Audio Processing in Police Investigations

David Luknowsky[1] and Jeff Boyczuk[1]
[1]Audio and Video Analysis Section, Royal Canadian Mounted Police, Ottawa, Ontario

1. INTRODUCTION

The Audio and Video Analysis Section of the Royal Canadian Mounted Police (RCMP) tries to reduce interferences and improve the intelligibility of audio recordings. The primary goal is to be able to clearly understand what was being said. The final result often still contains considerable noise because further noise reduction would also reduce signal information that is important for voice intelligibility. Common sources of the audio recordings are from hidden devices, interview rooms, hand held recorders, wire taps and 911 calls. Many different types of interference are encountered and each case is unique.

Some of the more common types of interference include tones, hum, motor noises, and music. These interferences may be reduced with notches which may be set manually or automatically, comb filters, spectral inverse filters, and adaptive filters. Common problems with voice which have to be corrected are low volume of one or more voices, reverberation, and clipping.

2. METHOD

One of the most common types of interference, and one of the simplest to reduce, is tones and hum. Hum typically comes from power lines, fluorescent lights or other electrical sources. It is seen in the spectrum as spikes at 60 Hz and harmonics thereof (Figure 1).

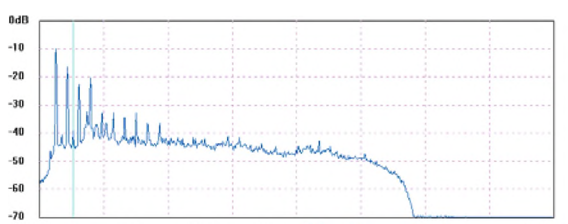

Figure 1. Average voice spectrum with strong power line hum.

60 Hz hum can easily be removed with a comb filter which applies notches every 60 Hz up to a maximum amplitude (Figure 2). The amplitude of the harmonics often falls below the average signal beyond 360 Hz so the maximum of the comb filter may not need to go much beyond this.

A series of individual tones at various, unequally spaced frequencies or tones whose frequency varies over time are also common tonal interference

Figure 2. Spectrum from figure 1 with hum reduced using a comb filter.

problems. For this type of interference, we would probably use an adaptive filter. There are two algorithms which are likely to be used for cases like this: 1) an algorithm which analyzes the amplitude over time of all frequencies in small, discrete bins and identifies frequencies whose amplitude is not changing over time, within certain defined parameters. 2) an adaptive predictive de-convolution filter which de-correlates the input signal and removes long term correlated components [1]. Essentially the filter delays the input signal, attempts to predict the next part of the signal using a least-mean-square estimator, and then subtracts the predicted signal thereby removing predictable components such as hum which remain consistent over time. Adaptive filters such as these can also be useful for reducing other types of predictable interferences such as motor noise from engines, fans, or air conditioners, and with the appropriate parameters can be set to reduce music since it is generally much more predictable than voice.

If tones are strong yet stationary and unevenly spaced (so that a comb filter would not work), a spectral inverse filter may be used. The average spectrum is analyzed and divided up into nearly a thousand equally spaced frequency bands. Amplitude adjustment can then be applied to each of these bands to produce a flat or shaped average spectrum as if one were automatically setting a very high resolution graphic equalizer. While a spectral inverse filter can be used to reduce tonal noises, it is more likely to be used to shape the average spectrum to increase the clarity of the voices. This is especially useful for analogue recordings with poor high frequency response such as those made at a very low record speed. Such recordings generally have too much bass and sound muffled. A spectral inverse filter with shaping will provide significant amplitude boost to higher voice frequencies and then shape the spectrum to a typical averaged voice spectrum[2]. While this may boost high frequency noise, it can also greatly reduce the muffled sound and increase the crispness, clarity, and intelligibility of the recorded voices.

In addition to problems with interferences, there are often problems with the level of one or more voices. On recordings of interviews, it is common to have the interviewee's voice at a very low level even when the investigator's voice is loud and clear. Similar problems are seen with body-worn recorders where the voice of the person wearing the recorder is clear but the other voices are at a very low level. Sometimes playing the audio through an automatic gain control, which attempts to bring all sounds to about the same record level, is sufficient. When the level adjustment needs to be dramatic or sudden, a digital audio editor is used. This allows very accurate selection of the section that needs to have its level changed and very quick and precise changes in level.

Audio levels which are too high and which result in distortion or clipping are also seen. In some cases where the clipping is not too bad, the waveform can be reconstructed to eliminate the clipping at that point. For example, Figure 3 shows a stereo audio waveform with clipping on the top and just below shows the same waveform reconstructed to remove the clipping effect by interpolating the waveform based on the surrounding data points.

The quality of the recordings we deal with is often quite poor. Sometimes the quality of the media itself is marginal. Before processing, we want to get the best possible signal from the media. The first step in processing forensic audio is to make sure we are dealing with the original recording. We do not want to work on a degraded copy. It is then necessary to ensure the quality of the media. For tape, this may mean cleaning the tape and/or placing it in a new cassette if there is damage. Good playback equipment with clean heads that can be azimuth adjusted for the best signal on playback is important for tape recordings. Such steps to get the best available signal from the media can sometimes make a big difference in the end product when dealing with forensic audio.

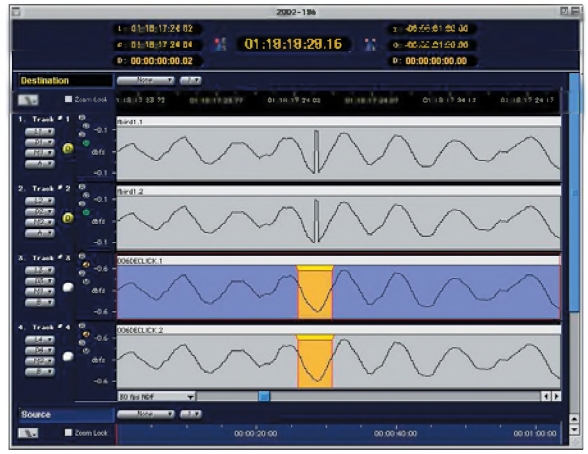

Figure 3. Audio waveform with clipping (top) and corrected (bottom). Total length of audio shown is 9.6 ms.

3. DISCUSSION

The exact combination of filters and filter parameters which provides the best intelligibility may be different for each recording. The analyst is always tasked with determining which processing to use in order to achieve the highest level of intelligibility. This is not the same as producing a nice sounding or noise free recording. Higher frequencies which often make voice recordings sound noisy need to be included (and sometimes even amplified) because voice information crucial to intelligibility would otherwise be lost. Being able to understand what was said is always the most important thing.

REFERENCES

[1] Paul, J.E., "Adaptive Digital Techniques For Audio Noise Cancellation," *IEEE Signal Processing Magazine,* Vol. 1, No. 4, pp. 2-7, 1979.

[2] J. L. Flanagan, Speech Analysis Synthesis and Perception, New York: Springer-Verlag, 1972, p. 163.