final stage would be a multistage voltage - controlled filter bank. Gradually a new picture emerges: we now have in block form the basis for a simple voice synthesiser. What Tim Orr of EMS has done for vocoders is rather the same that Robert A Moog did for the synthesiser. The old vocoders were enormous rambling heaps of machinery, plugged together with a nightmare profusion of cables-the analogy with the early breed of synthesisers, such as the BBC used in their Radiophonic Workshop complex, is obvious. Mr Orr has conveniently packaged all the necessary circuitry into a single ergonomically viable unit measuring about 5 x 6 x 20 cm.

Operation

The EMS Vocoder, in order to produce a synthesised voice, must first of all convert

the input signal into readable information. The live or recorded voice to be treated is, in the first stage, routed via a filter bank. This filter-bank consists of 20 bandpass filters plus one high and one lowpass filter. These are spaced over an average vocal spectrum of 200 to 8k Hz. The analysing filter-bank is directly coupled through a patchbay to the synthesiser filter-bank from which the final synthesised signal is derived. In order to produce the final control voltages necessary to control the synthesiser filter-bank, the input signal must be converted into a control voltage that will command the oscillators. These will, in combination with any other non-speech input (if required), produce the end 'excitation signal' that is sent to the synthesiser filter-bank. The first voltage necessary is voice-pitch. This is produced by a device known as a 'pitch-extractor', which acts as a specialised pitch-to-voltage converter reading the glottal pulses of the speech input. It includes a 'quality' control enabling the pitch voltage to be exaggerated for special effects. The output of the pitch-extractor is fed to one or both of the two voltage-controlled oscillators available in the machine that provide a sawtooth signal. I believe there are plans to incorporate a squarewave facility into the circuit to provide different harmonic possibilities. The input signal is also sent to a voiced/unvoiced detector, which has the function of deciding whether the oscillator or noise generator should be used at a given instant in the excitation signal. Thus, the excitation signal is made up of four separate signals all of which pass through a master control unit. These four signals are as follows: the controlling signal from the voiced/unvoiced detector; oscillator output; noise generator output; and an external non-speech input. The latter facility is one of the main features of the EMS Vocoder. By using a speech signal and a second signal from the non-speech input, the Vocoder will literally encode any recorded sound with any speech sound-this is how the machine can create, for example, talking musical instruments. The Vocoder also incorporates other less important but very useful effects devices. Nearly all the vc signals can be replaced with externally derived command signals, and there is a slew/ freeze control that will sample at any given moment the output signal as a constant tone. There is also a frequency shifter linked to the main output mixer of the device.

Applications

I was able to use the machine in my studio for about two weeks, and this enabled me to get a pretty fair idea of what it will do in a studio situation-working not only with electronic music, but also conventional pop and spoken special effects. There is no question that it is a very fine piece of machinery, and its limitations are literally those of the operator. Like any complex piece of equipment, it takes a bit of getting used to, but the front-panel layout is straightforward and well thought out. There are meters to read input, excitation or non-speech, and output signals. Those fitted to the review machine had vu faces with ppm ballistics, which I found a bit confusing, but as this was only the prototype it's hardly important since the machine can be supplied with either. Each of the 22 filter input levels has an associated led which makes it possible to read very efficiently the signal processing. Leds are also fitted to the voiced/unvoiced detector and the mode of operation is visible at a glance. The machine is capable of modulating any two audio signals, given that one of them is a voice or falls within the same frequency range. The possibilities in a studio situation are infinite; given a multitrack tape this machine can be hooked-up through the desk during remix, and almost any signal can be combined with a speaking or singing voice. To get the machine to 'sing' in pitch takes a few minutes of careful tuning between pitch extractor and oscillator controls. On its own (without a non-speech processing signal) the voice quality can be changed at will. The whole quality of a lead voice can cover a range, in terms of

frequency and timbre, that almost exceeds human capabilities. On its own, the voice sounds synthetic-it is not possible to create a replica that sounds absolutely authentic because, like all synthesisers, the sound is too clean, too free from natural imperfections. (A cough sounds like someone talking whilst trying to gargle!) To simply encode a voice is of little or no value in practical terms. The purpose of the machine, however, is to combine two signals, and there are many things that can be done with a single voice in this context. Firstly, a voice can speak in a flat monotone with no sibilants-this is done by switching out the pitch extractor and noise signals. A variation of this is to use the noise generator alone, ie cutting out the oscillators, to produce a very realistic whisper. By using an external vc source such as a synthesiser keyboard, and connecting in two oscillators tuned a third or fifth apart, a very interesting plainsong sound can be achieved. A normal speaking voice reading a rather dull passage from a book can be made to swoop theatrically in an overexcited manner. Very interesting musical sounds can be produced by linking up keyboards or a fast-moving sequencer pattern and varying the degree of melody to voice. Taking any instrument from a multitrack tape-or even a group of instruments-and feeding it, for example, through a foldback line into the machine, makes it possible to instantly assess the feasibility of different combinations. Depending on musical patterns, combinations such as drums, organs and especially the bass guitar, can provide totally new sound dimensions through the Vocoder. If the machine is linked to a complex synthesiser, such as a Moog 3c, the tonal variations are endless. If the synthesiser is confined to the frequency range of the voice, the other 'normal' instruments will actually sound as if they are being played or processed by the synthesiser. Thus a Hammond organ, in conjunction with a fast-moving sequence on the Moog, will produce a sound that is obviously a Hammond, but being played by a lunatic virtuoso. It is possible to see from the above comments that any recordable sound can be made to talk, whisper, sing or shout. Combined with even a modest sound effects library, thunder, trains, animals, traffic, etc can be created that sound intelligibly human. The limitations of the machine are very few to all practical intents and purposes, but the major one is the price that stands at present at £10 500. Whether this will come down if the machine catches on as a commercial proposition, no one knows. There is certainly a demand for machines of this kind but not, I would have thought, as standard recording studio equipment. However, studios or workshops specialising in sound effects and electronic music will find the machine an exciting and challenging proposition-as would radio stations and perhaps universities who would wish to make an investment of this kind. It is my belief that in terms of all kinds of music synthesis, this machine will be the fore-runner of the final stage of musical technological development-and perhaps it is at this time that the question should be asked: Where do we go from here? Nik Condon

TECHNICAL REVIEW

VOCODERS are generally associated with scientificthe creation of synthetic speech for specialised purposes, and with the analysis of speech. However, for the purposes of these notes there is little point in delving into finer details of the vocoder, and STUDO SOUND is not really an appropriate place to analyse the scientific aspects of such a device. As has been pointed out, the main use of a vocoder in the studio is the creation of unnatural sounds rather than the analysis of sounds, be they speech or other sounds. In this context the review has not mentioned a number of special features of the EMS Vocoder, such as a computer interface, which may be of little immediate interest to studio engineers.

FIG 1/2 soon!

Likewise, these notes on the technical features of the Vocoder are aimed at its studio application as an effects generator, rather than its application as a scientific instrument. Foremost in studio applications are the possible problems of interfacing the Vocoder with other equipment, followed by noise performance and, to a certain extent, distortion. Out of a number of inputs there are two that are likely to be used for effects generation - the speech input and the excitation input - both of which have an associated input level control and peak level meter. The speech input and the two excitation inputs have associated input gain controls that control the input sensitivity for 'ppm 6' from a minimum of -11 dB (ref 0.775V) for the speech input and -7 dB (ref 0.775V) for the two excitation inputs, with the maximum input being effectively infinite. As is common with input gain controls that appear to be connected to the input socket, the input impedance varies with gain setting: the speech input varying from 7570 ohms at maximum gain to 10 600 ohms at minimum gain, and the excitation input from 5230 ohms to 10 560 ohms, both being an undesirably wide impedance variation. The available output level at the onset of clipping was +19 dB (ref 0.775V) with the ppm indication 6 corresponding to + 1 dB (ref 0.775V) output, thus providing a very wide margin for peaks. The output, like the inputs, was single-ended but had a very low source impedance, which is always desirable; I do feel, however, that in view of the large number of available input and output connections a floating configuration would be an advantage. Returning to metering, for some reason the sensitivity of the excitation inputs at 1 kHz was higher than the speech input at -7 dB (ref 0.775V) for 'ppm 6', but this is of little significance; however, the frequency response of the meters was alarmingly variable, and the calibration between marks on the poor side. It was pleasing to note that the meter ballistics gave an attack time of around 10 ms and a fall time of 2.5s, which gives a good indication of level. (Provided that one can accept the poor frequency response?) Checking the overall frequency response from the speech input to the Vocoder output at a level corresponding to 'ppm 6' shows that the response was satisfactorily flat, as shown in fig. 1. This also shows that the third harmonic distortion value was very low, the second harmonic being even lower. On the other hand the frequency response through the filters at the equaliser output is somewhat lumpy, as shown in fig. 2, which was made with all the filter gains at maximum. It will be noted that the response extends well above the centre frequency of the highest filter (7888 Hz); however, when this filter output is eliminated the frequency response falls very rapidly above 9 kHz. The noise at the output with the mixer inputs closed and the mixer output open was found to be -84.5 dB(A)-ref 0.775V-increasing to -82.5 dB(A) with the speech channel open, or -76 dB(A) with the vocoder channel open; all these figures are quite adequate. Generally it is felt that the performance as briefly reviewed here is more than adequate for studio use, but the large ripple in the filter outputs will obviously have a substantial effect upon the final sound.

Hugh Ford