# Best Practices in the Acquisition, Processing, and Analysis of Acoustic Speech Signals.

Bartek Plichta, Michigan State University

This article is intended to assist researchers in implementing several important aspects of audio technology in the field, studio, and research lab. It presents a set of best practices in the recording, processing, and analysis of acoustic speech signals. The author has no commercial interest in recommending particular pieces of hardware or software. Some brand names and manufacturers have been mentioned to help researchers find appropriate tools more easily.

## 1. Recording

### 1.1. Technical basics

The purpose of recording is to capture the best possible signal. Regardless of the recording situation, one should consider at least three important acoustic parameters: frequency response, dynamic range, and signal-to-noise ratio (SNR). Frequency response describes the range of frequencies captured in the recording, measured in Hertz (Hz). Dynamic range is the ratio of the loudest to the softest part of the signal, measured in Decibels (dB). SNR is the ratio of signal amplitude to the amplitude of noise usually generated by the circuit, measured in Decibels (dB). In the audio digitization process, signal-to-noise ratio refers to the ratio of the maximum signal power to the quantization noise generated by the analog-to-digital converter. Table 1 summarizes the frequency response and dynamic range characteristics of speech as well as some common recording media.

| Medium | Frequency response: | Dynamic range: |
|---|---|---|
| Speech | < 10000 Hz | 30-40 dB |
| Analog tape | < 10000-15000 Hz | 45 dB |
| DAT (Digital Audio Tape, 16 bit, 48 kHz) | < 24000 Hz | 96 dB |
| Digital telephone (ISDN) | < 4000 Hz | 48 dB |

Table 1 Frequency response and dynamic range characteristics of speech and common recording media.

As we can see, the speech signal is not particularly demanding in terms of frequency response and dynamic range. In theory, a modern, high-quality analog or digital recording device should be able to capture the speech signal fairly accurately. The only exception is the phone line (analog or digital), which lacks detail in frequency and adds substantial amounts of noise to the signal. However, we can also see that digital recording devices, such as DAT and hard disk recorders, produce a considerably richer signal; one whose frequency spans across the entire audible range and contains enough amplitude variation to allow for very detailed acoustic analysis.

### 1.2. Recording hardware setup

The recording process requires, minimally, the following set of hardware devices listed in Fig 1. Each of the types of devices used in the process may have a crucial impact on the overall quality of the acquired signal.

| Typical analog recording setup |
|---|
| microphone – (phantom power) – pre-amplifier – analog recorder – analog tape |
| Typical digital recording setup |
| microphone – (phantom power) – pre-amplifier – A/D converter – HDD or DAT |

Figure 1 Minimal recording hardware setups

### 1.3. Microphones

Two basic microphone types are most typically used for recording speech. The dynamic microphone has a diaphragm that consists of Mylar plastic that has a finely wrapped coil of wire (so-called “voice coil”) attached to its inner face. This coil is suspended within a strong magnetic field. Whenever a sound wave hits the diaphragm, the coil is displaced in proportion to the amplitude of the wave, causing the coil to cut across the lines of magnetic flux supplied by the permanent magnet. Since the mass of the diaphragm and the coil is quite large, compared to the pressure changes in the sound wave, the dynamic microphone may not respond well to sharp transient sounds and it may fail to record minute changes in voice intensity. This does not mean that dynamic microphones should not be used for speech recording. On the contrary, there are several low-cost, rugged microphones, such as Shure SM58 or Shure SM48 (once recommended by Kay Elemetrics), that can be used successfully for many speech recording applications. Dynamic microphones do not require any external power supply, which makes them a good match to several field recorders (e.g., Marantz PMD 222).

The condenser microphone works on a different principle. A thin plastic diaphragm coated on one side with gold or nickel is placed at a close distance from a stationary backplate. Once a polarizing voltage (from a 48 V phantom power supply) is applied to these plates, the two surfaces create capacitance that varies as the diaphragm moves in response to a sound wave. Since the diaphragm is very light, the response of the condenser microphone is very accurate, often producing a recording that is extremely rich in both frequency and dynamic response. Condenser microphones are usually more fragile than dynamic microphones and require a 48 V phantom power supply. Some of them can use battery packs, but some rely on an external power source. This makes the condenser microphone a bit more cumbersome to use in the field, though several DAT and HDD recorders have an on-board phantom power supply.

### 1.3.1. Frequency response.

Each microphone type has a unique frequency response. It is important to remember that many manufacturers tailor a microphone's frequency response to accentuate particular parts of the spectrum to function optimally within a specific recording application. For the purposes of acoustic analysis, one should always choose microphones with a wide and flat frequency response curve.

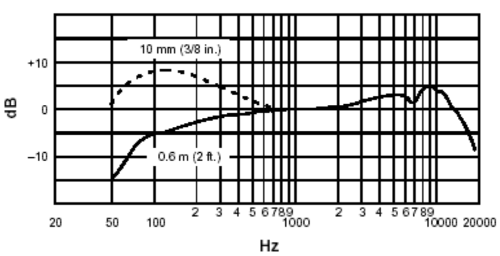



Figure 2 Frequency response of a Shure Beta 87a microphone. Note the proximity effect at 10 mm.

### 1.3.2. Polar patters

The microphone's polar pattern should play a crucial role in choosing a microphone for a specific recording application. The polar pattern is a plot of the sensitivity of a microphone as a function of the angle around that device (Fig 3). There are several common polar pattern types used in microphones today. The omnidirectional microphone records sound equally from all directions. Such microphones are most commonly as built-in or lavalier types. They seem to be very good for recording interviews, though their 360-degree pick-up range, introduces too much noise to the signal for it to be used reliably in acoustic analysis.

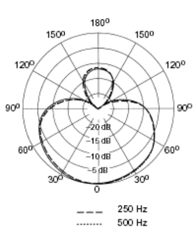



Figure 3 Cardioid polar patter of a Shure Beta 87a microphone

The cardioid (heart-shaped) pattern is most sensitive to sounds coming from the front. It is 6 dB less sensitive to sounds from 90 degrees to the sides, and, in theory, is completely insensitive to sounds coming from the rear. The most important attribute of a cardioid (directional) microphone is its ability to discriminate between direct sounds (coming from the direction in which it is pointed) and reverberant, unwanted sounds from all other directions. This type of polar pattern usually produces signals that are substantially less noisy than those captured with an omnidirectional microphone (Fig 4).

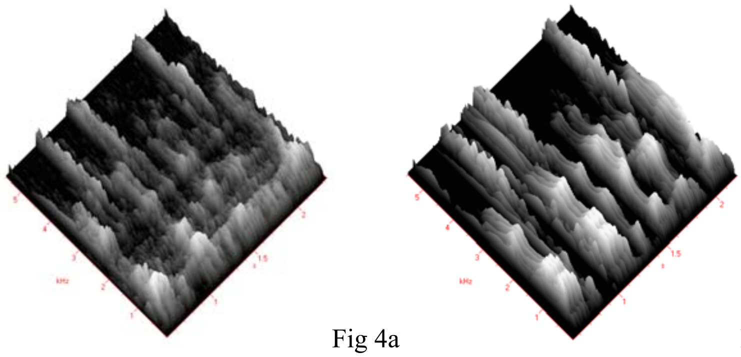


Fig 4a Fig 4b

Figure 4 Waterfall plot of a recording of the phrase "Bob was positive that his wife"
Figure 4a - recording done with a built-in, omnidirectional microphone. Note the amount of room noise and lack of detail in spectral prominences.
Figure 4b - recording done with a head-worm, directional microphone. The plot shows negligible amounts of noise and superb detail in the speech spectrum

### 1.3.3. Proximity effect

Usually, high-quality speech recordings require the sound source to be fairly close to the microphone's diaphragm. This may trigger a so-called proximity effect (Fig 2). Proximity effect is the increase in the low-frequency sensitivity of a microphone when the sound source is close to it. This is particularly true of cardioid, directional microphones. To counter that, most high-end directional microphones use a low-frequency roll-off filter to restore the response to its flat, natural balance. Some microphones have a user-selectable switch to control the filter. The proximity effect may be responsible for speech spectra showing emphasis in the low-frequency range, around the first and second harmonics.

### 1.3.4. Cabling and phantom power

Cabling is a common cause of recording problems. In order to avoid noise (60 Hz hum) and phase problems, it is recommended to use professional quality balanced XLR (two conductors for the signal with neither connected to the shield) cables to connect the microphone to the pre-amplifier. If the pre-amplifier does not have balanced XLR inputs, one should use a balanced to unbalanced transformer (Ebtech Line shifter, CP8201 In-Line Transformer). The transformer's primary side matches the impedance of the microphone and is balanced, while the secondary is unbalanced and has high impedance that matches most unbalanced pre-amplifier inputs. Condenser microphones require a 48 V phantom power. While some microphones, such as AKG C1000 can work from an internal battery source, others require an external phantom power supply, which many good pre-amplifiers and mixing consoles come equipped with.

## 1.4. Recording techniques

Much of the success of a speech recording depends on the recording environment and microphone placement. Ideally, speech recordings should take place in soundproof studios or labs. If those are not available, one should try to find a relatively quiet room with as little low-frequency noise as possible. Most typical sources of low-frequency noise include 60-Hz hum from electrical equipment, heating and air-conditioning ducts, elevators, doors, water pipes, computer fans, and other mechanical systems in the building. If possible, those devices should be switched off during recording. Fig 4 illustrates the spectrum of a typical ambient room noise. The low-frequency prominence of this kind of noise may interfere with acoustic analysis of the fundamental frequency and the first formant. It is also quite difficult to filter out any such extraneous noise without removing information from the speech signal itself. High frequency noise should also be avoided, though it can be more easily filtered out of the signal before analysis.

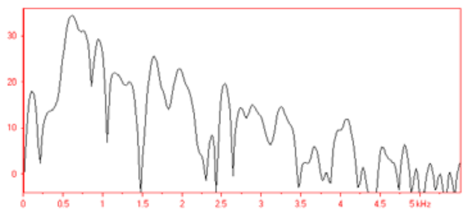


Figure 5 Spectrum of typical room noise. Note the prominence at around 600 Hz.

The placement of the microphone directly affects the intensity of the recorded signal as well as the signal-to-noise ratio. The inverse square law guarantees a loss of approximately 6 dB per doubling of distance from the sound source. A typical hand-held microphone is usually placed at a distance of 30 cm or so from the talker's lips. This, relative, to a close placement (say, 4 cm) represents the loss of about 18 dB and an increased possibility of noise leaking into the recording. For this reason, it is recommended to use a head-mounted condenser microphone (such as AKG C410 or AKG C420) to maintain a close, and constant distance to the source. Speech signals acquired this way are characterized by a high SNR and a broad range of intensity. It may also be useful to use a linear phase high-pass rumble filter (60 Hz cutoff and 24dB/octave attenuation), unless low-frequency components are expected in the signal.

### 1.5. Pre-amplifiers

The main function of a pre-amplifier is to accept a very low-level signal (such as that from a microphone) and amplify it without adding noise. Good pre-amplifiers are not easy to build, as they have to be immune to all kinds of potential noise and signal distortion. It is, therefore, important to use the best possible pre-amplifier, particularly one that has a fairly high gain, broad dynamic range, high SNR, phantom power, and balanced XLR inputs. Good field pre-amplifiers are particularly hard to find. Sound Devices MP2, USB Pre, and Shure FP 23 are among some of the best, yet affordable field pre-amplifiers. The gain on the pre-amplifier should set relatively high, but the speaker's voice amplitude range should be tested first to avoid signal overload. If intensity measurements are not indented, a soft compressor-limiter may be used to maximize amplitude and help prevent signal clipping.

### 1.6. Recording devices.

There are several types of recording devices used today. Some of the most common include analog cassette recorders, digital audio tape (DAT) recorders, minidisk recorders, and hard disk recorders. Some of these devices, especially field recorders, come with a built-in pre-amplifier and two independent input channels. When recording in the studio or lab, a hard disk recorder is highly recommended. In the field, however, other types of devices may also be considered.

#### 1.6.1. Analog recorders

As we saw in Table 1, the analog tape is able to provide the necessary frequency response and dynamic range to capture a fair amount of detail in the speech signal. In fact, if a premium-grade microphone and pre-amplifier have been used, a professional cassette recorder, such as Marantz PMD 222, is capable of capturing fairly high-quality acoustic signals. Cassette recorders are usually quite durable and inexpensive. They also operate on regular alkaline batteries and use standard audio cassettes, which are inexpensive and widely available. However, analog recorders typically have noisy transport mechanisms and offer no time code, which makes logging and analysis more difficult. In addition, the analog tape is a rather fragile medium and should be avoided by researchers interested in long-term preservation of their recordings. Perhaps, the most serious disadvantage of analog recordings is that they have to be digitized before any computer-assisted analysis can be performed. Even though, the audio digitization process can offer an extremely high quality of transfer, it is often done poorly, which results in a considerable degradation of the original material.

#### 1.6.2. Digital recorders

Digital recorders are becoming more and more common. They come in a variety of flavors, such as DAT, minidisk, solid state PC card, CD, and hard disk. What distinguishes them from one another are the recording medium and the recording format. For the purposes of acoustic analysis, one should use digital recorders that capture sound in an uncompressed, PCM (pulse code modulation) format.

At the heart of a digital recorder is the Analog-to-Digital converter (ADC). In digital systems, the analog audio signal must first be converted to digital form before it can be further processed. This entails sampling the signal at very short, successive time intervals and converting the value of the sample to a binary number representing the amplitude of the waveform at that instant. The output of the ADC is a series of digital "words", typically at a rate of 44,100 words per second (sample rate). Each word contains a certain number (typically 16) of bits (binary digits). The sample rate and bit depth are the most important factors determining the accuracy of the digital representation of the analog waveform. The most typical digitization settings (so-called "CD quality") of a 44,100 Hz sample rate and a 16-bit word length render a recording that has a broad frequency response (about 21,000 Hz) and an impressive dynamic range of 96 dB. Many modern ADCs are capable of sampling at the rate of 96 kHz and a 24-bit resolution, which produces a very highly accurate signal with negligible quantization noise. After the signal has been quantized the values of each sample have to be stored on the storage the medium that the recording device happens to be using.

### 1.6.3. DAT recorders

DAT recorders use magnetic tape as a storage medium. Typical DAT tapes allow the storage of 120 minutes of uncompressed, high-quality mono recording at 48,000 Hz/16-bit. TASCAM DA-P1 offers sophisticated time coding, rugged construction, and a pair of good pre-amplifiers. It also comes with a built-in limiter and microphone/line inputs. It features the S/PIDF digital I/O interface. The migration of digital audio from DAT to a personal computer hard drive for analysis is a lossless process, provided the correct digital transfer interface is used. The DAT tape is as fragile as a regular analog cassette tape, which, again, raises the issue of long-term preservation. Most digital-born recordings stored on DAT tapes will eventually have to be migrated to either spinning disk or optical storage.

### 1.6.4. Minidisk recorders

Minidisk recorders are quite popular among field workers, particularly due to their small size and ease of use. However, most minidisk recorders do not produce the quality required for detailed acoustic analysis. First of all, most of the portable minidisk recorders lack a quality pre-amplifier, which makes it difficult to interface them with premium-grade microphones. Minidisk field recordings are usually shipped with small, inexpensive lavalier condenser microphones whose frequency response and dynamic range typically rather poor. In addition, there is virtually no control over the incoming signal. The A/D converter is inferior to that found on larger DAT or hard disk recorders. Many portable minidisk recorders lack the S/PDIF I/O interface, which makes it necessary to use a stand-alone minidisk deck for digital audio transfer. Finally, to achieve a small disc diameter, MD uses data reduction based on psychoacoustic principles. Prior to storage, the audio data rate (bit rate) of 1.41 Mbps (uncompressed PCM) is compressed using a perceptual coder (proprietary algorithm developed by SONY) to lower the bit rate to 292 Kbps (1/5 of the original). ATRAC preserves the sample rate (44,100 Hz), but decreases the word length (resolution), which results in a less faithful representation of the analog original and quantization noise, which is then masked by the algorithm, yet, technically it is still there, though we can't hear it. A lower word length, naturally, results in decreased dynamic range, which, of course, the algorithm cannot compensate for. As a result, we have a recording that has the full 20-20000 Hz frequency response, but has a decreased resolution, increased quantization noise, and decreased dynamic range.

### 1.6.5. PC card and CD-R recorders

Marantz has developed a line of professional quality solid state PC card filed recorders. They have impressive features, one of which is the ability to record uncompressed PCM 16-bit audio at the sample rate 44,100 Hz. These recorders are compatible with both compact flash and ATA-size PC cards. The data is stored as a Wav audio file, which can be easily downloaded to a PC with a PC card reader. Unfortunately, the cost of a PC card capable of containing an hour of uncompressed audio is still prohibitively high, and it is uncertain whether the electronic industry will continue to be interested in developing high-capacity, inexpensive memory cards. Recently, Marantz has released a portable CD recorder, PSD 300, which boasts the same professional features as other Marantz portables, but writes data directly to a CD-R or CD-RW. Since this is a new release, it still remains to be seen how the PSD copes with quality issues as well as the demands of fieldwork.

### 1.6.6. Hard disk recorders

Hard disk recording is usually associated with expensive recording studio multi-track recording devices connected to SCSI hard disk farms and racks full of processing hardware. However, recently, small USB devices have appeared on the marked. USB Pre 1.5 is an impressive piece of hardware. It features two independent studio quality microphone pre-amps, phantom power, 24-bit A/D converters, 106 dB dynamic range and a variety of inputs and outputs (tape, line, instrument, microphone, and S/PDIF). The device is powered through a computer USB port and can be used with most modern PC and Macintosh machines, including a variety of laptops. Recordings obtained with USB Pre are superb. Whenever the recording situation allows the use of a laptop, USB Pre (or similar device) should be used to acquire speech signals. At the time of writing this article, USB Pre is one of the best USB digital recording devices, though Digidesign has just released a similar product, the Mbox, which offers similar features to USB Pre, in addition to being fully compatible with Pro Tools 5.2 LE software.

## 2. Processing

### 2.1. Analog-to-Digital conversion and digital data transfer.

In order to perform acoustic analysis on recorded speech data, the audio signal has to be converted into a digital audio file format, such as Wav or Aiff. Analog recordings have to be digitized and digital recordings need to be transferred to a personal computer via a digital audio file transfer interface. This is an important, yet often underestimated, stage in the process of preparing audio data for analysis.

The main goal of A/D conversion (digitization) is to obtain the best possible digital representation of the original analog waveform. Without going into too much technical detail of the digitization process, one should choose a sample rate that will capture a broad range of frequencies and a bit-depth that will allow a wide dynamic range and a negligible amount of quantization noise. These goals can be achieved by means of a premium-quality, stand-alone A/D converter operating at the sample rate of at least 48,000 Hz and a 24-bit resolution. It is absolutely crucial to not to use a PCI multimedia sound card, as they are built from inferior-quality electronic components and, more importantly, allow electrostatic noise and distortion to leak into the captured acoustic signal (Fig 6)

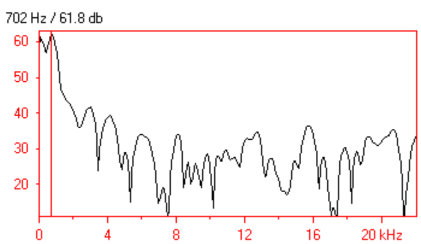


Figure 6 Spectrum of typical electrostatic noise generated by computer circuitry.

The A/D converter, such as Lucid AD 9624, should offer a variety of sample rates, oversampling, high quality anti-aliasing filters, and AES/EBU and S/PDIF digital outputs. Both AES/EBU (Audio Engineering Society/European Broadcasting Union) and S/PDIF (Sony/Philips Digital Interface) are fairly common on high-end digital audio devices. In addition, S/PDIF is used on a variety of consumer-level products, such as CD players, minidisk players, etc. It is also a common interface used on PCI digital I/O cards, which is why it is probably a better choice for most digital audio transfer applications.

| | AES/EBU | S/PDIF (IEC-958 |
|---|---|---|
| Cabling | 110 ohm shielded | TP 75 ohm coaxial or fiber |
| Connector | 3-pin XLR | RCA (or BNC) |
| Signal level | 3..10V | 0.5..1V |
| Modulation | biphase-mark-code | biphase-mark-code |
| Max. Resolution | 24 bits | 24 bits |

Table 2 Comparison of AES/EBU and S/PDIF interfaces

The analog playback device (such as TASCAM 122 mkIII) should be connected to the A/D converter. One should make sure that the output levels on the tape deck match the input levels on the A/D converter. It is recommended to use balanced XLR line level interface (+24 dBu min. gain, +7 dBu max. gain, 65k ohm impedance). If the tape deck does not have this kind of output interface, a signal level transformer (such as Ebtech Line shifter) and a pre-amplifier should be used (Table 3).

| | Clip Level (1% THD) | Sensitivity (typical, for 0 dB FS) | | Impedance (actual) |
|---|---|---|---|---|
| | | min. gain | max. gain | |
| MIC | -12 dBu (195 mV rms) | -10 dBu | –53 dBu | 2k ohm active-balanced |
| LINE | +24 dBu (12.3 V rms) | +24 dBu | +7 dBu | 65k ohm active balanced |
| DI | +9 dBu (2.2 V rms) | +8 dBu | –9 dBu | 10k ohm unbalanced |
| TAPE | +9 dBu (2.2 v rms) | +8 dBu | –9 dBu | 110k ohm unbalanced |

Table 3 Summary of typical signal level types.

The A/D converter needs to be connected to a PCI (though USB and FireWire are becoming common) digital audio I/O card (such as Midiman Delta DiO 2496 via a S/PDIF interface. The digital I/O card should be selected as the recording interface in the audio recording software (such as Sonic Foundry Sound Forge 5.0 on a PC or BIAS Peak VST on a Mac). The digital audio signal should be captured with this software and saved either as Wav (PC) or Aiff (Mac) file at the sample rate and bit depth that the A/D converter was set to. It is also possible to capture digital audio signal directly into acoustic analysis software, such as CSL or Praat, though it is not recommended due to the fact that specialized recording and processing software offers considerable more control over the incoming signal. It should also be mentioned that USB Pre may be used as a high-quality, stand-alone A/D converter. In this case the digital audio signal is transferred to a PC via the USB interface, which eliminates the need to install a separate PCI digital I/O card and makes it possible to capture digital audio on a laptop. In addition, USB Pre has a pair of tape-level inputs, to which a cassette deck can be directly connected (see Table 3).

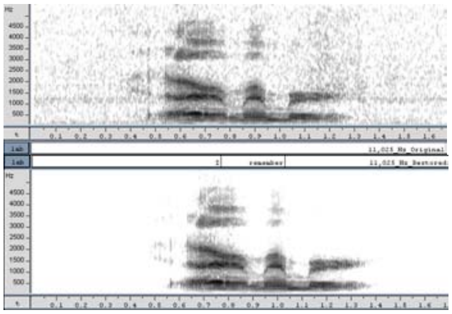
## 3. Analysis

### 3.1. Preparing files for analysis

48,000 Hz/24-bit PCM audio files are not suitable for acoustic analysis, though they should be stored master, preservation copies of the recordings. Such files should be described in simple metadata terms (such as Dublin Core or OLAC) and stored on a dependable optical storage medium, such as CD-R or DVD. In order to prepare files for acoustic analysis, several digital signal processing (DSP) techniques should be used. First, the files have to be downsampled to 11,025 Hz and their resolution should be changed to 16 bits. The downsampling process should always include the use of anti-aliasing filters, and the resolution change should include dithering to minimize the effects of quantization noise. Both Sonic Foundry Sound Forge 5.0 and BIAS Peak VST offer adequate DSP tools to do that. The Nyquist theorem guarantees that the frequency response of a 11,025 Hz is exactly half that value. This is adequate to analyze most speech sounds. If we are dealing with female and child voices, a higher sample rate (e.g., 22,000 Hz) may be used to make sure that the spectrum contains higher frequency transient sounds.

### 3.2. Restoration

Often, linguists find themselves dealing with old, noisy recordings. The digitization process itself does not improve the quality of the recording, nor does it remove any of its imperfections. However, there are a few simple DSP techniques that can be used to clean up and enhance the recording. The example below shows how several DSP techniques have effectively removed some of the unwanted noise from an old DARE recording. The original signal contained low frequency noise whose spectrum overlapped with that of speech, particularly around the f1 area. After the restoration procedures have been applied, the amplitude of noise has been decreased by over 20 dB, which separated it from the speech signal, thus making the file significantly more appropriate for reliable acoustic analysis.



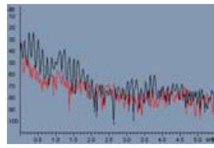


Fig 7a

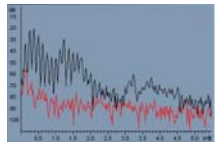


Fig 7b

Figure 7 Spectrograms and power spectra of the phrase “I remember” from a DARE, DM 0735 – S1 tape. Processing used: (1) original digitized at 48/24, (2) converted to 16 bit, (3) downsampled to 11,025 Hz, (4) 2:1 compressed starting at –15 dB, (5) volume adjusted. Figure 7a shows the file before restoration, and Figure 7b after.

### 3.3. LPC in acoustic analysis

Linear Predictive Coding (LPC) is often used by linguists as a formant extraction tool. There are a few important details about LPC that may help avoid common analysis errors. LPC analysis assumes that a signal is the output of a causal linear system. It also assumes that the vocal-tract system is an all-pole filter and that the input to the system is an impulse train. Because of these assumptions, LPC analysis is usually most appropriate for modeling vowels which are periodic and for which the vocal-tract resonator does not usually include zeroes (e.g., in nasalized vowels). The order of an LPC model is the number of poles in the filter. Usually, two poles are included for each formant + 2-4 additional poles to represent the source characteristics. For adult speakers, average formant spacing is in the 1000 Hz range for males and in the 1150 Hz range for females. The LPC order is related to the sample rate of the audio file: 10000 Hz – LPC order = 12-14 (males) and 8-10 (females); 22050 Hz - – LPC order = 24-26 (males) and 22-24 (females)

LPC usually requires a very good speech sample to work with. Many recordings done with omnidirectional microphones contain too little speech detail and too much noise to ascertain reliable LPC readings. Figure 8 shows two spectrograms of the word “Shannon” uttered by the same speaker. One recorded with an omnidirectional microphone, the other with a head-worn cardoid one. The corresponding LPC results show quite different formant frequency and bandwidth values. Clearly, the sample taken with a head-worn microphone contains much more substantially more detail and is more appropriate for analysis. The additional difficulty comes from the fact that the vowel /ae/ in this sample is nasalized due to the following consonant /n/. Only a very high quality recording makes it possible to use LPC in this case. By comparing the values obtained from spectrogram reading to those obtained by LPC, we can safely conclude that the LPC values in Fig 8b are fairly accurate.

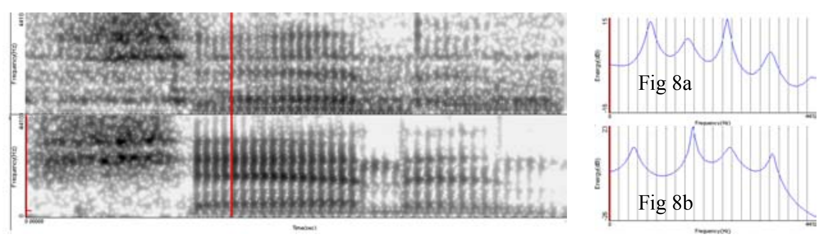



Figure 8 LPC analysis of two speech samples. Figure 8a shows a spectrogram and an LPC graph of the word “Shannon” recorded with a built-in, omnidirectional microphone; the bottom image shows a sample recorded with a directional, head-worn microphone. LPC parameters (frame length = 20ms, filter order = 12, pre-emphasis = 0.9) were kept constant. The following LPC results were obtained:

| | f1 (formant & bandwidth Hz) | f2 (formant & bandwidth Hz) |
|---|---|---|
| built-in microphone | 871; 138 | 1670; 263 |
| head-set microphone | 521; 146 | 1770; 45 |

## 4. Conclusion

Computer-assisted analysis of acoustic speech signals can be dramatically improved if appropriate sound acquisition and processing tools and methodologies are used. It is often true that great results can be obtained with simple means and inexpensive hardware. It is our hope that this article will help researchers achieve that goal.

## References


AKG. http://www.akg-acoustics.com/

A Division Of Sound Enhancements, Inc. http://www.swizzarmy.com/EbtechSite/products.htm

Harris, M.N. and R. Kelly, D.A.McLeod, M.J.Story. 1998. "Effects in High Sample Rate Audio Material"**.** DCS Ltd.Hubber, David M. and Philip Willams. 1998. Modern Microphone Techniques. Vallejo, CA: Mix Books

Johan, Lilkencrants. 1997 Speech Signal Processing. In *The Handbook of Phonetic Sciences*. Eds. Hardcastle William and Laver John. Oxford: Blackwell.Karl, J.H. 1989. *An Introduction to Digital Signal Processing*. Academic Press.

Kay Elemetrics. http://www.kayelemetrics.com

Plichta, Bartek. 2001. Digitizing Speech Recordings for Archival Purposes. Working Paper. Matrix, Michigan State University.

Pohlmann, Ken. 2000. *Principles of Digital Audio*. New York: McGraw-Hill.

Rabiner, Lawrence R., and Ronald W. Schafer. 1978. *Digital Processing of Speech Signals*. Englewood Cliffs: Prentice-Hall.Shure. http://www.shure.com

SONY Corporation. 1993. Minidisk Specifications. Digital Audio Disk Corporation.Sound Devices. http://www.sounddevices.com

Stevens, Kenneth N. 1998. *Acoustic Phonetics*. Cambridge, Mass.: MIT Press. Story, Mike and R. Kelly, D.A.McLeod. 1998. Resolution, Bits, SNR and Linearity. DCS Ltd.Story, Mike. 1997. "A Suggested Explanation for (Some of) the Audible Differences between High Sample Rate and Conventional Sample Rate Audio Material". DCS Ltd.

Titze, Ingo. 1994. *Workshop of Acoustic Voice Analysis*. National Center for Voice and Speech.

Vanderkooy, J., and S.P. Lipshitz. 1984 "Dither in Digital Audio". *JAES*, vol.32, no. 11.Winholotz William S. and Ingo R. Titze. 1997. Conversion of a Head-Mounted Microphone Signal into Calibrated SPL Units. *Journal of Voice*. Vol. 11, No. 4, pp. 417-421.