VOSIM—A New Sound Synthesis System*

WERNER KAEGI AND STAN TEMPELAARS

Institute of Sonology, University of Utrecht, Utrecht, The Netherlands

VOSIM (VOice SIMulation) sound synthesis is based on the idea that by employing repeating tone-burst signals of variable pulse duration and variable delay, a sound output of high linguistic and musical expressive power can be obtained. Conventional means such as oscillator and filter banks were therefore entirely dispensed with, basically simpler apparatus being used instead: two tone-burst generators and a mixer. Starting with these rudiments, the aim is an optimal representation of the input data and a progressive application of logical calculus to the sound-synthesis process.

1. INTRODUCTION: In 1972 Werner Kaegi embarked on an investigation, the purpose of which was to obtain insight into the sign repertory on which communication via music and speech, the two chief communication systems that make use of acoustic signals, is based. The attention was primarily directed toward the linguistic sign repertory, one argument being that this repertory, although not completely investigated and familiar, is less obscure and easier to work with than the repertory of musical signs. An important point was that it is possible in this case to decide whether a series of basic signs forms a unit (word, sentence) belonging to the relevant language or not. This is far more difficult in the case of musical signs.

A second argument was that a synthesis method able to produce natural speech sounds would be flexible enough to be used in a sophisticated sound-synthesis system for musical use.

There is moreover an unmistakably close relationship between speech and music, meaning that insight into the linguistic sign repertory might lead to insight into the musical sign repertory.

Summarized briefly, the investigation was first aimed at developing a model with which it would be possible to describe and generate the Indo-European linguistic basic signs in a metrical fashion, and then to expand and adapt this model so that musical signs could also be approached in the same way. For this aim Kaegi started by carefully observing the waveforms of speech sounds and looking for "invariants": characteristics of these signals that are essential to the nature of the sign. Once these are known, it ought to be possible to reduce the signal to a more simple one without affecting its relationship to the sign. An important first result [1] was that it turned out to be possible to reduce the waveform of one period of the vowel-sound "a" to the simple form of a $\sin^2$ pulse (Fig. 1). This gives us the basic principle of VOSIM sound synthesis: to generate sounds exclusively by means of $\sin^2$ pulses of variable duration and variable delay. Signals of this kind will in the following be referred to as VOSIM functions or VOSIM signals.

2. THE VOSIM MODEL

Experiments were made with various other pulseforms, but the $\sin^2$ pulse finally became the point of departure for the VOSIM model albeit in a more complicated form: a series of N $\sin^2$ pulses of equal duration T, with decreasing amplitude (starting with value A) and followed by a delay M (Fig. 2).

The amplitude decrease does not occur exponentially, as in the case of natural systems, but in steps. This means that each pulse is a pure $\sin^2$ pulse with an amplitude that is a particular, constant factor b smaller than that of the preceding pulse. This was chosen because it is simpler in form than an exponentially damped series. It can be shown, moreover, that this approach has only negligible consequences for the signal's spectrum [2].

* Presented at the 53rd Convention of the Audio Engineering Society, Zurich, Switzerland, March 1976; revised November 1977.

The problem in generating signals with the complexity of speech and music signals is approached in the VOSIM model via the time domain. It is, of course, always possible to have a look at VOSIM functions via the frequency domain. This will be done later in this paper. At the moment we only mention the most striking fact occurring, that as well as a strong component corresponding to the repetition frequency f_0 of the signal [$f_0 = 1/(NT + M)$], the spectrum also exhibits a peak corresponding in position to the frequency of the $\sin^2$ pulses. There is consequently a "formant" area (the form depending on N and b), and this accounts for the applicability of the signal for the simulation of speech and music signals when this problem is approached via the frequency domain.

Experience and theoretical consideration have shown that the small number of only two synchronized VOSIM functions, added together by means of mixing, is sufficient for synthesizing speech sounds of probably all Indo-European languages, and that on this basis richer resources of musical sounds can be synthesized. With only three synchronized VOSIM functions we were able to produce violin sounds.

T, M, N, A, and *b* are the primary variables. In order to obtain vibrato, frequency modulation, and noise sounds, *M* has to be modulatable, either sinusoidally or randomly. This leads to the following three variables: *S* (the choice as to sine or random), *D* (the maximum deviation of *M*), and *NM* (the modulation rate expressed in the number of periods in which a complete modulation cycle has to be performed). Finally, four more variables were introduced for the description of transitional sounds: *NP* (the number of periods) and *DT, DM*, and *DA* (the positive or negative increments of *T, M* and *A*, respectively, in the given number of periods *NP* via linear interpolation) [3].

The complete list of variables is therefore as follows:

T (μs)	pulsewidth
DT (μs)	in- or decrement of *T*
M (μs)	delay
DM (μs)	in- or decrement of *M*
D (μs)	deviation
A	amplitude of first pulse (511 = 10 volts)
DA	in- or decrement of *A*
b (%)	attenuation constant
N	number of pulses per period
S	type of modulation (1 = sine, 0 = random)
NM	modulation rate
NP	number of periods.

Based on this refined VOSIM model, a minimum description was aimed at for the description of the linguistic signs; this means that of the (eventually two or three times) 12 variables, those were traced whose value determines the relevant sign, and their value ranges were determined. The minimum description has been completed and parts of it were published [1], [3]. The investigation into musical signs has begun in the meantime; the description for a large number of musical instruments is already finished.

Here are some examples [data lines (1)–(14)]. They are not only intended to facilitate understanding of the text, but also to enable the reader to reproduce the described sounds himself. For the sake of simplicity, our choice of examples involves the use of only one "generator" each time. The input dimensions are now arranged on a line from left to right. Each line accordingly describes a state of the sound to be described. By joining several lines together we can describe sequences of sound states (such as attack, steady state, decay, stop) and also sequences of different kinds of sounds, etc. The VOSIM model thus makes it possible to describe the development of sound in time:

T DT M DM D A DA b N S NM NP.

a) Vowel /a/ (as in Italian "Rom*a*"). Frequency 110 Hz ± 60 cents; frequency modulation (FM) speed 0.2 Hz; total duration 1000 ms. The pulse duration of 960 μs corresponds to 1041.7 Hz and acts as formant:

960 0 1410 0 655 511 0 75 8 1 550 110. (1)

b) Unvoiced fricative / ʃ / (as *Sh*anghai). Duration 240 ms. The deviation of the random modulation in the first

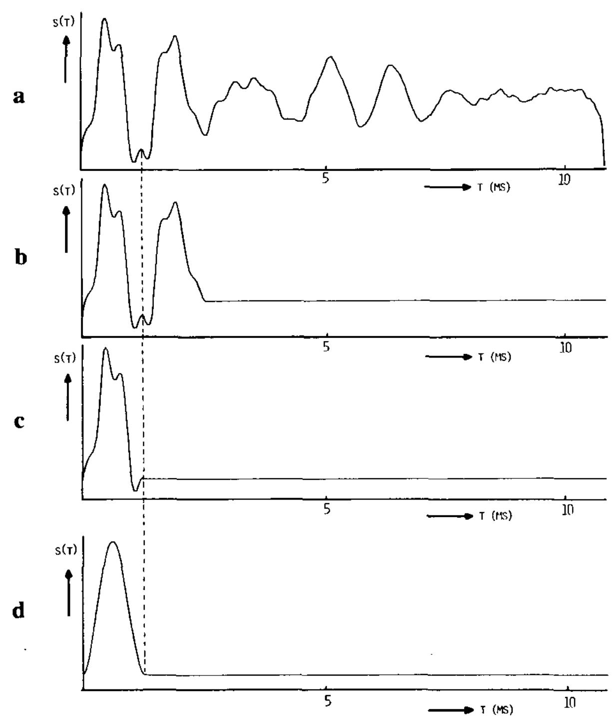


Fig. 1. Data reduction. **a.** Waveform of a very low spoken vowel "a" (92 Hz). **b.** Reduction 1. **c.** Reduction 2. **d.** $\sin^2$ pulse with T = 1.2 ms. **b, c,** and **d** also generate vowel "a."

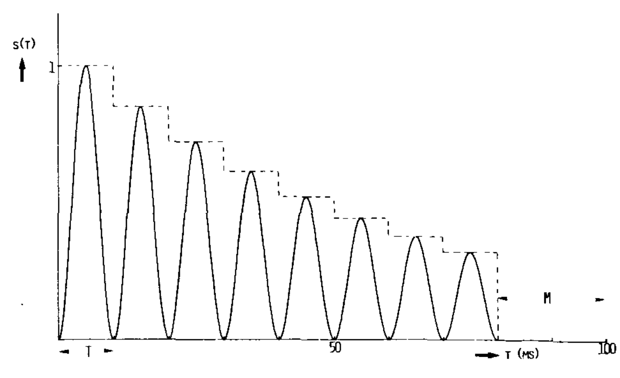


Fig. 2. The VOSIM time function. N = 8, b = 0.85, T = 10 ms.

data line satisfies the condition $2D = T = M$ (with $T = 200\ \mu s$ corresponding to 5 kHz), and generates the characteristic noise band of the fricative:

200	0	200	0	100	50	250	75	1	0	120	600.	(2)

c) Tone a″ of a soprano, with vibrato. Total duration 2006 ms; frequency 909 Hz; FM ca. 6 Hz; duration of attack 332 ms (first data line). See (13). The timbre is an /a/ vowel as in 1):

960	0	140	0	60	450	61	75	1	1	152	302	(attack) (3)
960	0	140	0	116	511	0	75	1	1	152	1522.	

d) Syllable /ʃ a/ (as *Sha*nghai). See (2) and (1). Total duration 1000 ms; frequency 110 Hz; FM deviation ± 60 cents, FM speed 0.2 Hz (second data line); duration of the fricative 240 ms (first data line):

200	0	200	0	100	50	250	75	1	0	120	600	[see (2)] (4)
960	0	1410	0	655	511	0	75	8	1	550	18	[see (1)].

e) Syllable /sa/ (as in French *Sa*bine). Frequency 220 Hz; other data as in d). Here the random modulation in the first data line satisfies the condition $2D = T = M$ (with $T = 80\ \mu s$, corresponding to 12.5 kHz):

80	0	80	0	40	50	250	75	1	0	120	1500	(5)
960	0	705	0	327	511	0	75	4	1	1100	36	[see (1)].

f) Syllable /fa/ (as in Italian Al*fa*). Frequency 329.7 Hz; other data as in d). Here the random modulation in the first data line satisfies the condition $D = M$ (with $T = 80\ \mu s$, corresponding to 12.5 kHz):

80	0	120	0	120	50	200	75	1	0	120	1200	(6)
960	0	1113	0	218	511	0	75	2	1	1648	54	[see (1)].

g) Tones c sharp″ (7), e″ (8), and a″ (9) of a clarinet in B flat. Frequencies 554.6 Hz, 659.6 Hz, and 880.3 Hz; total duration in all cases 1000 ms; attack 50 ms (first data line), decay 90 ms (third data line). In all three examples the condition $|NT - M| < 15$, that is, $NT \simeq M$ is satisfied; therefore only odd-numbered harmonics occur:

448	0	907	0	4	200	311	100	2	1	923	27	(attack) (7)
448	0	907	0	4	511	0	100	2	1	923	478	
448	0	907	0	4	511	−400	100	2	1	923	49	(decay)

376	0	764	0	3	200	311	100	2	1	1098	32	(8)
376	0	764	0	3	511	0	100	2	1	1098	568	
376	0	764	0	3	511	−400	100	2	1	1098	59	

284	0	568	0	2	200	311	100	2	1	1466	44	(9)
284	0	568	0	2	511	0	100	2	1	1466	757	
284	0	568	0	2	511	−400	100	2	1	1466	79.	

h) Tones c sharp″ (10), e″ (11), and a (12) of a harpsichord. Frequencies 554.6 Hz, 659.6 Hz, and 220 Hz: total duration in all cases 1000 ms; attack 10 ms; decay 20 ms; stop (fourth data line) 5 ms. The characteristic striking of the string is realized by incrementing T, from $T = 184\ \mu s$ in the first data line to $T = 252\ \mu s$ in the second line; the amplitude envelope plays a role too.

184	68	515	−476	0	500	0	75	7	1	5543	5	(attack) (10)
252	0	39	0	0	300	−200	75	7	1	5543	536	
252	0	30	0	0	100	−100	75	7	1	5543	11	(decay)
−5	0	0	0	0	0	0	0	0	0	0	0	(stop).

(A negative number in the first column indicates a pause duration in milliseconds.)

184	68	596	−340	0	500	0	75	5	1	6592	6	(11)
252	0	256	0	0	300	−200	75	5	1	6592	637	
252	0	256	0	0	100	−100	75	5	1	6592	13	
−5	0	0	0	0	0	0	0	0	0	0	0.	

184	68	1417	−1156	0	500	0	75	17	1	2200	2	(12)
252	0	261	0	0	300	−200	75	17	1	2200	213	
252	0	261	0	0	100	−100	75	17	1	2200	4	
−5	0	0	0	0	0	0	0	0	0	0	0.	

i) Portato of a soprano from a″ (909 Hz) to a′ (476 Hz). See (3):

960	0	140	0	60	450	61	75	1	1	152	302	(attack) (13)
960	0	140	0	116	511	0	75	1	1	152	1522	
960	0	256	1100	0	511	0	75	1	1	152	610	(portato)
960	0	180	0	150	511	0	75	2	1	76	762.	

j) The first nine tones from Stravinsky's score "The Rite of Spring" (high bassoon):

948	0	15	0	5	0	511	75	2	1	150	55	(attack) (14)
948	0	15	0	6	511	0	75	2	1	150	1200	
948	0	128	0	0	250	0	75	2	1	1	47	
948	0	15	0	0	150	0	75	2	1	1	49	
948	0	128	0	0	150	100	75	2	1	1	47	
948	0	655	0	0	200	0	75	2	1	1	116	
948	0	1138	0	0	200	0	75	2	1	1	98	
948	0	128	0	0	200	0	75	2	1	1	147	
948	0	376	0	0	300	−50	75	2	1	1	526	
948	0	15	0	0	255	0	75	2	1	1	206.	

[For (14) see Figs. 9 and 10.]

3. DESIGN OF EXPERIMENTS

Use was made of both analog and digital equipment for the experiments carried out in the course of the investigation. Analog equipment was used for a primary determination of the value ranges of the variables. This equipment is ideal, because of its flexibility, for a first approach, and what is more, VOSIM signals can be simply produced with standard generators. A special piece of apparatus was developed for the stepwise decay mentioned before [2]. The experiments were then continued systematically with digital equipment, which is of course indispensable when series of, and transitions between, VOSIM signals have to be generated. At first the signals were generated with software (programs VOSIM1 to VOSIM3) with the aid of a PDP-15 computer, but from the moment that the variables DT, DM, and DA cropped up, this was no longer possible. The number of calculations to be carried out was simply too high, which is why a digital oscillator was developed [4].

the sound to be described. Since relevant dimensions are noticeable for the invariance of their arguments, they can be represented by constants and/or terms. For the non-relevant dimensions, on the other hand, any value from the corresponding range of values can be used as an argument, and we consequently say that the variables of the non-relevant dimensions are tied by a universal quantifier $\forall$. We can now apply this procedure to the above examples (1) to (14), for instance, and in the new form they no longer describe individual particular cases of sounds (individual description) but their characteristic properties; that is, in general the sound (15) of a syllable /ʃa/ [generalization of (4)], (16) of a syllable /sa/ [generalization of (5)], (17) of a syllable /fa/ [generalization of (6)], (18) of a B-flat clarinet [generalization of (10)–(12)]:

T	DT	M	DM	D	A	DA	b	N	S	NM	duration (ms)	
200	0	T	$-N \cdot DT$	½T	50	250	75	1	0	120	240	(15)
960	0	$\forall$	$-N \cdot DT$	$\forall$	511	0	75	N^*		$\forall$	$\forall$	
with $M \leqslant 2T$ and $N^* = (10^6/F - M)/T$,												
80	0	T	$-N \cdot DT$	½T	50	250	75	1	0	120	240	(16)
960	0	$\forall$	$-N \cdot DT$	$\forall$	511	0	75	N^*	$\forall$	$\forall$	$\forall$	
80	0	120	$-N \cdot DT$	M	50	200	75	1	0	120	240	(17)
960	0	$\forall$	$-N \cdot DT$	$\forall$	511	0	75	N^*	$\forall$	$\forall$	$\forall$	
T	0	NT	$-N \cdot DT$	0	200	311	100	$10^6/2FT$	$\forall$	$\forall$	50	(18)
T	0	NT	$-N \cdot DT$	0	511	0	100	$10^6/2FT$	$\forall$	$\forall$	$\forall$	
T	0	NT	$-N \cdot DT$	0	511	−400	100	$10^6/2FT$	$\forall$	$\forall$	90	
with F being the pitch frequency, $284 \leqslant T \leqslant 476$,												
184	68	$\forall$	$-N \cdot DT$	0	500	0	75	N^*	$\forall$	$\forall$	10	(19)
252	0	$\forall$	$-N \cdot DT$	0	300	−200	75	N^*	$\forall$	$\forall$	$\forall$	
252	0	$\forall$	$-N \cdot DT$	0	100	−100	75	N^*	$\forall$	$\forall$	20	
−5	0	0	0	0	0	0	0	0	0	0	0.	

At any rate, the equipment's task is to apply VOSIM functions, described in the input dimensions, one to one in an $s(t)$ representation. The apparatus can therefore be regarded very generally as an operator P mapping the input V onto a VOSIM signal function $s(t)$:

$$V \xrightarrow{P} s(t).$$

The model's goal is to describe VOSIM functions in an optimal representation with regard to the desired output. We shall refer to this description as the input V since it can be represented as an n-dimensional vector, as shown in the above examples:

$$V = (x_1, x_2, x_3, \cdots, x_k, \cdots, x_n)$$

The totality of all possible m input vectors V can be regarded as an $n \times m$ table or matrix M (set of all vectors V) where n is the number of dimensions or columns and m the number of vectors or lines, m being dependent on the definition of the value ranges of the variables x_1 to x_n, for example, on the sampling rate. Each generated sound is then described by either a single line or a sequence of lines of the matrix. These lines are the protocol of the generated sound, or its metric description within the VOSIM model. As already stated, the chief objective of the experiments is to find the input dimensions relevant to the description of It is usual today to refer to a function which assumes one of the two logical values (*true* or *false*) as a predicate. A relation $R(e_1, e_2, \cdots, e_n)$ which according to the arguments assigned to the expressions $e_1, e_2, \cdots, e_n$ is *true* or *false*, is thus a predicate. Examples (15)–(19) can, in the same fashion, be regarded as relations which describe the sounds to be described either truly or falsely. We therefore speak of VOSIM predicates and call the ones which assume the *true* value calculable VOSIM predicates since, when the variables bound by universal quantifiers are replaced by constants, they always result in true sound description [such as in examples (1)–(14)]. With this principle as a basis, work is at present not only being done on a list of calculable VOSIM predicates, but the sound synthesis programs MIDIM8 and MIDIM9 have also been developed (MInimum DescrIption of Music). These programs can easily be expanded to become very versatile composing programs [5]–[7].

The same applies to work being done in the field of VOSIM speech synthesis [see examples (1)–(6) and (15)–(17)].

4. THE SPECTRUM

The signal function shown in Fig. 2 is described by

$$s(t) = f(t) \cdot p(t).$$

In this, $f(t)$ is the staircase-shaped envelope (for convenience's sake we assume A to be equal to 1):

$$f(t) = u(t) - b^{N-1}\cdot u(t-NT) + (b-1)\sum_{n=1}^{N-1} b^{n-1}\cdot u(t-nT)$$

$u(t)$ being the step function:

$$u(t) = \begin{cases} 1, & \text{if } t \geq 0 \\ 0, & \text{if } t < 0 \end{cases}$$

and $p(t)$ is the sinusoidal carrier:

$$p(t) = \tfrac{1}{2}(1 - \cos \Omega t)\ (= \sin^2 2\Omega t)$$

with $\Omega = 2\pi/T$.

The Fourier transform of this signal is

$$S(\omega) = \frac{j(1 - e^{-2\pi j\omega/\Omega})}{2\omega(\omega^2/\Omega^2 - 1)} \cdot \frac{1 - b^N \cdot e^{-2\pi jN\omega/\Omega}}{1 - b\cdot e^{-2\pi j\omega/\Omega}}$$

$$S(0) = \begin{cases} \tfrac{1}{2}T\cdot \dfrac{1-b^N}{1-b}, & \text{if } b < 1 \\ \tfrac{1}{2}TN, & \text{if } b = 1 \end{cases}$$

$$S(\Omega) = \begin{cases} \tfrac{1}{4}T\cdot \dfrac{1-b^N}{1-b}, & \text{if } b < 1 \\ \tfrac{1}{4}TN, & \text{if } b = 1. \end{cases}$$

From this it follows for the magnitude spectrum:

$$|S(\omega)| = \left| \frac{\sin \pi \frac{\omega}{\Omega}}{\omega(\frac{\omega^2}{\Omega^2} - 1)} \right| \sqrt{\frac{1 - 2b^N \cdot \cos 2\pi N\frac{\omega}{\Omega} + b^{2N}}{1 - 2b\cdot \cos 2\pi\frac{\omega}{\Omega} + b^2}} \qquad (A)$$

$$|S(0)| = \begin{cases} \tfrac{1}{2}T\cdot \dfrac{1-b^N}{1-b}, & \text{if } b < 1 \\ \tfrac{1}{2}TN & \text{if } b = 1 \end{cases}$$

$$|S(\Omega)| = \begin{cases} \tfrac{1}{4}T\cdot \dfrac{1-b^N}{1-b} & \text{if } b < 1 \\ \tfrac{1}{4}TN & \text{if } b = 1 \end{cases}$$

and for the phase spectrum:

$$\phi(\omega) = \tan^{-1}$$

$$\frac{(1+b)\sin \pi\frac{\omega}{\Omega} - b^N \sin(2N+1)\pi\frac{\omega}{\Omega} + b^{N+1}\sin(2N-1)\pi\frac{\omega}{\Omega}}{(b-1)\cos \pi\frac{\omega}{\Omega} + b^N\cos(2N+1)\pi\frac{\omega}{\Omega} - b^{N+1}\cos(2N-1)\pi\frac{\omega}{\Omega}}. \qquad (B)$$

Fig. 3 shows the magnitude spectrum and Fig. 4 the phase spectrum of the signal in Fig. 2.

A few characteristic properties of the magnitude spectrum can be observed. It is roughly described by the product of a $\sin x/x^3$ factor with a periodic function. The consequence of this factor is that the spectrum outside the frequency range $0-2\Omega$ can be ignored. There is also a strong dc component and finally, the most important property, the spectrum exhibits a peak in the vicinity of Ω. When the signal is repeated periodically, the spectrum changes into a discrete spectrum, the envelope of which if formed by Eq. *(A)*. This signal has an obvious formant character. Besides the pitch, which is determined by the repetition frequency, frequency components in the vicinity of Ω are accentuated. As we already know, many speech and musical sounds have a formant structure, and this explains the usability of the model (for a more detailed description see [8]).

For certain special values of N and b which are of particular importance for speech sounds, simpler expressions can be found for the spectrum:

a) $N = 2, b \leqslant 1$:

$$|S(\omega)| = \left| \frac{\sin \pi\frac{\omega}{\Omega}}{\omega(\frac{\omega^2}{\Omega^2} - 1)} \right| \sqrt{1 + 2b\cdot \cos 2\pi\frac{\omega}{} + b^2}.$$

The time function is shown in Fig. 5, the magnitude spectrum in Fig. 6.

b) $b = 1$:

$$|S(\omega)| = \left| \frac{\sin \pi N\frac{\omega}{\Omega}}{\omega(\frac{\omega^2}{\Omega^2} - 1)} \right|.$$

See Figs. 7 and 8 for an illustration of the time function

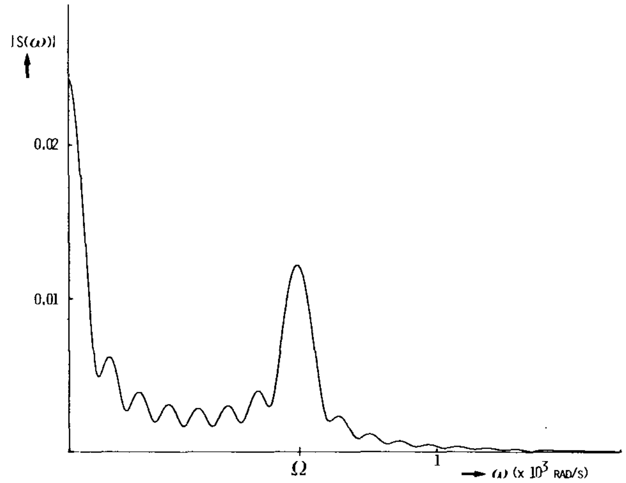


Fig. 3. Magnitude spectrum for $N = 8, b = 0.85, T = 10$ ms.

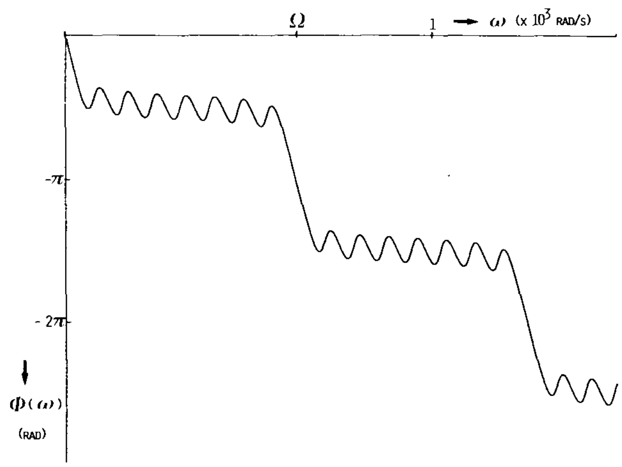


Fig. 4. Phase spectrum for $N = 8, b = 0.85, T = 10$ ms.

and the spectrum.

When the signal in Fig. 7 is repeated periodically (repetition frequency Ω'), any harmonics which coincide with the zero crossovers are missing from the spectrum. From

$$\sin \pi N \frac{n\Omega'}{\Omega} = 0$$

it can easily be derived that the above will be the case with harmonics whose number is a multiple of $(1 + M/NT)$, possibly with the exception of the formant frequency $n'\Omega'$ $(=\Omega)$.

A generally valid procedure for deriving the parameters of the description from a given signal does not (yet) exist. With a little experience, however, it is perfectly possible to make an assessment of these parameters from characteristics of the waveform and from the spectrum, and then to verify by means of experiment. Fig. 9 shows the spectrum of a bassoon tone, and Fig. 10 that of a VOSIM approximation of that tone, realized in this way [see example (14), second line]).

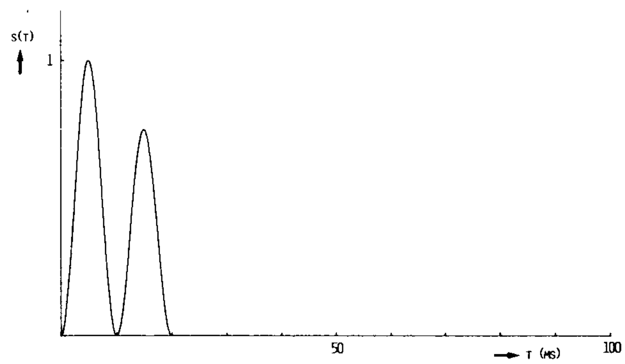


Fig. 5. Time function for $N = 2, b = 0.75, T = 10$ ms.

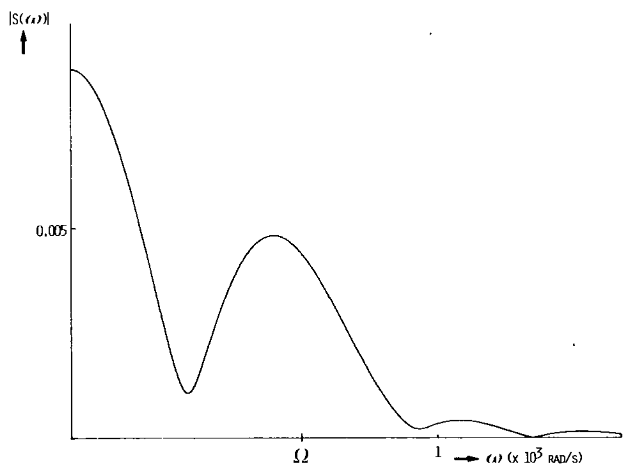


Fig. 6. Magnitude spectrum for $N = 2, b = 0.75, T = 10$ ms.

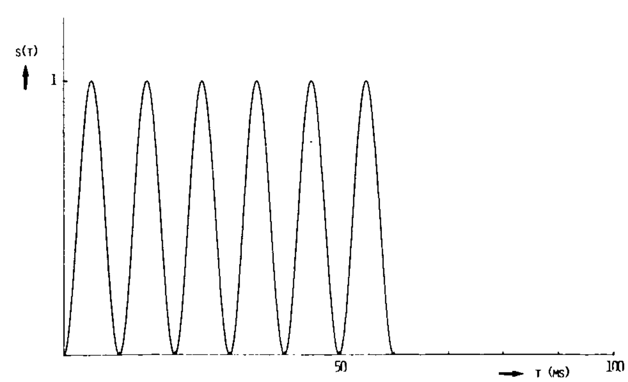


Fig. 7. Time function for $N = 6, b = 1, T = 10$ ms.

5. THE VOSIM OSCILLATOR

Experiments with the signals described here have been performed with both analog and digital equipment. The value ranges of the parameters were determined with the flexible analog apparatus. For more precise and reproducible signal generation, a computer (PDP-15) was used. A computer is indispensable for the production of longer series of VOSIM signals (for example, synthetic speech).

It is easy to generate (software) stationary VOSIM signals using a computer. This becomes more difficult as the number of variables increases, certainly when two independent VOSIM signals are to be generated simultaneously, and so it was finally decided to make use of a digital oscillator. The point of departure here was that any

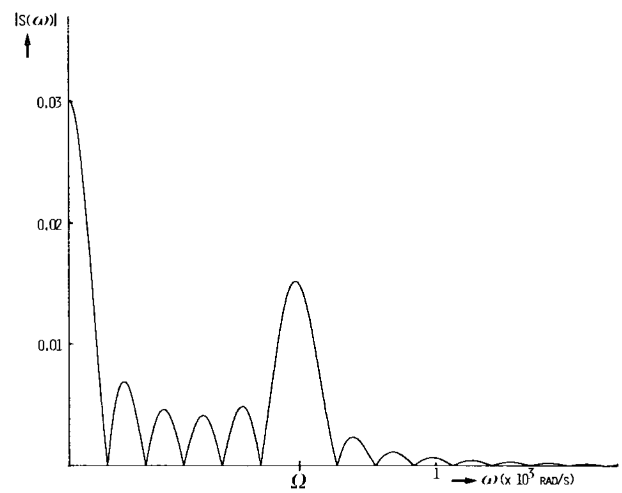


Fig. 8. Magnitude spectrum for $N = 6, b = 1, T = 10$ ms.

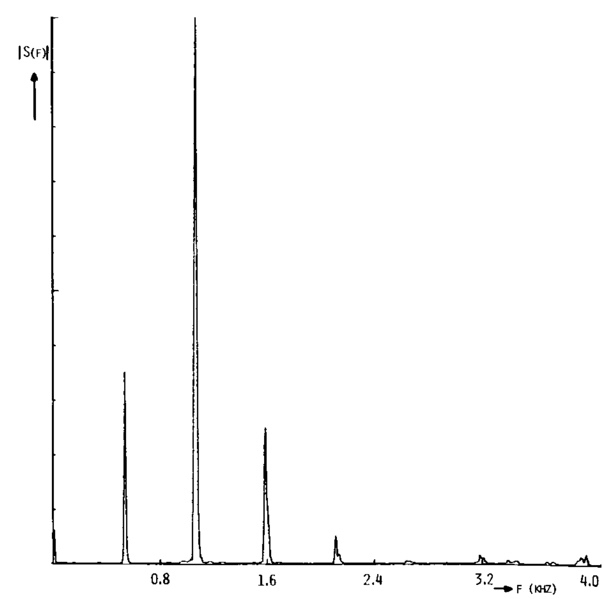


Fig. 9. Spectrum of a bassoon tone (253 Hz).

VOSIM signal can be regarded as a series of single $\sin^2$ pulses with a particular amplitude and duration, and followed by a delay (that can be 0 in the case of connecting pulses). We decided to construct an oscillator capable of supplying this kind of elementary signal and thus requiring three control data items. Two such oscillators were built. The input data (lines with the 12 variables according to the list discussed above) are read from disk and converted by a program into the corresponding T, M, and A values for the two oscillators. Interpolation and modulation are thus software realized. Fig. 11 shows the block diagram of the oscillator. The core is an 8-bit 128-word random access memory (RAM) in which the sampling values of a $\sin^2$ pulse are placed. (In connection with other possible applications of the oscillator, a RAM was used instead of a

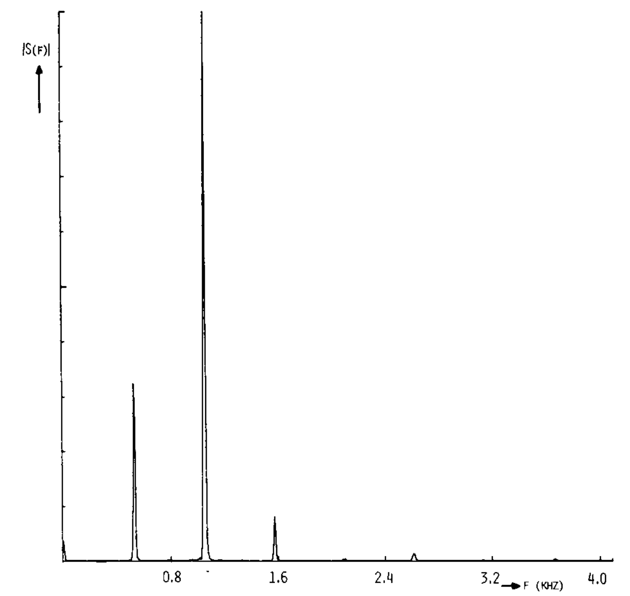


Fig. 10. Spectrum of the VOSIM simulation of the bassoon tone of Fig. 9.

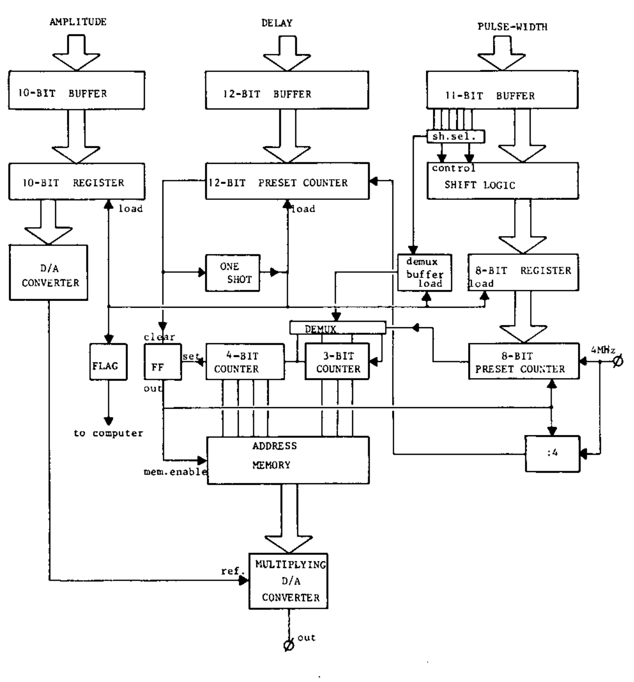


Fig. 11. The VOSIM oscillator.

read-only memory.) The circuits required for reading in the samples are disregarded here. A 7-bit address counter which counts from 0 to 127 is responsible for sending the samples one by one to a digital-to-analog converter where they are converted into an analog voltage. The speed at which the address counter is increased thus determines the pulse width T. In order to make this pulsewidth variable, the trigger pulses for the address counter are derived from a preset counter in which a number depending on T is placed. As soon as the preset value is reached in this counter, which is triggered by a 2-MHz clock, a pulse is produced, triggering the address counter. Practically the same system is used for the realization of the delay. Once the RAM has been run through completely, a memory-disable pulse is given. The delay preset counter is increased by a fixed frequency until the set value is reached, and the cycle starts running through the RAM again. In order to make it possible to control the amplitude, the digital-to-analog converter employed is a multiplying type. The number in the amplitude buffer is converted into an analog voltage which serves as reference voltage for the output converter.

There are of course several other ways to design an oscillator with the described properties. An example is Christiansen's microprocessor-based design [9].

ACKNOWLEDGMENT

The authors wish to thank Jo Scherpenisse for his contribution to the construction of the oscillator. This research was supported in part by the Fonds National Suisse de Recherche Scientifique and the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

REFERENCES

[1] W. Kaegi, "A Minimum Description of the Linguistic Sign Repertoire (part 1)," *Interface,* vol. 2, pp. 141-156 (1973).

[2] S. Tempelaars and J. Scherpenisse, "Exponential Decay," *Interface,* vol. 5, pp. 207-224 (1976).

[3] W. Kaegi, "A Minimum Description of the Linguistic Sign Repertoire (Part 2)," *Interface,* vol. 3, pp. 137-158 (1974).

[4] S. Tempelaars, "The VOSIM Oscillator," Proc. 1st Int. Conf. on Computer Music, Massachusetts Institute of Technology, Oct. 28-31, 1976 (in press).

[5] W. Kaegi, "Ein musikalisches Zukunftsmodell. Wissenschaft und Technik im Dienste des zukünftigen Komponisten," Proc. Int. Conv. for Micro-electronics, Munich, Germany, Nov. 29-December 1, 1976.

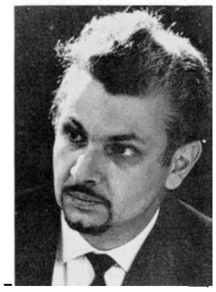
[6] W. Kaegi, "Le Modèle de Synthèse de Son 'VOSIM' et ses Aspects Psychoacoustiques," Proc. Symposium sur la Psychoacoustique musicale, IRCAM, Paris, July 11-13, 1977 (in press).

[7] W. Buxton, "A Composer's Introduction to Computer Music," *Interface,* vol 6, pp. 57-71 (1977).

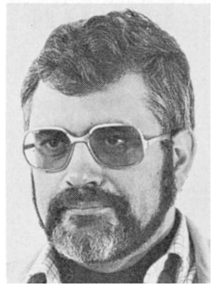
[8] S. Tempelaars, "The VOSIM Signal Spectrum," *Interface,* vol 6, pp. 81-96 (1977).

[9] S. Christiansen, "A Microprocessor-Controlled Digital Waveform Generator," *J. Audio Eng. Soc. (Project Notes/Engineering Briefs),* vol. 25, pp. 299-309 (1977).

THE AUTHORS



W. Kaegi



S. Tempelaars

Werner Kaegi was born near Zurich, Switzerland, in 1926. He studied musical composition, musicology, and mathematical logic in Zurich, Basel, Heidelberg, and Paris. He received the Ph.D. degree (cum laude) in 1952 from the University of Zurich with a dissertation on the structural problems in the inventions and fugues of J.S. Bach. Dr. Kaegi was a student of Paul Hindemith, Arthur Honegger, and Louis Aubert. He came in contact with electroacoustical music for the first time in 1951 at the Paris radio.

His numerous compositions of instrumental and electronic music have been performed at many international festivals (including International Festival Lucern, Berlin, Avignon, Zagreb, Nordlyd Oslo). He won numerous competitions and on commission from the Swiss government wrote the electronic music for the Swiss Pavilion at the Osaka World's Fair in 1970. Dr. Kaegi is the author of technical books about electronic music: *Was ist elektronische Musik* (Zurich 1967) and *Vom Sinuston zur elektronischen Musik* (Stuttgart 1970) and of numerous articles in professional journals. He worked at the Centre de Recherches Sonores de la Radio de la Suisse Romande in Geneva from 1963 to 1970. Since 1971 he has been at the Institute of Sonology of the University of Utrecht, Holland, as a researcher and teacher.

●

Stan Tempelaars was born in Breda, Holland, in 1938. He studied physics and musicology in Utrecht.

Since 1961 he has been working at the Institute of Sonology of the University of Utrecht (formerly known as the Studio for Electronic Music), first as a student, later as a staff member and teacher. Originally he concentrated on the development of analog studio equipment, but after the institute's purchase of a computer his work turned to the application of digital signal processing in the field of sonology.