# Analysis and synthesis of bird vocalizations: An FFT-based software system

CHRISTOPHER CLARK
*Rockefeller University Field Research Center, Millbrook, New York 12545*

and

ROBERT J. DOOLING and TIMOTHY BUNNELL
*University of Maryland, College Park, Maryland 20742*

An FFT-based software package has been designed and implemented in order to analyze and synthesize bird vocalizations. The system is flexible so that screen editing can be performed on either the original waveform or the Fourier coefficients of the transform. The edited coefficients of selected notes, syllables, or songs are stored for later modification or combination with other song components to compose entirely new songs. These analyses and synthesis methods have proved to be extremely useful in the elucidation of birds' perceptive capabilities and the biological significance of their songs.

For anyone interested in the study of acoustic phenomena in animals and, in particular, the study of their communication signals, there are obviously certain analytical tools that can not only provide improvements over old techniques but also open up entirely new approaches to the problems at hand. In this report, I will briefly discuss the application of some computer signal processing techniques as they relate to animal acoustic signals.

The programs that I will be discussing were developed in Peter Marler's laboratory at the Rockefeller University Center for Field Research in Millbrook, New York. The research group has been intensely interested in delineating the various communicative functions of animal acoustic signals, with particular emphasis on birdsong. However, the programs have also been successfully applied to a variety of other vertebrate groups, including primates and whales. In order to illustrate this versatility, I will refer to the work that has been successfully accomplished with birdsongs.

Let me begin with the assumption that any tool that allows you to decompose a signal (in this case acoustical) into constituent parts, reverse the process, and completely recover the original signal is very useful. The usefulness is twofold.

From the decomposition analysis, one can elucidate the mechanisms of sound production. This has been beautifully illustrated in the early work on human speech, in which engineers at Bell Laboratories (Flanagan, 1972) were able to decompose the speech waveform and, from that, build models of the human vocal system. A similar approach motivated Greenewalt (1968) in his classic study of birdsongs. By meticulously dissecting the songs, Greenewalt began to uncover the mechanisms of avian sound production.

A second usefulness of this analytical tool is in synthesis. That is, if one can synthesize sounds that contain known amounts of certain acoustic dimensions, one can play these sounds back to the animals and obtain a great deal of information on their perceptive abilities as well as the biological significance of those sounds.

There are several ways to approach the synthesis problem. A model can be built to serve as the basis of the synthetic vocalization, or the vocalizations produced by the animals can be decomposed into basic units and then the desired vocalization can be synthesized by composition from these basic elements. Since there is no good model of the avian vocal production system and since sound synthesizers have tended not to produce very good imitations of birdsongs, this second approach of synthesis by analysis has been adopted.

The basic concept in the operation of our analysis/synthesis technique is the discrete Fourier transform. We apply the fast Fourier transform (FFT) to a digitized representation of an acoustic waveform. Modifications in phase, amplitude, frequency, or duration are then made by directly manipulating the Fourier coefficients. An inverse Fourier transform converts the modified coefficients back into a digital waveform, which is then converted to analog and recorded on a tape recorder (see Zoloth, Dooling, Miller, & Peters, 1980).

There are six program groups in the analysis/synthesis algorithm. These can be summarized by function.

(1) SNDIN is a program that digitizes analog input from an A/D converter at a sample rate of 20 kHz. Other sampling rates are possible with minor program changes. By the technique of double buffering, it is possible to write the sampled data directly to hard disk

at 20 kHz or even to RX-02 floppy disk at lower sampling rates. The digitized train of samples (representing amplitude by time) is given a file name with a suffix denoting pulse code modulation (PCM).

(2) EDPCM allows cursor editing of the PCM waveform prior to conversion to spectral representation. The beginning or end of the sound segment of interest is selected by viewing sections of the input waveform on a graphics terminal in an amplitude by time display. Two successive 256-point segments (12.8 msec of the sampled waveform) are displayed at once. The user "steps" through the file, finds the initial part of the signal of interest, and marks this position with the cursor. The terminal part is determined in a similar fashion. The sound segment bounded by the cursor positions is saved in a new PCM file.

(3) PCMFFT is a program that converts the PCM waveform (amplitude by time information selected by program EDPCM) to a spectral representation. Using the radix 2, time decomposition, bit reversal method, a disk file is produced with interleaved amplitude and phase information for each FFT performed on successive sets of the input sample train. A tapered Hanning window is used to compensate for the discontinuities introduced in violating the assumption of infinite periodicity. The PCMFFT program takes successive windowed spectra with half-frame overlap for the entire input sample train.

(4) EDFFT modifies the signal spectrum in the graphics mode. There are four different display modes: frequency vs. time (spectrograph), amplitude vs. frequency (power spectrum), phase vs. time, and amplitude vs. time. Editing changes include modification of signal phase, amplitude, and frequency, mixing of two sound segments, and deletion, insertion, or replication of selected time segments. Changes in the spectral content of the signal produce corresponding alterations in the FFT data file. Frequency changes are restricted to upward or downward shifts in multiples of the minimum frequency resolution (for a 128-point segment, that is 1/6.4 msec or 156.25 Hz).

(5) FFTPCM converts the altered FFT file back to PCM data using the inverse discrete Fourier transform. The new PCM file then contains a digital representation of an acoustic signal resulting from changes in the Fourier coefficients of the original signal.

(6) SNDOUT is a program that sends PCM data out through a D/A converter at the sample rate of 20 kHz. Both the input and output routines are written partially in assembly language. These I/O routines can be called from FORTRAN programs allowing a wide range of flexibility in output schemes. In addition, a series of PCM files can be concatenated using the system peripheral interchange program (PIP).

I would now like to illustrate some of the program uses just to provide a feeling for its flexibility and applicability in other areas.

Many songbirds acquire their normal songs through learning, and this vocal learning process occurs during a sensitive period when the birds are young (Mulligan, 1966). One question that we have asked concerns the specificity of this vocal learning process. That is, what is there about a species-specific vocalization that makes it more attractive as a model for learning than some other acoustic pattern?

A naturally occurring system exists for studying this selective learning process in birds. There are two closely related species, the song sparrow and the swamp sparrow, that also live and breed within the same habitats. Their territorial songs play an important role in species and individual recognition and mate attraction. The songs of the two species share certain common properties. For example, some of the individual utterances, or notes, are similar, but the songs differ in structure, note complexity, and, most strikingly, in the timing pattern.

The synthesis by analysis programs were used to generate songs that contained mixtures of features from the two species (see Figure 1). First, a song sparrow song is input into the computer. We then eliminate all but one note element, convert the signal into its Fourier transform, and store this single song sparrow element in its own file.

Next we input a swamp sparrow song, extract a pair of notes, convert them, and store these in another file. Then the two files are mixed to produce a syllable that is then repeated to produce what looks like and sounds like a normal swamp sparrow song; the only difference is that the center note of the three-note syllable is not normally sung by swamp sparrows.

These various synthesized songs are used to tutor young birds of both species in an effort to determine which of the various acoustic dimensions are most important in selective learning. The results demonstrate that swamp sparrows can learn the song sparrow notes if the notes are placed in a normal swamp sparrow song context (Marler & Peters, 1981).

This experiment exemplifies the usefulness of these FFT programs for manipulating sounds in ways that would have otherwise been extremely difficult.

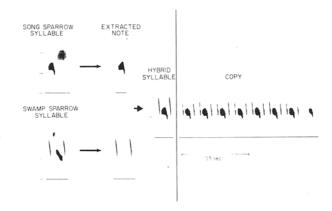Another extremely fruitful experimental approach



Figure 1. Example of synthesis by analysis procedure. Syllables from swamp sparrow and song sparrow songs were used to produce a hybrid syllable.

has been taken by Dooling and Searcy (1981). Dooling has successfully used synthesized songs in psychoacoustical experiments in order to test whether 1-day-old song sparrow or swamp sparrow chicks have a predisposition to learn their species-specific songs. As it turns out, these very young birds, despite having nearly identical hearing abilities, respond differently to the songs of their species than to the songs of a congeneric species.

## SUMMARY

We have developed a set of programs that allows the user to manipulate animal vocalizations for the purpose of signal analysis or signal synthesis. This set has been successfully used for experimentation not only with birdsong but also with primate and cetacean sounds. We believe that these programs provide a significant tool for others interested in the analysis and synthesis of animal vocalizations.

Two versions of these programs are presently in operation. One version is running on a PDP-11/23 processor with VT-125 graphics terminal, an AA VII-A A/D converter, and RX02 floppies. The second version is running on a PDP-11/24 processor with Tektronix 4006 graphics terminal, an AR-11 10-bit A/D converter, and RX01, RK05, and RP02 drives. Source programs are written mostly in FORTRAN IV, with a few written in macro assembly language. All are running under the DEC RT-11 operating system.

## AVAILABILITY

Source code listings for the first version (PDP-11/23, VT-125, and RX02) are available from Robert J. Dooling, Psychology Department, University of Maryland, College Park, Maryland 20742. Source code listings for the second version may be obtained from Christopher Clark, Rockefeller University Field Research Center, Tyrrel Road, Millbrook, New York 12545. There is a $5 charge for copying. Users interested in obtaining the software should send a floppy diskette (RX01 compatible, single sided, single density) to the appropriate address above. We request that publications resulting from this software formally acknowledge this source.

### REFERENCES

DOOLING, R. J., & SEARCY, M. H. A comparison of auditory evoked potentials in two species of sparrow. *Physiological Psychology*, 1981, *9*, 293-298.

FLANAGAN, J. L. *Speech analysis synthesis and perception.* New York: Springer-Verlag, 1972.

GREENEWALT, C. H. *Bird song: Acoustics and physiology.* Washington, D.C: Smithsonian Institute Press, 1968.

MARLER, P., & PETERS, S. Birdsong and speech: Evidence for special processing. In P. Eimas & J. Miller (Eds.), *Perspectives on the study of speech.* Hillsdale, N.J: Erlbaum, 1981.

MULLIGAN, J. *Singing behavior and its development in the song sparrow,* Melospiza melodia. Berkeley, Calif: University of California Press, 1966.

ZOLOTH, S. R., DOOLING, R. J., MILLER, R. C., & PETERS, S. A minicomputer system for the synthesis of animal vocalizations. *Zeitschrift für Tierpsychologie*, 1980, *54*, 151-162.