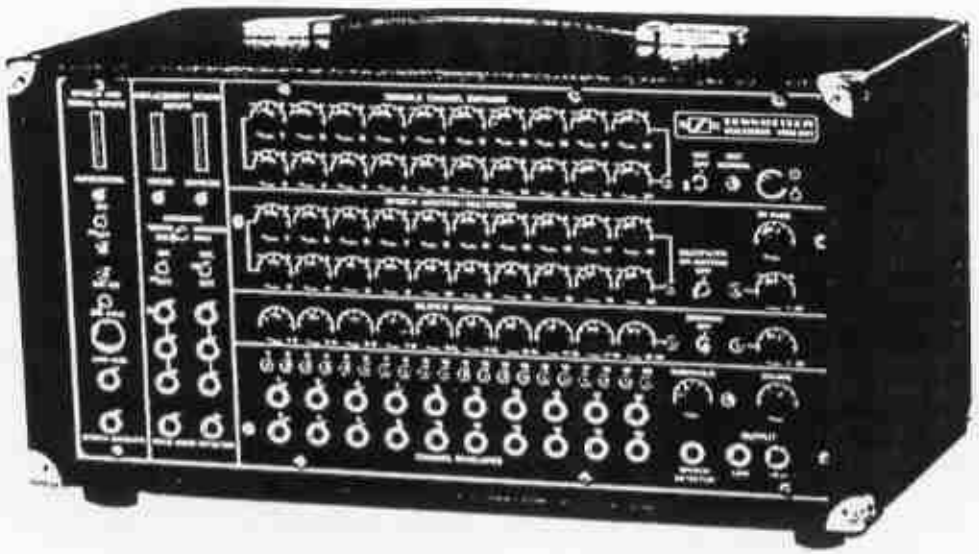
SENNHEISER

SOUND EFFECT VOCODER VSM 201



THE SENNHEISER SOUND EFFECT VOCODER VSM 201

In the continuing search for new acoustical impressions, a large number of electronic instruments has been created which are capable of generating unusual and unknown musical sounds, as well as sound effects. Now, the Sennheiser Sound Effect Vocoder VSM 201 affords the possibility of utilizing the human voice in a new dimension of sound extension.

The basic function of the Vocoder entails a spectral speech analysis which is fed into the circuit from either a microphone or a tape recording. Specific channel envelopes are formed and, at the same time, the Vocoder decides whether the speech contains voiced or unvoiced sounds.

The second major part of the Vocoder circuitry synthesizes a signal based on the characteristic channel envelopes and a replacement signal from a second external source. The result is a new synthetic sound effect with good diction and high intelligibility.

The Sennheiser Sound Effect Vocoder VSM 201 can be combined with all popular synthesizers and interfaced with usual studio equipment.

THE VOICE

The basic ingredient of a voice is the stream of air that the lungs push through the vocal cords, the pulsating movement of which is perceived as sound. Using tools such as the mouth, lips, tongue, and teeth, speech is formed "around" this matrix, composed or voiced and unvoiced sounds. Common speech is a rapid succession of both types of sounds.

The resonant cavities of the mouth and position of the tongue determines the so-called "voiced" sounds – the vowels A-E-I-O-U – and also parts of almost all consonants as well. The "unvoiced" sounds are formed by restricting the flow of air by means of the lips ("P" or "B"), the tongue ("T", "D", and "G"), or the teeth ("C", "V", or "Z"). Sounds such as "Sssssss" or "Shhhhhh" are completely unvoiced.

ENTER THE VOCODER

The Vocoder makes it possible to separate the elements of speech from the airstream and to substitute any audible signal for the air, thus allowing speech to be formed "around" a nearly infinite number of sounds.

It is obvious, then, that the Vocoder needs at least two inputs: a microphone, through which vowels and sounds are introduced to the "Speech Input," and the substitution for the airstream (whatever it is), which enters through the "Replacement Sound Input" and is made to appear "talking".

The harmonic frequencies of the human voice range from 75 Hz for bassists to 1000 Hz for trained sopranos, though the usual frequencies are 100-400 Hz as the "fundamental" with harmonics up to 4000 Hz.

The incoming signal at the Speech Input is analyzed in 20 channels of bandpass filters. The full range of the bandpasses is 100 Hz to 8000 Hz (the unvoiced portions ("Sssssss") comprise the highest frequency content). Twenty (20) LED's light up according to the strength (amplitude) in each channel, giving a visual display of use as a real-time spectrum analysis – an indispensable aid in learning how the speech input signal is processed and how to use the special features of the VSM 201 for optimum results.

Technically, the replacement signal is broken up into 20 channels as well, and in each channel is only opened as much as the corresponding channel of the speech

input signal happens to be. This presence or absence of special frequencies in the full harmonic spectrum is exactly what is perceived as vowels, just as the human resonating cavity (mouth) is nature's way of bandpass filtering.

This brings to mind the principle of the Talk Box: to form the vowels around, as in the famous case of Peter Frampton's guitar. A sentence in a Talk Box is reduced to its vowels: "Do you feel like I do" becomes "OO-OO-EE-I-I-OO." But the guitar is unable to produce anything of the unvoiced family (such as "D" of "Do") because, being a waveform with a definite pitch, it has no unvoiced quality. It is the same with the other consonants.

The Vocoder has capabilities far beyond these limitations.

Since speech demands different qualities in the replacement signal for voiced and unvoiced segments, the Vocoder's Replacement Signal itself has two inputs: anything with sufficiently high frequency content can replace the unvoiced, while the voiced can be replaced with any sound. The problem of deciding between the two, accurately and fast (as often as 10 times a second or more in normal speech), is solved principally by comparison: unvoiced signals range up to the 20th channel or 8000 Hz (voiced range only to 4000 Hz), so channel 19 is compared to a special channel Zero, which has a lowpass instead of the bandpass and covers lower frequencies. Depending on the result, that respective replacement is activated, and the vocoded output corresponds exactly to all phases and characteristics of the speech input, although consisting entirely and exclusively of the replacement input signal.

WHAT CAN BE DONE WITH IT?

The sky is the limit! There are quite a few obvious applications, but many more are yet to be discovered.

Known territory: put a guitar into the voiced replacement signal and talk through a mike into the speech input – the basic Talk Box. Advance by using the internal noise source, white or pink noise, to form all unvoiced parts, and all of the consonants can be articulated (see the next section: "The Vocoder and the Synthesizer").

If intelligible speech is the goal, a keen sense for detail helps: e.g., the peaks of a picked guitar probably should be clipped (compression). Take an outboard equalizer and shape a noise source into the external unvoiced replacement signal to match the "Sssssss" and "Shhhhhh" to the rest and again make good use of the special features as described in the next section.

Any sound can be used, via microphone, direct or recorded. Violins, bass, percussion – even football crowds – each can be modulated by the speech input, with radically different results. Of course, the unvoiced segments can be internal noise, but how about wind, thunder, jet planes, explosions or . . .? If one uses the guitar or other sound to form the "Sssssss" and talks through the noise as the voiced replacement, the result is the sound of a very sore throat, with generally weird effects and unheard-of timbres.

Nothing prohibits the use of another sound in place of voice in the speech input. Any instrument or any sound has resonances (though not A-E-I-O-U) and these natural resonances do affect whatever is the replacement signal. The result is not speech, but incredible multimodulations and previously non-existent audio signals.

One of innumerable experiments: speech input a miked cymbal, replacement the simple internal noise. The result is a strange phasing-type swoosh sound, varying with location, motion, type of mike, the way the cymbal is hit (and with what), and of course the kind of replacement. Just substituting the unvoiced sources at hand provides endless variations: the external noise could be flanged, phased or otherwise modulated even before entering the Vocoder and so could the cymbal – which, again, could be substituted with anything audible. If, for instance, one uses an oscillator as speech input signal and sweeps it through its range, say 20-10,000 Hz, then any sound at the replacement input is swept through its frequency spectrum, just like an infinite notch, one would have a distinct, and if one so desired, distinctly different, flanger.

Most applications will not be known until they have been tried, but imagine the sound of breaking glass modulating gongs . . . or percussion on an orchestra . . .screams through tearing paper . . . or everything vice versa . . . or just two instruments multimodulated. We have discussed how the airstream is replaced . . . how about taking one person's airstream (saying "Aaaaaaa") and another's voice and speech? If carefully adjusted, the results can be very effective: the man with the female voice, the child with its grandfather's voice, and many other unnatural possibilities.

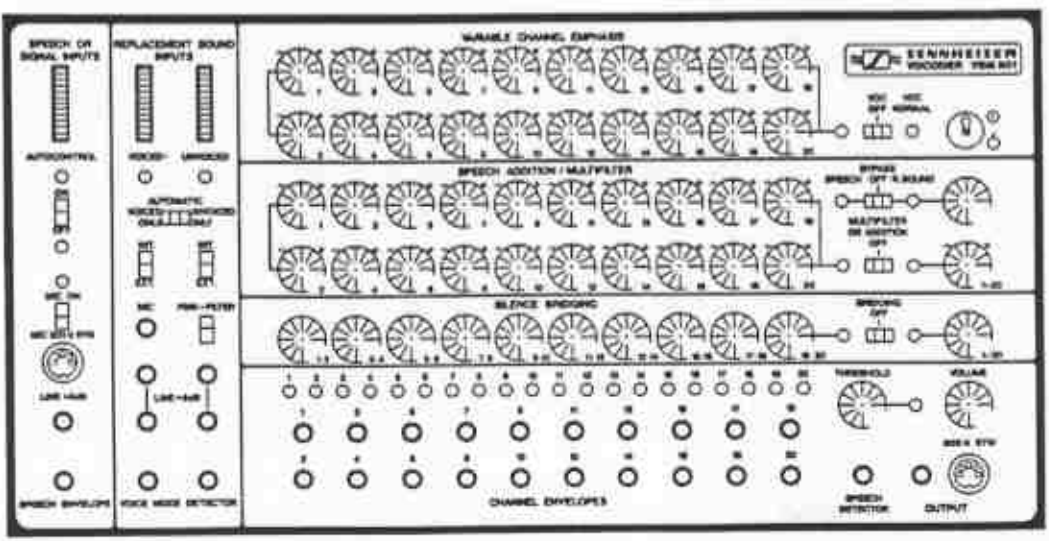
THE VOCODER AND THE SYNTHESIZER

Of the myriad sounds possible, one particular family holds enormous potential: the Synthesizer. Although it is an electronic device, the Vocoder is not limited to electronic sound sources, and by its nature is absolutely multi-para-polyphonic. But in case one does use a synthesizer, extra benefits result.

The basic source, of course, is one or several oscillators. If a key is hit, a constant pitch of one of the common synthesizer wave forms enters the voiced replacement signal input. Adding simple internal noise for the unvoiced parts, and talking into the microphone, the result is "instant robotness," or what is commonly believed to be the sound of a talking computer: constant pitch. The initial raw signal determines if the characteristics are smooth, metallic, or harsh; individual taste is the rule.

The advantage of an oscillator is that all frequency modulations in the full eight - octave range are easily accessible: deep Darth Vader-type voices or high-pitched soprano or mice talking – all are capable of being produced from any spoken word. Of course, one could play along on the keyboard and the constant pitch becomes a true singing voice, or if one used a 2-4-8 voice synthesizer, the result would be instant choir. Any amount and speed of vibrato or tremolo or filter sweeps or "sample and hold" can be treated to the desired effect. The more refined applications make use of complex patches or additional devices. A frequency shifter, for instance, in combination with a reasonably high-pitched oscillator makes anyone capable of Donald Duck speech without straining muscles. If one's needs include alien creatures or monsters making the usual mistake of speaking English, one does not have to fake it with a bit of noise and an inadequate voice, nor does one have to spend weeks (or months, as in some famous examples) fiddling with tape machines and other devices.

Another advantage of the synthesizer over, say, an organ, is that there are provisions in the Sennheiser Vocoder for a "Feedback;" the results of a vocoding process can be patched into the source of the sound again, affecting the next step of operation.



FEATURES AND SPECIFICATIONS

The VSM 201 has 3 gates and 21 control voltage outputs that can be utilized in any system. In particular, the panel mesurements are exactly those of MOOG studio series modular synthesizers, so it fits right into any 35, 55, etc., cabinet.

The synthesizer interface outputs have a speech detector gate, switching from 0.5 Volts to 11 Volts as soon as a certain threshold between pause and speech is reached. The trip point is continously variable, with an LED indicating speech. The result of the comparison between voiced and unvoiced segments of speech input signal is gated out in two jacks, with 11 Volts for the active side and 0.5 Volts for the other. The strength or amplitude of the speech input signal is analyzed in an envelope follower and a proportional control voltage of 0-7 Volts can be patched out. After the division into 20 channels each of these channels has its own envelope following output, 0-7 Volts. A variety of uses for these are self-evident, but many more are yet to be discovered. The main envelope follower or just parts of the spectrum can open or close filters, VCA's, change the frequency, or trigger events. Patching it into the FM control input drives the oscillator pitch higher, with increasing level at the speech input, resulting in old woman-like speech – quite a funny effect. The only other outputs now are the two audio outputs, either 6dB, 100 kOhms unbalanced or 2.5mV, 200 Ohms balanced.

The speech input has a 6dB line or 200 Ohms microphone input with an On/Off switch and LED for Mic On. Amplitude of the incoming signal is displayed on an LED bargraph with 10 green and 2 red LED indicators. There is an auto-control that can be switched on, limiting the gain automatically and preventing overload, with a green LED lighting when these peaks occur. The replacement sound input actually has 5 inputs: 2 summing line-in 6dB for unvoiced and 2 for voiced plus a microphone input. A three-way switch decides if the replacement consists of voiced or unvoiced or, in the middle on automatic, it decides back and forth as described earlier. Two green LED's show the decision. Both sides can be switched to internal, replacing unvoiced segments with a selectable white or pink noise source and voiced with a simple constant pitch pulse oscillator.

There is a 20 channel graphic equalizer, boost and 100% cut, in which each channel of the vocoded signal can be individually adjusted. Keeping in mind that the speech input signal opens specific frequencies of the total spectrum of the replacement signal, the spoken vowel "A," which is dominant in channels 6-11, does not activate the higher harmonics of the replacement signal, even though they may exist. To achieve the bright end and intelligible speech it might help boosting channels 12 to 20; in other cases it might be more helpful to cut away specific bands or restrict the whole operation by cutting out all but a few channels 100%.

A three-way switch determines if the 20 channel EQ labeled "Variable Channel Emphasis" is "On," "Off," or at the "Vocoder Normal" setting, equalling mid-position of all potentiometers. There is a green LED for this and a red one for the EQ. For the optimum effect, there is another 20 channel EQ, providing the possibility of adding fractions of the normal speech or whatever is the speech input signal to the vocoded result. This circuit is labeled Speech Addition/Multifilter because it acts as a separate 100% cut 20 channel (almost third octave) Equalizer. This stage adds part of the speech input after the vocoding process to the final audio output. There is, again, a three-way switch for "On," "Off," or one potentiometer cut or boost for all channels at the same time, with red LED's for both "On" positions. Both speech and replacement (voiced) signals can also be directly bypassed -- the potentiometer controls the amount for one of the two or can be switched off.

A feature special to the VSM 201: In normal operation, the replacement signal is only audible when the speech input is present, giving no output signal at all for the pauses of a voice. In special effects it is necessary to fill those pauses with the unaltered replacement signal in order to identify the otherwise "talking" sound. Proportional to the speech level, a pause-filling signal in selectable amounts is added for silence bridging. The continuously variable silence bridging can be set with one potentiometer for two channels each (10 channels) or with one potentiometer for all 20 channels at once. Trip point of the logarithmic operation is determined by the threshold setting.

Although conceived for the most demanding task – intelligible speech under almost any circumstances – the Vocoder's special features are capable of many exotic effects of their own. As with any sound modification tool, the ingenuity originates in an imaginative mind, rather than in the hardware. The Vocoder VSM 201 is not a singing voice one buys and just turns on. It takes careful and patient use. It is by no means a shortcut to talent, but it can supply an exciting new dimension of sound for the creative recording or sound engineer.

If you are interested in purchasing or renting the Sound Effect Vocoder VSM 201, or have any questions, please contact **Sennheiser Electronic Corporation, 10 West 37th Street, New York, New York 10018, (202) 239-0190.**