SOUND SYNTHESIS BY RULE

Ercolino Ferretti

University of Utah

Abstract

Synthesis by rule is the process of synthesizing digitized sound entirely by mathematical representation and rules based on the physical, acoustic and psychophysical principles of generating sound, coupled with the parameters of speech, the concepts of performance and the principles of composition. The objectives of synthesis by rule are to approximate conventionally produced sounds with a high degree of accuracy at very low bit rates, and also to find new and interesting sounds. There are three major categories of information which pertain to synthesis by rule - the sound generating system or instruments, the performance and the composition. Only the first category, information pertaining to the instrument, will be considered in this paper.
SYNTHESIZED SOUND

Electrical signals which are representative of synthesized sound appear as analog or digitized signals. Sound in the form of digitized signals has associated with it bit rates. Direct transmission of digitized signals requires about one half million bits per second for high fidelity music and about one tenth that rate for telephone quality speech.

The general categories of digitally synthesized sound are speech synthesis and sound synthesis by rule. Speech synthesis is the result of utilizing a vocoder. A vocoder is an analysis-synthesis system whereby the original speech is presented at the input, an analysis is performed producing a set of parameters which are encoded and transmitted as binary information. The information is decoded at the receiving end from which synthesized speech is produced. The goal is to produce digitized speech at relatively low rates so that transmission is both economically feasible and can be made secure. For example, typical vocoder speech is transmitted over telephone lines at 2400 bits per second. The speech is highly intelligible and of fair-to-good quality.

Synthesis by rule is the process of synthesizing digitized sound entirely by mathematical representation and rules. There is no original sound at the input, consequently no waveform analysis. The mathematical representations and rules are based on the physical, acoustic and psychophysical principles of generating sound, coupled with the parameters of speech - elisions, inflection, intonation, intensity - the concepts of performance - nuances - and the principles of composition - harmony, counterpoint, rhythm and other related parameters. The objectives of synthesis by rule are to approximate conventionally produced sounds and music with a high degree of accuracy, and also find new and interesting sounds as well as new ways of organizing sound. A successful system of synthesis by rule results in data rates an order of magnitude less than typical vocoder
rates and leads to a more fundamental understanding of sound properties.

One of the major tasks in establishing synthesis by rule is the development of a model in which the parameters of waveforms can be easily related to the corresponding sound in a familiar sense of familiar sound categories and at familiar information rates. From an information point of view, a most interesting observation regarding the performance of a musical instrument is the apparent nearly one to one correspondence between the action of the performer and the resulting sound. There appears to be only a handful of operations that can be executed by the performer in the span of say, one second... the placement of a finger, the twist of a lip, the positioning of the tongue. A reasonable assumption regarding the information supplied by the performer to instrument or voice, is one of low order bit rates.

NON-PERIODICITY

In establishing a low bit rate model, a better understanding of the intrinsic characteristics of sound waveforms as related to physics is a good place to begin. While most waveforms produced by conventional musical instruments (violin, trumpet, etc.) appear to be nearly periodic, they, in fact, also possess clear cut non-periodic characteristics. Approximation of these waveforms by models based on periodicity appear not to have the subtle characteristics of the original. Vocoder models can produce fair-to-good approximations of the original sound, but they must present new information at nearly pitch rates which represents a relatively large amount of information and also the parameters are not readily related to the psychologically sensed parameters, and furthermore the original signal is also required.

Since non-periodicity is of major importance, the primary task is to
seek out its origin. One consideration is to assume the waveform as being fundamentally periodic and that the non-periodicity is a result of perturbations about the periodic waveform caused by nuances such as vibrato and tremolo, breathing in wind instruments, unsteady hands in bowing stringed instruments, attack transients and reverberation.

Undoubtedly the above effects contribute to non-periodic behavior, but the fact remains that with no contribution made by the performer whatsoever, the sound emanating from a musical instrument still will display non-periodic characteristics. A case in point is the monochord. A plucked monochord is one of the simplest instruments to consider and yet one of the most obvious examples of non-periodicity. Measurements of the monochord clearly indicate the obviousness of non-periodicity including beat-like phenomenon.

MONOCHORD MEASUREMENTS

The monochord was constructed by attaching both ends of a piece of piano wire to harpsichord pegs which were implanted into hardwood, Fig. 1. The string was displaced at the center in an upward position about one inch from the quiescent position and then released. The resulting acoustic wave was recorded on one channel of a dual channel audio tape recorder. An endless tape loop was formed for continuous playback with a timing pulse recorded on the second channel for synchronizing all measurements.

Fig. 2.1 illustrates the experimental setup for making the measurements. The GR 1900-A Wave Analyzer was set for a bandwidth of 50 cycles and the time base of oscilloscope #2 was set for 0.2 sec/cm. The analyzer was tuned to twelve successive harmonics beginning with the fundamental of 175 HZ.

A set of twelve components each represented by a dual trace on oscilloscope #2 with corresponding photographs are shown in Fig. 3.1 through 3.12. The lower
trace in each photograph illustrates the restored output, amplitude versus time, of the GR 1900-A Wave Analyzer, and the upper trace illustrates the zero crossings of the corresponding restored output.

The output of the zero crossing meter can be interpreted as discrete frequency changes as a function of time represented by the points of transition which fall on the dashed lines in Fig. 2.2. The twelve upper traces of frequency versus time, are represented more conveniently by smooth curves in Fig. 4.1 through 4.12 with calibrated frequency scales for each component.

The zero crossing meter is unstable for certain classes of input, as seen in Fig. 3.2. However, there is sufficient stability to indicate the pertinent features. Quite clearly, there is independent frequency as well as independent amplitude information as a function of time from component to component.

Beating is also apparent, and can be seen as well in the amplitude versus frequency curve, Fig. 5. The time waveform illustrated in Fig. 6.1 represents a duration of 0.5 sec., from which a magnified slice, Fig. 6.2 through 6.4, shows three contiguous intervals indicating approximately one cycle per division or five cycles per photograph. Clearly observed is the character of non-periodicity.

SOUND SYNTHESIS MODELS

As a result of experimental observations, a general model for synthesizing sound has been developed, eq (1).

\[ S(t) = \sum_{n=1}^{m} A_n(t) \cdot F_n(t) \]  

(1)

A tone or sound, \( S(t) \), is represented by a sum of components where \( A_n(t) \) is defined as the amplitude variation term and \( F_n(t) \) as the frequency variation term, allowing for independent amplitude and frequency information for each
component. The objective is to have the availability of an experimental model for the study of non-periodic behavior by synthesis.

A first phase of experimentation consisted of summing a set of sines with relatively slowly varying amplitude and frequency characteristics. The model included non-integral multiples spaced relatively close to neighboring integral multiples or non-integral multiples replacing integral multiples, see Fig. 7.1 and 7.2. Non-integral frequencies are represented as \( kf_n/f_1 \), with typical values for \( k \) ranging from .95 to 1.1.

Approximations to live sounds, flute, hand drums, female voice (singing vowels) and other sounds have been fairly successful for \( 4m<12 \). These signals have the right sound and the right look, but they have one prime deficiency... lack of high frequency information. Thus synthesis by this technique is very limited. For example, a signal with a fundamental frequency of 50 HZ requires one hundred components to include information at 5 KHZ. The specification of one hundred components with independent amplitude and frequency information is utterly impractical.

A second phase was an attempt at amplitude and frequency modulation with the modulating frequencies at the same rate as the fundamental. While the modulation versions produce sufficient high frequency information, there is little possibility for independent frequency information.

A more recent phase has been the development of a model which inherently produces non-periodicity and is perceived as being more closely related to natural sounds. The model is based on a modification of frequency modulation where the instantaneous frequency is exponentiated. The time waveform in Figure 8.1 is an illustration of frequency modulation with \( f_c = f_m \), and Figure 8.2 shows the same relation for the exponentiated version.
Fig 2.1

CONSTANT FREQUENCY

INCREASING FREQUENCY

DECREASING FREQUENCY

Fig. 2.2
Fig. 7.1

Fig. 7.2
Fig. 8.1

Fig. 8.2
CONCLUSIONS

Several instruments of the string, brass and percussion family as well as voices were examined and all exhibited similar characteristics as the monochord in relation to non-periodicity and beating. Undoubtedly, there is a common behavior in all mechanical vibrating systems which manifests itself as a natural non-periodic phenomenon. Further research may reveal a simplistic model which defines the behavior of mechanical vibrating systems. At present, the author can only speculate as follows: the tension of a vibrating system is not constant, but varies as a function of the vibration itself, which suggests a type of vibrato function in which the percent change in frequency modulation is logarithmic, (Fig. 8.2).
ACKNOWLEDGEMENTS

The information in this paper covers a period from 1962 to 1975.

(1) The measurements of the monochord were conducted in 1966 with the aid of two of my students, Tony Lin and Jonathan Sachs, while enrolled in an interdisciplinary program sponsored by the Research Laboratory of Electronics and Department of Humanities at the Massachusetts Institute of Technology.

(2) The design of the zero crossing meter was executed by William F. Kelly, a graduate student, 1962, in the Electrical Engineering Department at the Massachusetts Institute of Technology.

(3) Laboratory equipment and computing facilities were made available by the Research Laboratory of Electronics, and the Department of Humanities at the Massachusetts Institute of Technology.

(4) Computer facilities from 1970 to 1975 were provided by ARPA supported research at the University of Utah.
REFERENCES


