Appendix A Digital Representation of Sound

About this appendix

This appendix provides a brief explanation of how sound is represented digitally. An understanding of the basic principles introduced here will be helpful in using Canary.

Digital sampling

Before a continuous, time-varying signal such as sound can be manipulated or analyzed with a digital computer, the signal must be *acquired* or *digitized* by an *analog-to-digital (A/D) converter*.[1] The A / D converter repeatedly measures or samples the instantaneous voltage amplitude of an input signal at a particular sampling rate, typically thousands or tens of thousands of times per second (Figure A.1). The digital representation of a signal created by the converter thus consists of a sequence of numeric values representing the amplitude of the original waveform at discrete, evenly spaced points in time.

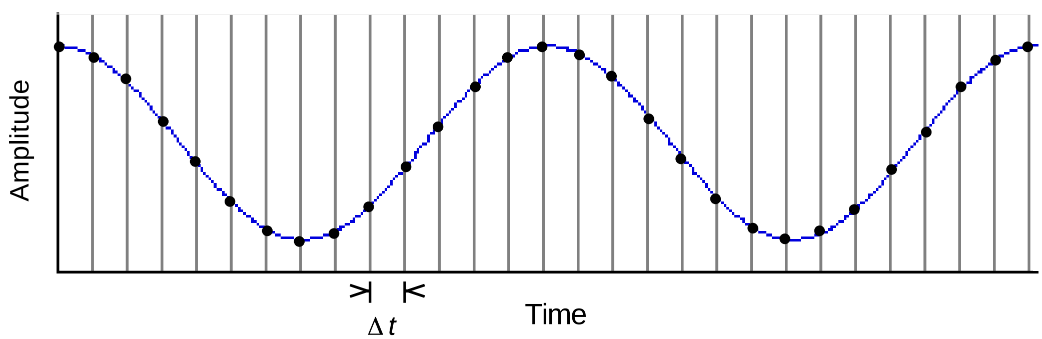

Figure A.1. Sampling to create digital representation of a pure tone signal. Measurements of the instantaneous amplitude of the signal are taken at a sampling rate of 1/Δt. The resulting sequence of amplitude values is the digitized signal.

The precision with which the digitized signal represents the continuous signal depends on two parameters of the digitizing process: the rate at which amplitude measurements are made (the *sampling rate* or *sampling frequency*), and the number of bits used to represent each amplitude measurement (the *sample size*).

[1]Recent Macintosh models (including Quadra, Performa, Centris, PowerMac, and Powerbook models) come equipped with a built-in A / D converter, which takes its input from the Mac's microphone jack. A / D converters can also be purchased from third-party manufacturers (e.g., MacRecorder from MacroMind Paracomp) for use with these or other Mac models.

Sampling rate

Canary's Sound Recording dialog enables you to choose the sampling rate at which a signal is to be digitized. The choices available are determined by the A/D converter hardware and the program (called a *device driver*) that controls the converter; most converters have two or more sampling rates available.[1] The highest frequency available with the Macintosh built-in A/D converter depends on which model of Macintosh you are using. Commercial digital audio applications use higher sampling rates (44.1 kHz for audio compact discs, 48 kHz for digital audio tape). Once a signal is digitized, its sampling rate is fixed.[2]

The more frequently a signal is sampled, the more precisely the digitized signal represents temporal changes in the amplitude of the original signal. The sampling rate that is required to make an acceptable representation of a waveform depends on how rapidly the signal amplitude changes (i.e., on the signal's frequency). More specifically, the sampling rate must be more than twice as high as the highest frequency contained in the signal. Otherwise, the digitized signal will have frequencies represented in it that were not actually present in the original at all. This appearance of phantom frequencies as an artifact of inadequate sampling rate is called *aliasing* (Figure A.2). The highest frequency that can be represented in a digitized signal without aliasing is called the *Nyquist frequency,* which is half the frequency at which the signal was digitized. The highest frequency in a spectrogram or spectrum calculated by Canary is always the Nyquist frequency of the digitized signal. If the only energy above the Nyquist frequency in the analog signal is in the form of low-level, broadband noise, the effect of aliasing is to increase the noise in the digitized signal. However, if the spectrum of the analog signal contains any peaks above the Nyquist frequency, the spectrum of the digitized signal will contain spurious peaks below the Nyquist frequency as a result of aliasing. The usual way to guard against aliasing is to pass the analog signal through a low-pass filter (called an anti-aliasing filter) before digitizing it, to remove any energy at frequencies greater than the Nyquist frequency. (If the original signal contains no energy at frequencies above the Nyquist frequency or if it contains only low-level broadband noise, this step is unnecessary.)

---

[1]The device driver for the built-in Macintosh sound input port is included as part of the system software. Device drivers for other A/D converters are supplied with the converter, usually as a file that must be placed in the System folder.

[2]Some digital signal processing programs (e.g., MATLAB's Signal Processing Toolbox) enable you to resample a digitized signal at a lower rate than the original sampling frequency by discarding some samples, or to increase the nominal sampling rate by interpolating samples.

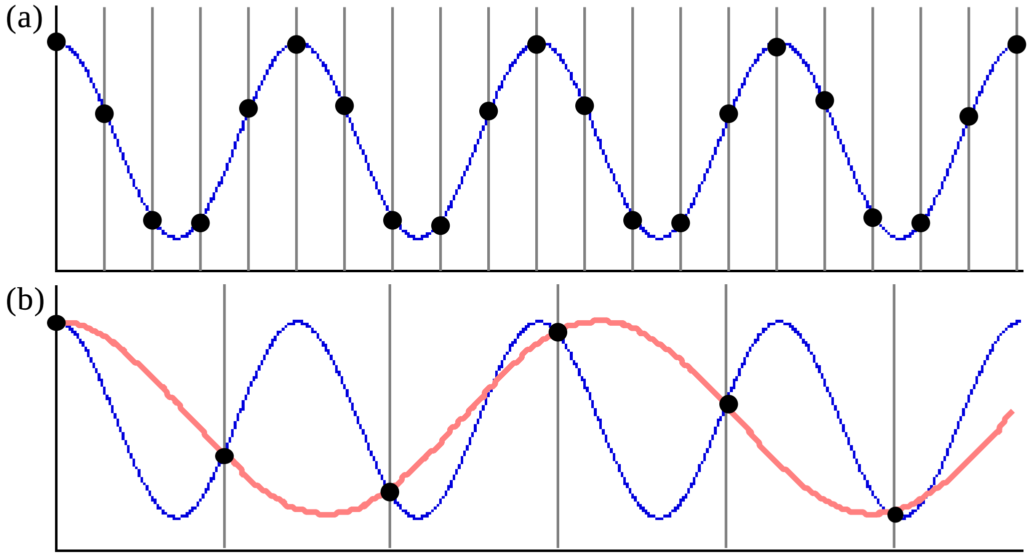

Figure A.2. Aliasing as a result of inadequate sample rate. The same analog waveform is shown in both figures. Vertical lines indicate times at which samples are taken. **(a)** Sampling frequency approximately five times the signal frequency. **(b)** Sampling frequency approximately 1.5 times the signal frequency. The resulting digitized signal (gray waveform) exhibits aliasing: it portrays a waveform of lower frequency than the original analog signal.

In order to interpret a sequence of numbers as representing a time-varying signal, one needs to know the sampling rate. Thus, when a digitized signal is saved in a file format that is designed for saving sound information, information about the sampling rate is usually saved along with the actual data points comprising the signal. If you try to open a file with Canary that contains sound data, but no information about the sampling rate, Canary asks you for the sampling rate.

Sample size (amplitude resolution)

The precision with which a sample represents the actual amplitude of the waveform at the instant the sample is taken depends on the *sample size* or number of bits used in the binary representation of the amplitude value. Some A/D converters can take samples of one size only; others allow you to choose (usually through software) between two or more sample sizes. Some Macintosh models provide only 8-bit sampling capability; others allow you to choose between 8-bit and 16-bit samples. An 8-bit sample can resolve 256 (=2^8) different amplitude values; a 16-bit converter can resolve 65,536 (=2^{16}) values. Sound recorded on audio CDs is stored as 16-bit samples. When a sample is taken, the actual value is rounded to the nearest value that can be represented by the number of bits in a sample.

Since the actual analog value of signal amplitude at the time of a sample is usually not exactly equal to one of the discrete values that can be represented exactly by a sample, there is some error inherent in the process of digitizing (Figure A.3), which results in *quantization noise* in the digitized signal. The more bits used for each sample, the less quantization noise is contained in the digitized signal.

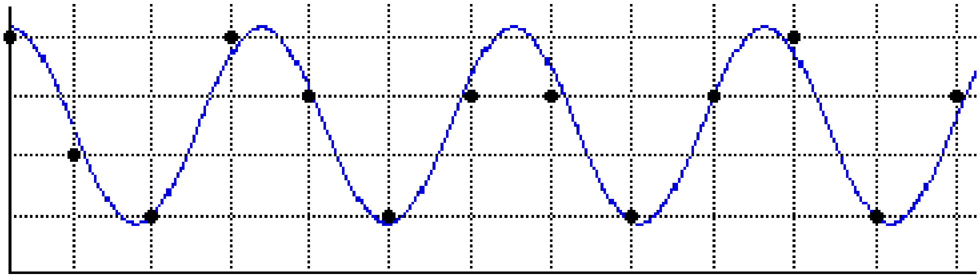

Figure A.3. Digitizing error with a hypothetical 2-bit sample size. 2-bit samples can represent only four different amplitude levels. At each sample time (vertical lines), the actual amplitude levels are rounded to the nearest value that can be represented by a 2-bit sample (horizontal lines). The amplitude values stored for most samples (black dots) are slightly different from the true amplitude level of the signal at the time the sample was taken.

The sample size determines the maximum dynamic range of a digitized sound. Dynamic range is the ratio between the highest amplitude and the lowest non-zero amplitude in a signal, usually expressed in decibels. The dynamic range of a digitized sound is 6 dB/bit.[1]

Storage requirements

The increased frequency bandwidth obtainable with higher sampling rates and the increased dynamic range obtainable with larger samples both come at the expense of the amount of memory required to store a digitized signal. The minimum amount of storage (in bytes) required for a digitized signal is the product of the sample rate (in samples/sec), the sample size (in bytes; one byte equals 8 bits), and the signal duration (seconds). Thus, a 5-second signal sampled at 22.3 kHz with 8-bit precision requires about 110 Kbytes of storage. The actual amount of storage required for a signal may exceed this minimum, depending on the format in which the samples are stored. For reasons of programming efficiency, Canary always uses a 32-bit format for its internal representation of a digitized signal, irrespective of the sample size with which the signal was digitized. The amount of memory (RAM) required by Canary to store a signal while working with it is thus four times the minimal requirement for a signal sampled with 8-bit resolution. When a signal is saved as a disk file, its storage requirements depend on the file format used and may be less than the amount of storage required by Canary's internal representation. For example, if a sound is recorded with 8-bit precision and saved in SoundEdit or AIFF format, it will be saved with 8-bit precision, (even though Canary uses 32 bits while working with the sound).

No matter what file format is use, digitized sound files take up a lot of storage space. If you plan on storing many or long digitized signals and need to save storage space, you might consider using a data compression program such as Compact Pro (shareware, from Cyclos, PO Box 31417, San Francisco, CA 94131-0417 USA; available from many users' groups and bulletin boards), StuffIt (Aladdin Systems, Inc.), or Disk Doubler (Salient Software, Inc.). These programs can often compress digitized sound files by 50% or more (the amount of compression will vary from file to file). Compressed files must be expanded before they can be used with Canary.

---

[1]The dynamic range of a signal in decibels is equal to 20 log(V_{max}/V_{min}), where V_{max} and V_{min} are the maximum and minimum voltage in the signal. For a digitized signal, $V_{max}/V_{min} = 2^n$, where n is the number of bits per sample. Since $\log(2^n) = n * .3$, the dynamic range of a digitized signal is 6 dB per bit.