Digitizing Speech Recordings for Archival Purposes

Bartek Plichta
Matrix, The Center for Humane Arts, Letters,
and Social Sciences Online
310 Auditorium
East Lansing, MI 48823 USA
+1 517 355 6187
plichtab@msu.edu

Mark Kornbluh
Matrix, The Center for Humane Arts, Letters,
and Social Sciences Online
310 Auditorium
East Lansing, MI 48823 USA
+1 517 355 9300
mark@mail.matrix.msu.edu

ABSTRACT

In this paper, we describe the best practices for digitizing speech recordings for archival purposes. Many current guidelines do not give adequate rational and are not specific enough about audio digitizing practices for spoken word digital repositories. By analyzing frequency response, dynamic range, formant bandwidths, noise, and psychoacoustic, and perceptual quality, spectral evaluation of recordings was used to develop digitization best practices. Given the available technologies and the demands of archival preservation, weighed against expense and storage, the following digitization practices were determined acceptable for digitizing spoken word resources; sample rate: 96,000 Hz; bit-depth: 24-bit; and oversampling delta-sigma A/D converter with dither added prior to sampling. Best practices for digitizing equipment and digital format are also described.

Keywords
digitization, audio, linguistics, audio recordings, speeches, preservation, audio digital repositories

INTRODUCTION
The issue of audio digitization has been of interest to many organizations building repositories of digital resources. Several of these organizations have completed extensive research in this area and published papers outlining suggested best practices for audio digitization. Although we have found these resources to be helpful, we have also found that they include information that is problematic and can make it difficult for organizations to decide on best practices. In particular, while these resources often contain good information, they usually lack depth in the rationale for particular choice of practices and do not focus on the demands of spoken word resources.

Cohen, for example, relies on the putative influence of supersonic harmonics on brain function as one of the primary reasons for choosing the sample rate of 192 kHz [1]. While the standard be an excellent goal for some collections, the rationale does not support an investment in high-end equipment and calls for a practice that would be difficult for most institutions to adopt. She is also primarily concerned with music resources. Similarly, Schuller outlines many important considerations but does not give any specifics beyond the statement that the digitization hardware must be of a high quality [9]. Moreover, some resources recommend practices that are fundamentally flawed. One of the most common of such practice is digitizing analog materials directly from a stand-alone CD burner to an audio CD (e.g., Colorado 2001[2]), a practice that can often lead to poor digital sound quality. Finally, few of the publications contain empirical data to support the recommended practices. Impressionistic testing is often used to assess the quality of the digitization process in place of a thorough acoustic analysis of the files obtained in the process. Thus, the following paper extends and enhances the work published on digitizing audio for archival purposes by both focusing on spoken word resources and outlining rationales for particular practices. The practices we recommend are based on a substantial body of empirical work and are meant to be applicable across a wide variety of analog audio resources.

The National Gallery of the Spoken Word is a five-year research project that focuses on the ability of digital libraries to preserve and deliver aural resources in a wide variety of useful formats. It is funded by Michigan State University, it's partnering institutions, and the National Digital Library Initiative Phase II spearheaded by the National Science Foundation. As part of this project MATRIX is building a digital repository of spoken word resources entitled Historical Voices. The current collection spans several decades of recording history. A large portion of it dates back to the very first days of recording; several recordings contain some of the most important moments in both US and world history. The preservation, storage, and online delivery of these resources are among the primary goals of MATRIX, and current technologies provide remarkable tools to achieve those goals. The first step in the process is the conversion of existing analog speech signals into the digital format. The purpose of this paper is to present the digitization process adopted at MATRIX along with a set of guidelines, or best practices, that MATRIX deems recommendable for other, similar projects.

Recordings

Archives, libraries, museums must deal with broad range of recording. From 1920s wax cylinders to contemporary DAT recordings, each recording type can possesses a different set of characteristic that can vary greatly: the original recording medium, recording conditions, bandwidth, frequency response, and so on. For our purposes, we have had to develop and instantiate digitization practices applicable to the entire Historical Voices collection. To achieve this goal we have been motivated by one primary factor: to provide the most truthful digital representation of the analog original.

To describe the different types of recordings that we deal with is beyond the scope of this paper. Instead, we will focus on two typical, yet quite different cases: a 1930s phonograph recording and a 1990s high fidelity, reel-to-reel recording. We have looked at several variables that are useful in describing a speech recording, such as frequency response, dynamic range, formant bandwidths, noise, and psychoacoustic, and perceptual quality. In doing so we have found that a thorough spectral evaluation of recordings is instrumental in developing digitization best practices.

Figure 1 below shows a spectrogram and waveform of a 1930s recording of a phrase "*is serial number one fifty eight*" digitized at a sample rate of 48,000 Hz and a 16-bit resolution. The spectrogram range is 0 - 24,000 Hz, which represents the entire frequency response of the digital audio file. As can be seen, the bandwidth of the actual speech signal has its upper limit at around 5,000 Hz. This is very characteristic of such, limited-bandwidth recordings. The relatively narrow bandwidth of this recording is a direct result of the original recording conditions, and the hardware used for recording and playback. At that time, recordings were not able to capture acoustic information above, roughly, the 5,000-6,000 Hz limit. What we also see in the spectrogram is two kinds of noise (i.e. non-speech information). The first one is located close to the bottom of the spectrogram (low frequency "rumble"); the other is spread evenly across the entire frequency range. It is safe to assume that the low frequency noise is a direct result of the hardware used to record and read the original token, while the other type of noise is most-likely noise added during the analog mastering of the recording (reel-to-reel transfer).
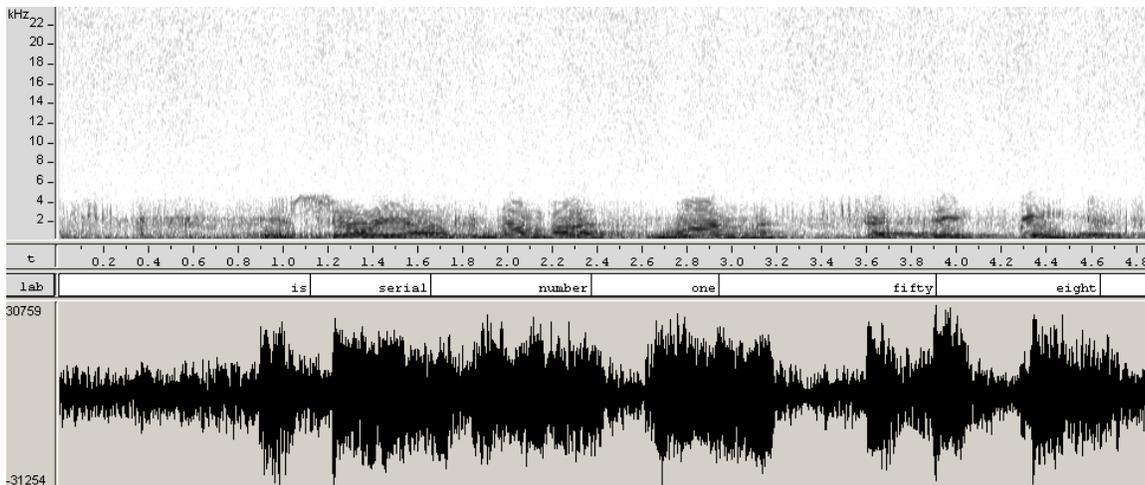

Figure 1 Spectrogram of a 1930s phonograph recording of the phrase "*is serial number one fifty eight*" digitized at 44,000 Hz/16-bit

Figure 2 shows a spectrogram of the phrase "*They got to the mall at exactly ten o'clock*" recorded in the 1990s on a high-fidelity reel-to-reel recorder with a professional grade dynamic microphone (Shure SM58) in a sound-proof, studio environment. Here we have a relatively clean recording, with a bandwidth and dynamic range far exceeding that of the file in Figure 1. This recording captures the entire

frequency response of an adult female talker, showing even the highest-frequency fricative sounds in the 17,000 Hz range. We do not see much extraneous information in the signal, either. There is a slight high-frequency hiss at around 16,000, which is probably the result of the noise generated by the microphone preamplifier, quite typical of this kind of recording.
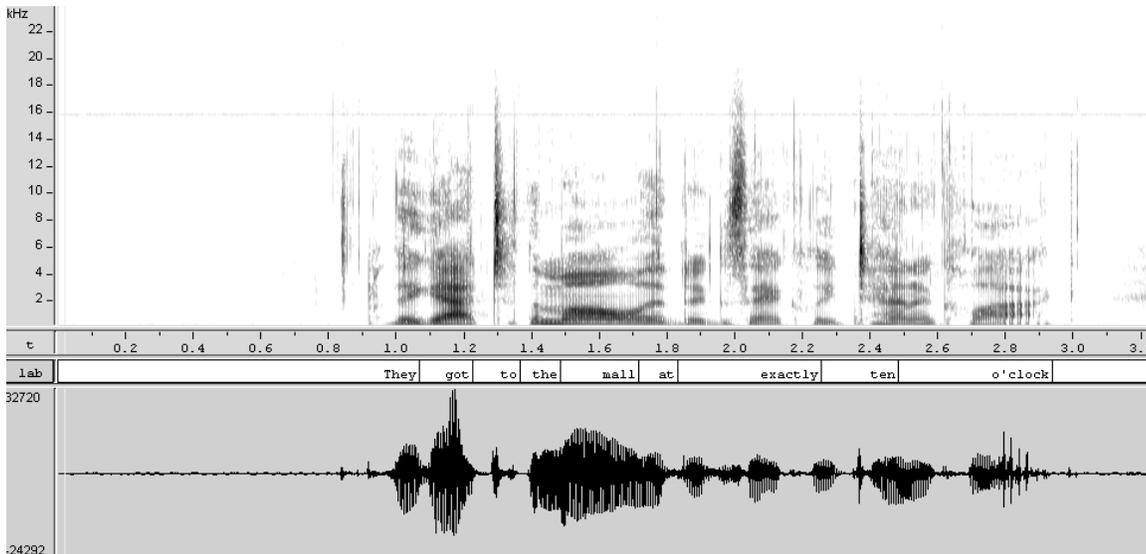
Figure 2 Spectrogram of a 1990s reel-to-reel recording of the phrase "*they got to the mall at exactly ten o-clock*" digitized at 44,000 Hz/16-bit

Figures 3 and 4 illustrate some of the most important contrasts between the two files even further. The waterfall plots give us a very good sense of formant values and bandwidth. As we can see in Figure 3, there is a low frequency spectral prominence across the entire file. It is located somewhere at around 1,000 Hz. There is another, much smaller peak running at around 2,500 Hz. Between 1050 and 1250 ms, we see a rise in the second peak. In a clean speech recording, we would have to conclude that we see two formants, f1 and f2, of a fronting and raising diphthong-like sound; however, this certainly is not what we hear when we play the clip. We hear an utterance that contains vowels, fricatives, and liquids; each normally

having quite different spectral characteristics. The spectrogram shows us much reduced spectral information, where noise cannot be effectively separated from speech, and where acoustic characteristics of otherwise distinct speech sounds merge into a continuous two-formant pattern. Not surprisingly, the recording is almost unintelligible.

Figure 4 shows us a much different picture. We see a plot that is rich and varied in spectral information. We see very strong, narrow-bandwidth formants, high-frequency turbulence of the fricatives, low-frequency nasal murmur, and well-defined formant transition with no interference from any harmonic noise. The recording sounds perfectly clear and intelligible.
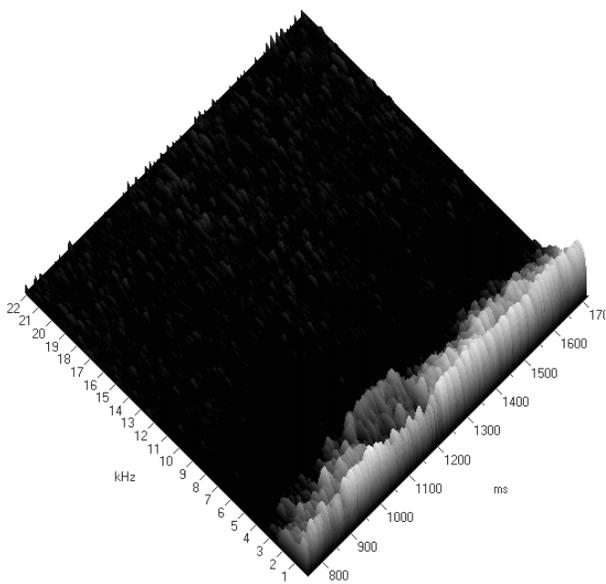

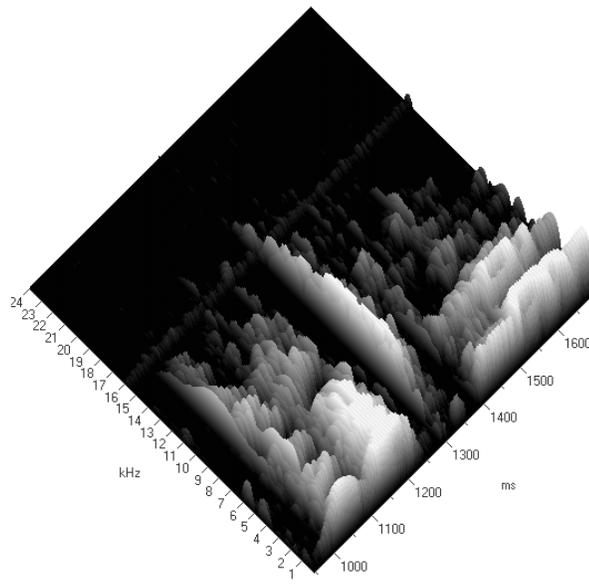**Figure 3** Waterfall plot of "*is serial*"

Figure 4 Waterfall plot of "*they got to the mall*"

Let us now take a closer look at noise characteristics in each of the files. Both Figures 5 and 6 show power spectra taken at a non-speech point (noise) and at a vowel nucleus. The spectra below (gray) represent the noise print. As we can see in Figure 5, the noise spectrum is almost undistinguishable from the vowel spectrum. The difference
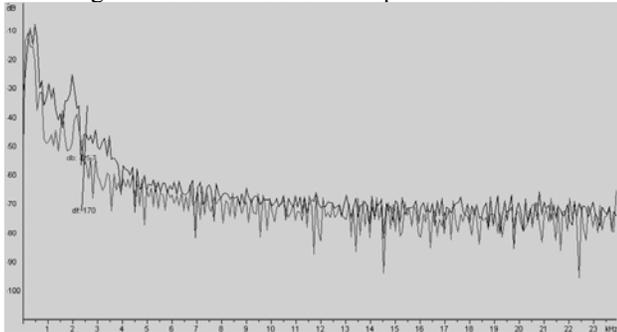
in power between the signal and noise is only about 15.7 dB. Figure 6, again, shows that speech and noise are well separated both in terms of their spectral characteristics and power, with Signal-to-Noise ratio of about 86.2 dB, which is quite impressive even for this kind of recording.
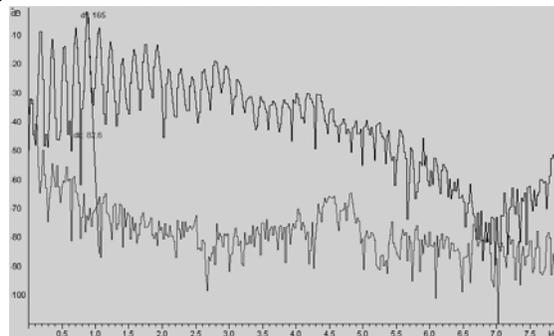

Figure 5 Noise and speech spectra of a 1920s phonograph recording

Figure 6 Noise and speech spectra of a 1990s reel-to-reel recording

The two digital audio files that we have just looked at represent probably two extreme ends of the NGSW continuum. However, it is crucial for NGSW to be able to develop standards that would uniformly apply to both samples, as well as to any other recordings that come in between. The two samples illustrate the most common challenges of the NGSW digitization practices. Having looked at both files in details, we now have a better understanding of those challenges, which can, no doubt, help us accomplish our digitization goals more effectively.

Choosing the standards

One important conclusion that we can already make at this point is that our digitization standards should be able to faithfully represent acoustic signals that are varied in dynamics and have a frequency response of, minimally, 0-20,000 Hz, which happens to coincide with typical human hearing range. Such minimal requirements can be accomplished, in theory, by using standard, "CD quality" settings of a sample rate of 44,100 Hz and a 16-bit resolution. However, it is also true that higher specifications, especially 24-bit word length, help capture more detail and minimize digitization noise and distortion. Before we decide to settle on any practice, let us first consider some basic, yet important, aspects of the digitization process that we already know.

Some basic concepts

Digitization

Digitization is a process of converting an analog, continuous, waveform to digital form by an *Analog-to-Digital converter* (*ADC*). The sound pressure waveform is a continuously varying signal. It changes from instant to instant, and as it changes between two values, it goes through all values in between. Using an ADC, the signal is sampled at a set of equally spaced times. The sample value is a number equal to the signal amplitude at the sampling instant. Computers always represent numbers internally in the binary number system, where each binary position can

be one of the two binary digits, 0 or 1. Each binary digit is called a *bit*, and a collection of eight bits is called a *byte*.

The astonishing thing about the ADC process is that, if done properly, every detail of the original signal can be captured. The original continuous waveform can be reconstructed exactly and, more importantly, powerful digital signal processing can be applied to the digital representation of the signal. Such a representation can then serve a variety of purposes, such as streaming, watermarking, speech recognition, and so on. A vast number of effects, such as noise reduction, bass emphasis, normalization, compression, etc. can be achieved by applying a mathematical transformation called "a digital filter." After processing, the resulting continuous signal can be reconstructed without any degradation. The process of converting from digital back to a continuous signal is called *digital to analog conversion*, or *DAC*.

When any signal is transmitted over a communications medium, it is inevitably distorted to some degree. The received signal, therefore, differs to some extent from the originally transmitted signal. For an analog signal, even small amounts of distortion lead to noise that is often impossible to eliminate. Examples of noise include "hiss," "static," and the poor quality (limited bandwidth) of signals transmitted over telephone circuits. For digital signals, the same signal distortion factors are present. However, since the signal at any instant of the represents a number, rather than an analog signal level, it is necessary only to unambiguously recognize the number that was transmitted. This choice is substantially eased by the fact that digital signals are binary. As long as noise is small enough so that the choice between these two outcomes is not confused, the received signal will represent the same numerical value as the transmitted signal. In fact, additional coding techniques, such as parity check, are used to detect and correct errors that might occur in the process. The result is unparalleled quality and robustness of digital signals.

The sampling theorem

It is not obvious that an exact reconstruction of an analog signal should be possible, since a complete continuous signal is replaced by a finite number of samples taken at equal time intervals. The problem is to have complete information between the samples. The answer lies in the mathematical result called the *Sampling Theorem*. In short, the sampling theorem states that if a band-limited signal is sampled at a rate greater than twice the highest frequency in the signal (the so-called *Nyquist frequenc [y]*), no information is lost and the original signal can be exactly reconstructed from the samples. Acoustic signals that humans can hear lie in a limited range of about 20 to 20,000 Hz. Thus, intuitively, in order to reconstruct exactly the original analog signal, one should use the sample rate of at least 40,000 Hz. This is, indeed, true. It is often recommended to use the practice of CD quality that has a sample rate of 44.1 KHz for all of spoken word digitization projects. It is important to note that sampling at a rate lower than twice the Nyquist frequency results in a phenomenon called *aliasing*, whereby the resulting digital signal lacks a significant number of bits of information and it may be severely distorted by the aliased components, making it audible as noise. Finally, even though the speech signal usually does not contain any information above 7 kHz, and, theoretically, the sample rate of 16 KHz should be sufficient to capture all details of the signal, it is nevertheless recommended to use the sample rate of 44.1 kHz for all of AD conversion for archival purposes.

Aliasing

Simply put, aliasing is a kind of sampling confusion that can occur during the digitization process. It is a direct consequence of violating the sampling theorem. The highest frequency in a sampling system must not be higher than the Nyquist frequency. With higher audio frequencies, the sampler continues to produce samples above Nyquist at a fixed rate, but the samples will create false information in the form of alias frequencies. In practice, aliasing can and should be overcome. The solution is rather straightforward. The input signal must be band-limited with a low-pass (anti-aliasing) filter that provides significant attenuation at the Nyquist frequency. The most "archetypal" anti-aliasing filter will have "brick-wall" characteristics with instantaneous attenuation and a very steep slope. This results in unwanted ringing-type effects and should be avoided. In practice, a system should use an oversampling (see below) A/D converter with a mild low-pass filter, high initial sampling frequency, and decimation processing to prevent output sampling frequency.

Quantization

For the sampling theorem to apply exactly, each sampled amplitude value must exactly equal the true signal amplitude at the sampling instant. Real ADCs do not achieve this level of perfection. Normally, a fixed number of bits (binary digits) is used to represent a sample value. Therefore, the infinite set of values possible in the analog signal is not available for the samples. In fact, if there are R

bits in each sample, exactly 2^R sample values are possible. For high-fidelity applications, such as archival copies of analog recordings, 16 bits per sample (65, 536 levels), or a so-called *16-bit resolution*, should be used. The difference between the analog signal and the closest sample value is known as *quantization error*. Since it can be regarded as noise added to an otherwise perfect sample value, it is also often called *quantization noise*. The effect of quantization noise is to limit the precision with which a real sampled signal can represent the original analog signal. This inherent limitation of the ADC process is often expressed as a *Signal-to-Noise ratio* (SNR), the ratio of the average power in the analog signal to the average power in the quantization noise. In terms of the dB scale, the quantization SNR for uniformly spaced sample levels increases by about six dB for each bit used in the sample. For ADCs using R bits per sample and uniformly spaced quantization levels, SNR = 6R - 5 (approximately). Thus, for 16-bit encoding about 91 dB is possible. It is 20 to 30 dB better than the 60 dB to 70 dB that can be achieved in analog audio cassette players using special noise reduction techniques.

Dither

A well-design digitization system must be able to suppress any audible effects of its quantization error. One could increase the number of bits in the quantizing word, which would result in a decrease in error amplitude of 6 dB per each additional bit. However, this is a rather uneconomical solution, as many bits are necessary to reduce effectively the audibility of the quantization error. Additionally, the error will always be significant with low-level signals, which are quite common. Dither seems to offer a more elegant solution. Dither is a small amount of noise added to the audio signal before sampling. This causes the audio signal to shift with respect to quantization levels. Quantization error is thus decorelated from the signal and the effects of the quantization error become negligible. Dither does not prevent the quantization error; instead, it allows the system to encode amplitudes smaller than the least significant bit.

Oversampling

Oversampling is another technique aimed at improving the results of the digitization process. As noted above, a brick-wall filter may produce unwanted acoustic effects. In oversampling A/D conversion, the input signal is first passed through a mild low-pass filter, which provides sufficient attenuation at high frequencies. To extend the Nyquist frequency, the signal is then sampled at a high frequency and quantized. Afterwards, a digital low-pass filter is used to reduce the sampling frequency and prevent aliasing when the output of the digital filter (e.g. an interpolating, phase linear "FIR" filters) downsampled to achieve the desired output sampling frequency (e.g., 44,100 Hz). In addition to eliminating unwanted effects of a brick-wall analog filter, oversampling helps achieve increased resolution by extending the spectrum of the quantization

error far beyond the audio base-band, rendering the in-band noise relatively insignificant.

What "Best Practices" should we choose?

The choice of appropriate digitization best practices for Historical Voices is influenced by two distinct factors - the technological and the archival. Each of these factors poses a different set of questions and challenges to the project. When deciding on particular specifications of sampling rate, quantization, hardware, etc., our primary technological goal is to provide a digital copy that closely matches the analog original. In the technical sense, we need to establish a process that, minimally, reconstructs the entire frequency response of the original while adding as little of the so-called "digital noise" as possible. To achieve this goal, it seems to be sufficient to use the 44,100 Hz sampling rate and a 16-bit resolution. The former ascertains that we capture the entire audible frequency range of the original (see Nyquist theorem above), while the latter, gives us a fairly good, 96 db SNR (signal to noise ratio). From the archival standpoint, it is our desire to preserve as much information (both speech and non-speech) of the original as possible. The technical specifications mentioned above promise to fulfill this goal.

In the ideal world, the discussion on choosing appropriate standards could end right here. However, there are a few other issues involved. Current technology makes it possible to use higher sampling rates and resolution rather inexpensively. One could fairly easily sample at 96,000 Hz and a 24-bit resolution. This would result in a much increased frequency response of the digital waveform –

from 0 to 48,000 Hz, and a dramatically improved SNR of 144 dB. At first glance, this technology appears to provide a very simple solution. After all, it is our goal to preserve as much information as possible, and using such high specifications does exactly that. The digital audio file captured at 96,000 contains over twice as much information as the one sampled at 44,000 Hz. Also, as we have also noted before, using an oversampling A/D converter can dramatically minimize the unwanted effects of the quantization error. Given the technological potential that we have at our disposal, the choice of digitization standards appears to be simple: use 96,000 Hz sampling rate and a 24-bit quantization. However, is the extra amount of information really worth preserving?

A common argument against specifications higher than the so-called "CD quality" 44,100 Hz/16-bit are based on the fact the analog speech recordings are narrow-bandwidth, with no significant high-frequency information present. The argument is additionally strengthened by the fact that even if there is high-frequency information present in the recording, it makes little sense to capture it, as the human ear is not capable of hearing anything above 20,000 Hz. Figures illustrate the point quite clearly. We have taken a typical analog cassette tape recording at digitized it at 44,000 Hz/16-bit and 96,000 Hz/24-bit (a new, audio DVD standard). We can see quite clearly that the lower standards, with the Nyquist frequency at 22,000 Hz, seem to capture adequately all of the speech frequencies present in the recording. The 96,000 Hz file has a wide band of virtual silence above, roughly, 15,000 Hz.
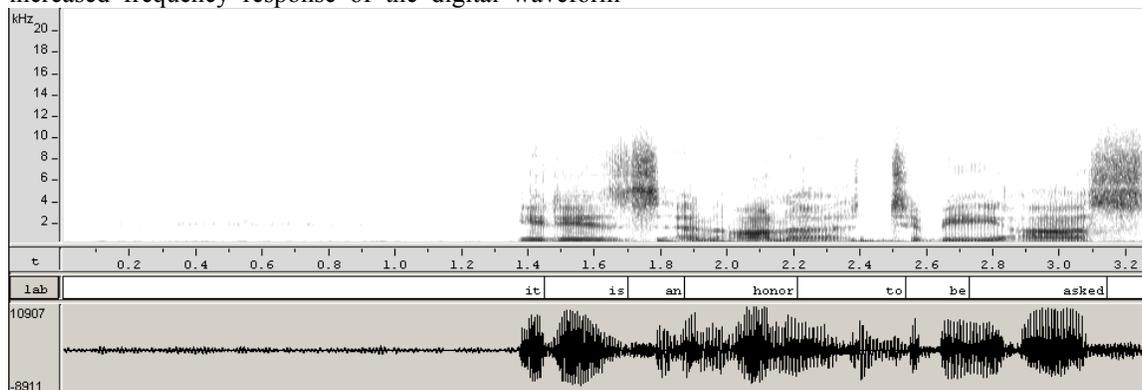

Figure 7 "*it is an honor to be asked*" digitized at the sample rate of 44,100 Hz.
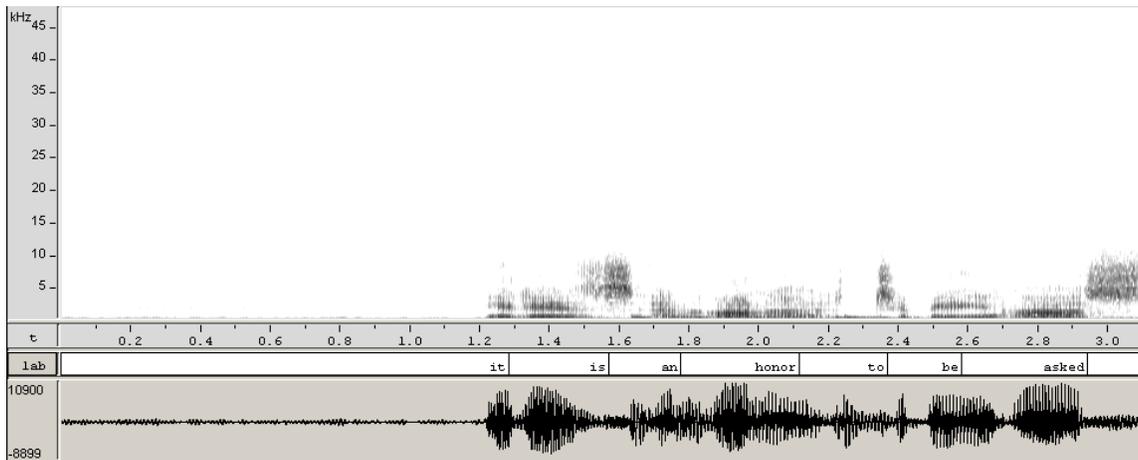
Figure 8 "*it is an honor to be asked*" digitized at the sample rate of 96,000 Hz.

The majority of Historical Voices recordings can be easily characterized as limited bandwidth. First of all, most of them are analog speech recordings. Speech, as such, is a rather narrow-bandwidth medium. In fact, the highest frequency speech sounds, voiceless fricatives, such as /th/ and /f/ have an upper frequency of about 6,000 – 7,000 Hz, depending on the age and gender of the speaker. Thus, any information in the channel that is above the 7,000 Hz is, potentially, non-speech information. Of course, from the archival standpoint, we may want to preserve this information, too. That could be any "environmental," ambient sound present in the recording. Second, most Historical Voices recordings have been mastered on a limited-bandwidth analog equipment that was not capable of capturing any of the high frequency information in the first place. The Nyquist frequency in such recordings varies, but it hardly ever goes beyond the audible range of 20-20,000 Hz. However, what we have also seen earlier (Figures 2, 4, and 6), some Historical Voices resources have characteristics that go beyond the most common analog recordings, not to mention materials that are born digital, such as DAT, hard-disk recordings and audio DVDs, which already constitute a decent portion of our collection and are very likely to dominate all future oral history recordings. We should also consider the possibility that our collections will continue to expand and that more types of recordings, including live music, TV broadcasts, round-table talks, will be added to the repository. As noted before, our digitization standards must be able to accommodate all of such demands, including the emerging hard-disk recording standards, as well as the audio

In light of the above arguments, Historical Voices has chosen the following digitization best practices for spoken word resources:

- Sample rate: 96,000 Hz

- Bit-depth: 24-bit

- Oversampling delta-sigma A/D converter with dither added prior to sampling.

Given the spectral evaluation of recordings, available technologies, and the demands of archival preservation -- weighed against expense and storage -- these practices have more than acceptable for digitizing spoken word resources.

Additional standards
The WAV file format

The WAV file format is recommended for storing digital versions of speech recordings. WAV files are uncompressed, thus preserving all bits of information recorded in the AD process. It is also widely used and easy to process and convert to a variety of streaming formats.

The WAV file format is a variant of the RIFF format for data interchange between programs. This format was designed so that data in a file is broken up into self-described, independent "chunks." Each chunk has a prefix, which describes the data in that chunk. The prefix is a four-character chunk ID that defines the type of data in the chunk, followed by a 4-byte integer, which is the size of the rest of the chunk in bytes. The size does not include the 8 bytes in the prefix. The chunks can be nested. In fact, a RIFF file contains a single chunk of type "RIFF," with other chunks nested inside it. Therefore, the first four bytes of a WAV file are "RIFF," and the four bytes after that contain the size of the whole file minus 8 bytes. After the RIFF header is the WAV data, consisting of the string "WAVE" and two important chunks: the format header and the audio data itself. There may also be other chunks in a WAV file that contain text comments, copyrights, etc., but they are not needed to play the recorded sound. The header and the audio data can be easily separated to facilitate migration to any other format, if necessary.

Hardware

A/D converter

It is recommended a digitizing system use a professional-level, hardware platform, preferably integrated with the

software that includes reliable signal acquisition, analysis, and playback. Such hardware should include the following features:

- 24-bit quantization (with a 16-bit option)
- 96,000 Hz maximum sample rate
- oversampling capability
- user-selected anti-aliasing filters
- a wide assortment of sample rates
- XLR (balanced) and RCA (unbalanced) inputs
- high gain preamplifier to accommodate low-level input signal levels
- digital (minimally, SPDIF and AES/EBU) inputs and outputs
- user-selected AC/DC coupling for all channels
- real-time indicator of signal overloading

PCI digital I/O Card

The analog signal is converted to digital by the external module and then captured by an internal PCI sound board. It is important to avoid using standard multimedia sound cards. Thus it is recommended that a digitizing system use a professional-level sound card that meets the following specifications:

- PCI Interface
- 8 to 24 bit resolution
- variable sample rates, including 11.025kHz , 44.1kHz, and 96 kHz
- S/PDIF digital in/out via gold-tipped RCA connectors

USB and IEEE 1394 (FireWire) A/D converters

Recently, a small number of stand-alone USB and IEEE 1394 (FireWire) compatible A/D converters have emerged on the marked. At first glance, they seem to present a viable alternative to the system described above. However, this type of hardware and the accompanying drivers need to be thoroughly tested before they can be reliably recommended.

Computer

- A top-of-the-line PC or Mac that meets the following specifications:
- 800 MHz or higher processor
- 512 MB of RAM

- High-capacity, fast HDD (preferably 2 Ultra 160 SCSII drives with a RAID controller card)
- Fast CD Writer
- Fast DVD Drive
- IEEE 1394 card for digital in/out (optional)

Software

A sophisticated sound editor is recommended. It should meet the following specifications:

- 96,000 Hz and 24-bit capability
- Multiple Document Interface for editing multiple files in one session
- Unlimited file size editing
- Real-time graphs (amplitude, spectrum, bar, and spectrogram) (optional)
- Real-time fast forward and rewind playback
- Numerous digital filters and effects
- Support for many file formats (WAV, AU, IFF, VOC, SND, MAT, AIFF, MP3, OGG, and raw data) and a batch converter.

REFERENCES
1. Cohen, E. Preservation of audio. *Folk Heritage Collections in Crisis*. (18 May 2001). Available at http://www.clir.org/pubs/reports/pub96/contents.html

2. Colorado Digitization Project - Digital Audio Standards. Available at http://coloradodigital.coalliance.org/digaudio.html

3. Harris, S., Techniques to measure and maximize the performance of a 120-dB, 24-Bit, 96-kHz A/D converter integrated circuit. *AES preprint* 4530, New York, (September, 1997).

4. International Association of Sound Archives. The safeguarding of audio heritage: Ethics, principles, and preservation strategy. (5 December 2001). Available at http://www.llgc.org.uk/iasa/iasa0013.htm

5. Johnson, K., *Acoustic and Auditory Phonetics*, Blackwell Pub, 1997.

6. Nyquist, H., Certain Topics in Telegraph Transmission Theory, *Trans. AIEE*, vol. 47, no. (2, April, 1928).

7. Pohlmann, K.C, *Principles of Digital Audio*, McGraw-Hill, 2000.

8. Rodman, R. D, *Computer Speech Technology*, Artech House Signal Processing Library, 1999.

9. Stevens, K. N., *Acoustic Phonetics*, MIT Press, 1999.

10. Woram, J. W., *Sound Recording Handbook*, Howard W. Sams, 1989.