Chapter 7

Active Audio Recording

As described in the previous chapter, several components have been added to E-Chalk's audio system because, in contrast to the scenario assumed for the older versions of World Wide Radio, one cannot just assume that the audio signal fed into the encoding computer is of broadcast quality. This chapter analyses some of the problems causing the quality distortion and presents techniques to improve the situation. These techniques were integrated into a wizard application that is used before the first lecture and several SOPA nodes that run during the recording. The entire system was integrated into E-Chalk under the name *Active Audio Recording*, which is also described in [Friedland et al., 2004b, Friedland et al., 2005c].

7.1 Audio Recording in Classrooms

Looking at the different usage scenarios of E-Chalk (see Chapter 3) and those of similar lecture-recording systems, several practical problems that deteriorate audio quality can be observed. The two key factors that most often degrade the audio quality are usage errors of the sound equipment, including a wrong assessment of the expected quality with given equipment and distortion sources that are typical for the situation in a classroom or a lecture hall.

7.1.1 Usability Problems

According to my experiences as well as user feedback, the usability problems that directly concern audio quality can be classified into three categories: Wrong assessment of the expected results using certain equipment and workload, usage errors that concern configuration and setup, and issues during the lecture.

Both teachers and students are accustomed to high fidelity, almost noise-free, broadcast-quality voice recordings that they are able to receive every day through radio and television or are able to buy on commercial compact discs. These recordings are mostly produced in a studio with very costly equipment and (several) sound engineers present surveilling every recorded sample both by listening to it and observing different measurement instruments. Just plugging a microphone into a notebook computer's sound input jack does not deliver the same results.

Of course, the experienced audio quality of a lecture recording also depends on the speakers and the equipment at the receiving end. In fact, most sound cards focus on sound playback but not on sound recording. Gaming and music replay are their most important applications. Many sound cards cannot generate studio-quality audio recordings. On-board sound cards, especially those in laptop computers, have often very restricted recording capabilities. Even when a decent sound card is installed, noise may be introduced to the sound equipment by hard disk motors and fans because of the very compact construction [Mack, 2002].

Many people also have problems setting up the computer for sound recording. Especially if it is an unknown notebook computer and the setup is made adhoc, e. g., directly before a presentation. The improper handling of the operating system's mixer often causes malfunctions. It differs from sound card to sound card and from operating system to operating system and usually has many knobs and sliders with potentially wrong settings. Selecting the right port and adjusting mixer settings can take even experienced users minutes. Often the microphone is put into the wrong input jack – an error that is noticed mostly only after the presentation or lecture. Even if the right input jack is chosen, the input level is very often not adjusted well because many people just do not how to adjust the input level correctly. The results are either overflows or the recording is not set to maximum gain.

During a lecture, the instructor's attention is entirely focused on the presentation, and technical problems can easily be overlooked. Often, lecturers just forget to switch the microphone on. In many lectures, weak batteries in the microphone caused a drastic reduction of the signal-to-noise ratio.

However, as described in Chapter 3 and discussed in Section 6.1.3, E-Chalk's philosophy requires that using the system must ideally not cause additional overhead. The lecturer has to do his teaching job, and controlling a battery lamp during the whole lecture or setting up the mixer to the right level is overhead. So the problems resulting in not doing these above-mentioned things right or even not doing them at all cannot be considered the lecturer's fault.

7.1.2 Distortion Sources

Besides the above-mentioned usability issues, there are also many direct audio distortion sources that can affect the audio quality of a lecture recording. This paragraph only concentrates on those that in my experience have the greatest impact on recording quality. For a more detailed discussion of these problems refer for example to [Dickreiter, 1997a, Katz, 2002].

In contrast to a recording studio, a classroom or a lecture hall is filled with multiple sources of noise: Students are murmuring, cellular phones ring, doors slam, etc. Several independent studies provide measurement methods and quantitative data, see for example [American National Standards Institute, 2002], [Hodgson et al., 1999], or [Bradley et al., 1999]. In lecture halls there may also be reverberation that depends on the geometry of the room and on the amount of people in the seats [Knecht et al., 2002]. As a consequence, the speaker's voice does not always have the same loudness as it adapts to the noise level of the audience. The loudness and the volume of the recording depends on the distance between microphone and the speaker's head, which is usually changing all the time. Coughs and sneezes or even movements of the lecturer result in irritating noise. Feedback loops can also be a problem if the voice is amplified for the audience. Long audio wires can cause electromagnetic interference that results in noise or humming.

7.1.3 Other Issues

Some issues are also directly caused by low-quality sound cards. Some analog-to-digital converters introduce a fairly high DC-offset. A DC component is unhearable during replay. However, when a lecture is appended to a lecture recorded with a sound card that has different offset or Exymen is used to edit the sound a changing DC-offset causes a "click" sound at the merging point.

Yet another problem is the incorrect timing of soundcards. E-Chalk's board events are timed using the clock of the computer, while the audio system relies on getting a certain number of samples per second. However, comparing the computer clock with the number of sound samples received sometimes results in noticeable differences. Using a small measurement application distributed to several E-Chalk users, I found that these timing differences differ from sound card to sound card[1]. Usually, better sound cards have a more accurate timing, but timing errors of up to 0.1 % appear in many sound cards. In a 90-minute lecture an error of this magnitude results in a desynchronization between board and audio track of about five seconds at the end of the lecture.

7.1.4 Ideal Audio-Recording Conditions

Even though many audio distortion sources exist in a classroom or lecture hall, this does not mean that broadcast-quality recording is impossible. Many commercial music productions for example have been produced at live concerts of rock bands. Usually, the noise level at these events exceeds the noise in classrooms or lecture halls by several orders of magnitude. However, the equipment and personnel required for such a recording also exceeds any effort expendable for regular lecture recording. It is impossible to describe a generic solution for all kinds of rooms, speakers, and situations in one section. Nevertheless, looking at what would be ideal helps to identify the problems of the status quo.

In contrast to a studio recording, the voice of the instructor is also important for an audience in the room. Either the instructor speaks aloud or additional speakers are used to amplify his or her voice. In the latter case, care must be taken that the speakers for the audience are both loud enough and cause no feedback with the recording microphone. Reverberation effects of the room should be avoided using proper absorption material, one of many typical methods is to pad the back of the audience seats.

Since noise is inevidentably present in lecture halls or seminar rooms, a directed microphone should be used for recording. This also eliminates other influences of the room acoustics, like reverberation. Directed microphones, however, are usually very sensitive against direct contacts or movements of the instructor which result in scratch sounds. The microphone should have some kind of pop-screen to avoid popping effects during the recording of explosive consonants. The distance to the microphone should be adjusted according to the specification of the microphone (usually about 30 to 60 cm from the mouth of

[1]Thanks to Stefan Flor and Thomas Klein of the Physics Department of University of Rostock for their initial reporting of this problem.

the speaker). As the volume of the speaker's voice changes dramatically with the distance, wireless microphone headsets are a perfect technical choice. If the distance between sound source and microphone is too small, the proximity effect boosts the lower frequencies and makes the voice sound unnatural. However, some lecturers feel constrained by having to wear a headset. The alternative is a wireless lavalier microphone. However, the distance has to be properly adjusted and scratch sounds due to movements or direct contact – for example by clothes that brush against it – must be avoided. The signal level can drop or raise dramatically if the speaker turns away or closer to the mic.

Before the recording, a sound check should be performed. The recording should then be monitored by a technician to ensure the equipment works as expected. The gain must be controlled continuously because the volume of the instructor can change rapidly. If the gain is too low, the signal-to-noise ratio becomes a problem; if gain is too high, the signal is clipped. Both problems result in distortions of the audio signal especially when lossy encoding techniques are used because they assume a perfect input signal. Furthermore, noise raises the entropy of the signal, decreasing the compression ratio. Cable length should be minimized and only shielded cables should be used. Although this seems to go without saying, we have experienced many problems with university lecture halls using long cables that introduced humming. Equalizers may be used to balance out certain frequencies and to tune the signal for a more pleasant listening experience.

The factor that should be optimized for a lecture recording is speech intelligibility. The two factors that mainly determine speech intelligibility are the upper-bound frequency and the signal-to-noise ratio [Dickreiter, 1997a, Allen, 1994]. The upper-bound frequency is cut off mainly by the sampling rate of the sound card. Given state-of-the-art codecs, however, the distortion introduced by noise and the upper bound frequency is virtually imperceptible, especially for speech (see for example [Hansen, 2002]). The most important factor is the sound equipment and from that, mainly the microphone and the sound card [Mack, 2002].

7.2 Improving Audio-Recording Quality

Given the problems discussed in the previous section together with the fact that in most cases no further personnel is available for consultation during or even before lecture recording, the question arrises how E-Chalk could be able to assist a lecturer in improving the audio quality. Providing state-of-the-art codecs for audio compression is important; however, if the signal is distorted even before the compression, satisfying results will not be achieved anyways. Therefore, improving audio recording for lectures means first and foremost improving the quality of the raw signal before it is processed by the codec. Ideally, it should be possible to produce satisfactory audio quality with standard hardware and without a technician necessarily present. In order to do this, the software must be able to help the lecturer in assessing the expected results, monitor the signal during the recording, and provide basic tools that automatically reduce the influence of distortion sources. Yet, sound recording is a profession and a research area of its own and, of course, the work presented in this chapter does not aim to replace them. However, the work in this chapter indicates that there are possibilities to assist an audio un-savvy user in producing higher-quality speech recordings.

None of the state-of-the-art Internet broadcasting systems such as Windows Media Encoder, RealProducer, or QuickTime (see Chapter 2) provide automatic monitoring or signal enhancing mechanisms. Like the old World Wide Radio system, they assume the typical streaming usage scenario where a high-quality audio signal is fed in by a radio broadcasting station with audio technicians being present. As already discussed in Section 2.3 as well as explained in [Mack, 2002], these software systems integrate into a recording work flow that requires both adequate equipment as well as trained personnel. Microsoft's *Real-Time Communications API* at least provides an audio-tuning wizard which provides a manual input-device selection and a dialog that helps a user to adjust the microphone distance. Most video-conferencing tools, such as Polycom ViaVideo (see Chapter 2), do at least have basic filters for echo canceling or feedback suppression. Octiv Inc. applies real-time filters to increase speech intelligibility in telephone calls and conferences. They provide hardware to be plugged into the telephone line. Cellular telephones also apply filters and speech enhancement algorithms but these rely on working with special audio equipment and knowing the properties of the underlying devices. Among other products, Octiv also sells a product called *Volume Pro* which acts like a kind of enhanced audio compressor [38].

In academic research, many projects seek to solve the *Cocktail Party Problem* [Haykin, 2003]. Most approaches try to solve the problem with blind source separation using extra hardware, such as multiple microphones. Itoh and Mizushima [Itoh and Mizushima, 1997] published an algorithm that identifies speech and non-speech parts of the signal before it uses noise reduction to eliminate the non-speech parts. The approach is meant to be implemented on a DSP and although aimed at hearing aids it could also be useful for sound recording in lecture rooms.

In practice, most recording and sound-editing tasks are still solved manually. Just as in many image manipulation programs, the tools provided by generic audio-editing applications are very powerful, yet they require a user who knows what he or she wants to do. Even consumer-level remastering software such as Steinberg "WaveLab" [51] usually require the user to visit an introductory seminar to be able to handle the software properly – not to mention more complex remastering software, such as "Magix Samplitude" [30] or "Logic Pro" [7]. Automation is still at the beginning and only available for special purposes. For example, Steinberg also provides software like "My MP3 Pro" [51] that facilitates the creation of MPEG-encoded files from a given audio source.

PEAQ [Thiede et al., 2000] is a sound-quality-assessment algorithm based on psychoacoustic models. It aims to emulate a sound quality test using audio experts at a degree of quality where distortions are not easily noticeably anymore. The algorithm is *intrusive*, i. e., it requires a reference signal. Additionally, the quality measurement is computationally expensive, so an online measurement is impossible at the moment[2]. However, the algorithm is actually a combination of several methods from which some can be used to assess audio quality during and before lecture recording, as described below.

[2]I thank Opticom GmbH for providing me a with a test version of their PEAQ and PQSM implementation distributed under the product name Opera.

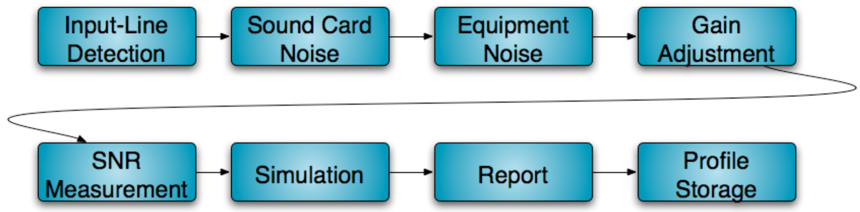

Figure 7.1: A conceptual overview of the steps of the audio diagnose wizard that is to be run before the first lecture recording.

Of course, not all of the audio-distortion problems discussed in Section 7.1 can be solved by software. E-Chalk's active audio recording component focuses on the special case of lecture recording and mainly concentrates on the automation of equipment configuration and sound hardware monitoring. The system assists in the assessment of the sound equipment and provides several basic methods for the oppression of audio distortions. The system relies on the lecturer using some kind of directional microphone, so that the influence of room geometry and of cocktail-party noise is already eliminated. A lecture-recording system has the advantage that information about speaker and equipment are already accessible before the recording. The approach is therefore divided into two parts:

1. An expert system analyzes sound card, equipment, and the speaker's voice and keeps this information for recording. It assists in assessing the quality of the audio equipment and makes the user aware of its influence on the recording.

2. During recording, a hardware monitor and some basic filters use the information collected by the expert system to improve the quality of the incoming audio signal.

7.3 Before the First Lecture

Before the first lecture in a new room or with new equipment is recorded, the lecturer creates a so-called audio profile. It represents a fingerprint of the interplay of sound card, equipment, and speaker. The profile is recorded using a GUI wizard that guides through several steps, see Figure 7.1. This setup takes about three minutes and has to be done once per speaker and sound equipment. Each speaker uses his or her audio profile for all subsequent recordings as long as the recording equipment remains unchanged. The GUI wizard does not only record the audio profile, it also simulates several tasks that a recording technician would do before a lecture recording. The program identifies and configures the sound hardware setup (once it has been installed into the operating system) and takes sample measurements of the sound card and the sound equipment. It then simulates E-Chalk Audio's processing chain, allowing a user to listen to the recording exactly as it will be broadcast. A final report gives a hint how the sound equipment compares to an ideal one based on the measurements. The results are saved in the audio profile to be be used by E-Chalk during recording. Because users asked for the diagnostic features of the wizard independently of E-Chalk, the wizard also exists in a diagnostic-only version that runs as a stand-alone application and does not create an audio profile. In the following paragraphs, the diagnostic steps of the expert system underlying the wizard are presented. Further technical details on the underlying methods are presented in Appendix E.

7.3.1 Detection of Sound Equipment

The first step of the setup consists of detecting the audio equipment. The user is asked to disconnect every input device from the sound card but the one he or she wants to record with and turn this on. Using the operating system's mixer API the sound card's input ports are scanned to find out the recording devices plugged in. This is done by shortly reading from each port with its gain at a maximum, while all other input lines are muted. The line with the maximum noise level is assumed to be the input source. For the result to be certain, the maximum must differ to other noise levels by at least 3 dB, otherwise the user is required to select the line manually. With a single source plugged in, this occurs only with digital input lines because they produce no background noise. At this stage several hardware errors can also be detected, for example if noise is constantly at zero decibel there is a short circuit. After the sound card and input line has been chosen, E-Chalk's audio system takes full control over the sound card mixer, both during the rest of the steps of the wizard and during recording. There is no need for the user to deal with the operating system's mixer.

7.3.2 Recording of Floor Noise

In theory, silence should not contain any noise or the noise level should at least be below the hearing threshold. The second step in the wizard is therefore the recording of the sound card's floor noise. The user is asked to unplug all input devices from the sound card[3]. The mixer control raises input gain to maximum and a few seconds of "silence" are recorded. The signal is analyzed to detect possible hardware problems or handling errors. For example, if the floor noise is too high, sound maybe introduced by electromagnetic interference. Harddisks or fans are typical sources of such problems. Overflows may be caused by short circuits at the sound card ports and wires. These critical noise levels result in descriptive warnings.

After recording sound card noise level, the user is asked to re-plug and switch on the sound equipment. Then the equipment floor noise is measured. The equipment floor noise is influenced by several factors and consists maybe of noise or humming introduced by the equipment or direct environment. Before analysis, the user is asked to play back the recording in order to verify that no accidental sounds have been recorded. Again, warnings will be shown if the floor noise level is too high. Furthermore, comparing this signal to the previous recording enables to detect handling errors. If the equipment noise level is lower than the sound card noise level, the measurement was performed wrongly. If

[3]On notebook computers this is not always possible because built-in microphones cannot always be switched off. The wizard then skips the recording of sound card background noise.

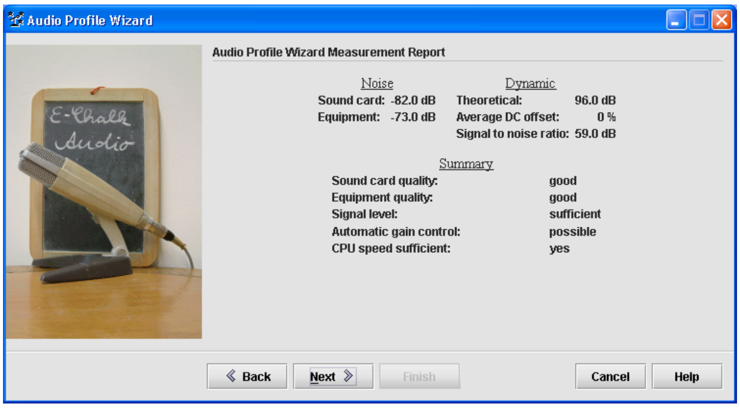

Figure 7.2: A report gives a rough estimation of the quality of the equipment.

both are the same, the equipment is likely to be switched off or connected to a wrong port. If overflows occur during the recording of floor noise there are very probably hardware misfunctions. A constant maximum level indicates a possible short-circuit. Both floor noise recordings are stored for later use.

7.3.3 Dynamic-Range Adjustment

In order to equalize the voice level, adapt to loudness variations, and to level out distortions easily detectable by signal peaks, Active Recording includes an automatic gain control (see Section 7.4). Since the dynamic range adjustable by the sound card mixer is mostly only a sub-intervall of the overall dynamic range of the equipment connected to the sound card, the equipment must be adjusted so that the automatic gain control is able to control the signal level optimally. For a perfect gain control, the recorded voice varies only inside the dynamic range of the mixer. The dynamic range adjustment is the next step of the GUI wizard. The user is asked to record a phrase containing many explosives. In English, repeating the word “coffee pot” seems to provide good results. Explosives form the peaks in voice recordings and are therefore able to provide the upper bound of the dynamic range. During the test recording, the sound card mixer's input gain at the current port is adjusted to control the gain of the signal. During recording, the average signal level should be maximized but overflows must be avoided. If too many overflows are detected or if the average signal gain is too low, the user is informed about possible improvements.

7.3.4 Measuring Signal-to-Noise Ratio

As discussed in the previous sections, noise is one of the predominant factors in speech intelligibility and also an important quality indicator for sound equipment. For an accurate measurement, a frequency generator has to be plugged into a sound card port and both the harmonic distortion as well as the noise added by the sound card have to be determined by measuring the distance between the input and the output at different frequencies and waveforms. Of course, this requires equipment and more importantly an educated technician.

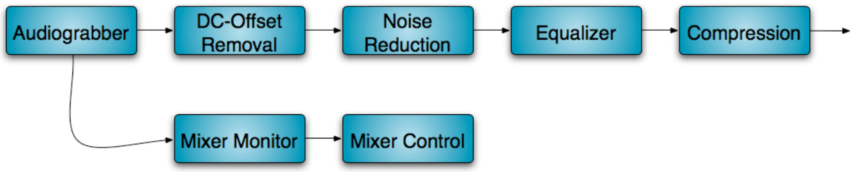

Figure 7.3: E-Chalk Audio's processing chain during lecture recording. The encoded signal is then transmitted or stored to a file.

Therefore E-Chalk Audio relies on an approximation that is measured more easily. In the next step of the wizard, the user is asked to record a predefined sentence. The gain is measured, without counting speech pauses between words.

The signal-to-noise ratio is then estimated by comparing the A-weighted gain [DIN EN, 2003] against the A-weighted floor noise. Although only an estimation, this is a very practical method that has often been used – for example to measure the signal-to-noise ratio of tape recorders [Dickreiter, 1997b].

7.3.5 Fine-Tuning and Simulation

When the tests and adjustments have been finished, the user is asked to record a typical start of a lecture. This final recording serves as the basis for a simulation and allows for some fine-tuning. The recording is filtered (as described in Section 7.4), compressed, and uncompressed again using E-Chalk's default codec. The user is able listen to his or her voice as it will sound after having been transmitted through the Internet. This step does not only provide a demonstration of the expected results, it is also useful for debugging certain problems before the actual recording of the lecture. If necessary, an equalizer (according to [ISO, 1997]) allows experienced users to further fine-tune the frequency spectrum of the recording. The time for filtering and compressing is measured. If this process takes the same time or even longer than the length of the recording, audio packets will be lost during real recording because of a slow computer.

7.3.6 Summary and Report

At the end of the simulation process a report is displayed, as shown in Figure 7.2. The report summarizes the most important measurements and grades sound card, equipment, and signal quality into the categories "excellent", "good", "sufficient", "scant", and "inapplicable". The sound card and the equipment are graded using the background noise and the estimated signal-to-noise ratio calculated from the recordings. A sixth grade, "improperly measured", is given for contradictory results, for example when equipment noise level is lower than sound card noise level alone. Of course, this is only a very rough grading. Sound quality is determined by many more factors, for example, frequency response, harmonic distortion, or phase distortion. Other factors can be ignored, for example crosstalk does not play a role since E-Chalk Audio only records mono sound. Many problems of the past are believed to be good enough in modern sound cards, such as the dynamic range (which is actually determined by the

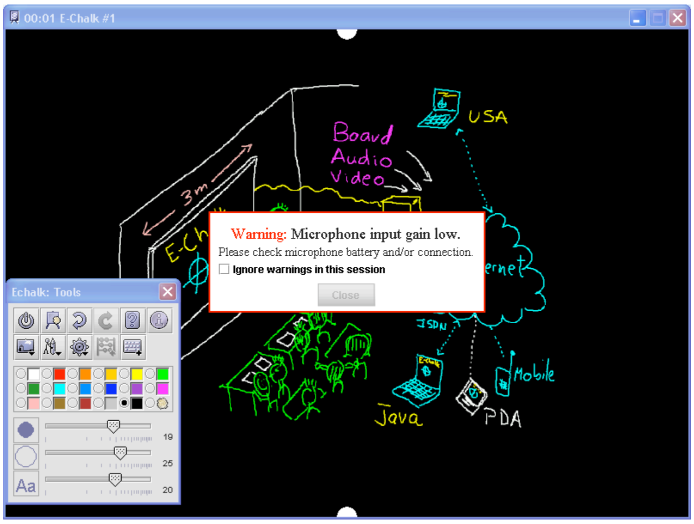

Figure 7.4: The microphone's floor noise level has sunk – batteries have to be changed. The warning dialog appears directly in front of the chalkboard and disappears without any need of interaction when the problem has been fixed.

bit depth) and jitter. The advantage of using this simple heuristics is that the measurement of silence noise can be easily performed by users without prior knowledge and still assists in identifying quality bottlenecks. Further testing would require the user to work with loop-back cables, frequency generators, and/or measurement instruments. In order to create a grading scale, practical experience reports were collected from the Internet and consumer computer magazines, for example from [56], and floor noise level seems to be an important indicator for the overall quality of sound cards. Further details on the grading heuristics can be found in Appendix E.

In the last step of the wizard the user is asked to enter an identification string for the audio profile. Among other information, the created profile contains all mixer settings, the equalizer settings, the recordings, and the sound card's name. The identifaction string then appears in E-Chalk's Startup Wizard and enables the selection of a certain profile for the Active Recording chain.

7.4 During Lecture Recording

This section describes the steps that are performed during the lecture that are also illustrated in Figure 7.3.

7.4.1 Mixer Monitor

During the lecture, the system relies on the profile of the equipment. If changes are detected, for example a different sound card, the system complains at startup. The mixer settings saved in the profile are used to initialize the sound card mixer. The mixer monitor complains if it detects a change in the hardware configuration such as using a different input jack. It supervises the input gain

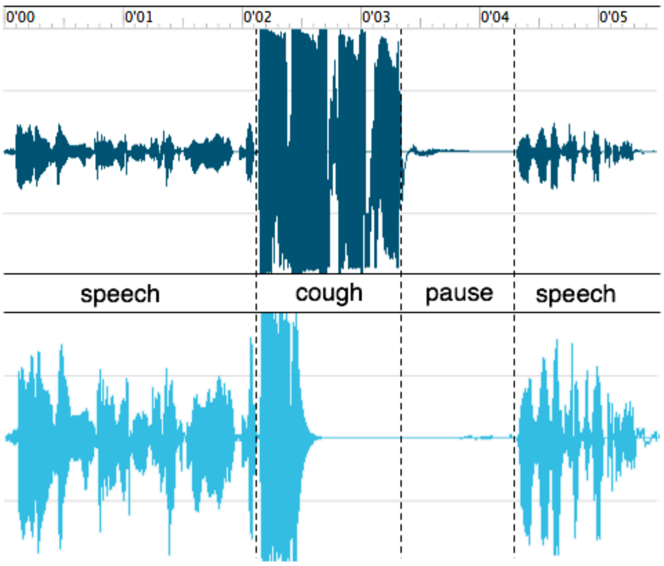

Figure 7.5: Without (above) and with (below) mixer control: The speech signal is expanded and the cough is leveled out.

in combination with the mixer control. A warning is displayed if too many overflows occur or if the gain is too low, for example, when microphone batteries are running out of power. The warning disappears when the problem has been solved or if the lecturer decides to ignore the problem for the rest of the session. Figure 7.4 shows a warning dialog presented during the recording.

7.4.2 Mixer Control

The mixer control levels out the input gain using the sound card's mixer. The analog preamplifiers of the mixer channels thus work like analog expander-compressor-limiter components used in recording studios. This makes it possible to level out voice intensity variations. Coughs and sneezes, for example, are leveled out (compare Figure 7.5) as well as gain variation caused by microphone distance changes (compare Figure 7.6). The success of this method depends on the quality of the sound card's analog mixer channels. Sound cards with high-quality analog front panels, however, are becoming cheaper and are getting more popular. Automatic gain control reduces the risk of having feedback loops. Whenever a feedback loop starts to grow, the gain is lowered. As in analog compressors used in recording studios, the signal-to-noise ratio is lowered. For this reason noise filters, as described in the next paragraph, are required. In order to be able to react quickly on gain changes, Steinberg's "ASIO" low-latency sound-recording interface is used if the sound card provides it.

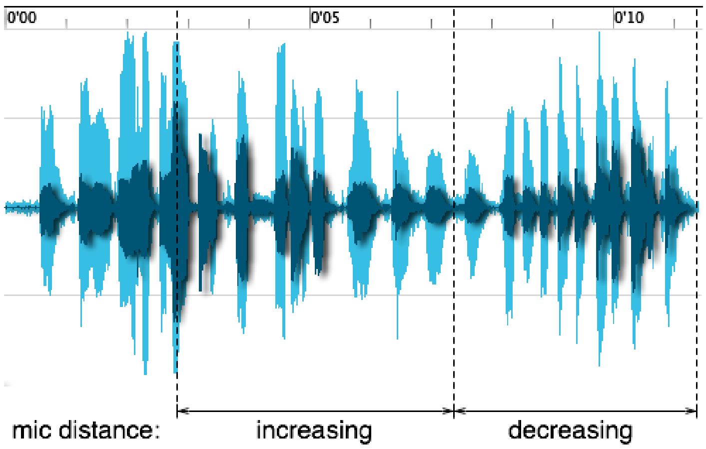

Figure 7.6: Another example of the mixer control in action. When the lecturer turns away from the microphone, the audio gain goes down and with the instructor's mouth approaching the microphone the gain raises again (darker signal). Using the mixer control, the overall gain is higher and the microphone distance differences are leveled out more effectively (lighter signal).

7.4.3 Filtering

According to the measurements the audiowizard creates a filterchain using a SOPA graph description (see Section 4.7.2) that is used during the recording to eliminate common problems. Of course, the chain can also edited manually. Mostly, however, the following chain is used. First, the signal's DC offset is removed. Then, the sound card's background noise level is used as threshold for a noise gate and the equipment noise as a noise fingerprint. This is an important step since the mixer control tends to raise the silence noise level. The fingerprint's phase is aligned with the recorded signal and subtracted in frequency space. This removes any humming caused by electrical interference. Because the frequency and shape of the humming might change during a lecture, multiple noise fingerprints can be specified. A typical situation that changes humming is when the light is turned on or off. The best match is spectrally subtracted as described in [Boll, 1979]. See Figure 7.7 for an example. It is not always possible to pre-record humming, but if so this method is superior to using electrical filters. Electrical filters have to be fine-tuned for a specific frequency range and often remove more than wanted. Any equalizer settings specified by the user are applied before the normalized signal is processed by the codec.

7.4.4 Final Processing

After the recording has finished the recorded samples are counted and compared to the time stamps of the board event file (if present). The mismatch due to any timing difference between sound card and real-time clock of the computer is calculated and logged. During later replay, the timing of the audio playback can then be adapted. The Java audio replay client, for example, adjusts the

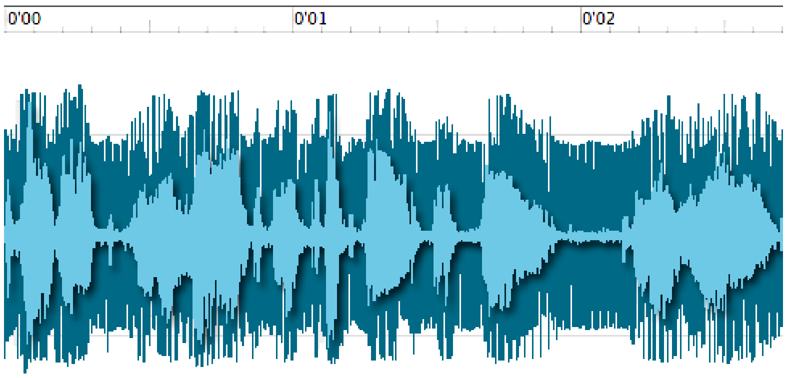

Figure 7.7: Three seconds of a speech signal with a 100 Hz-sine-like humming before (dark) and after spectral subtraction (light). Humming is a frequent audio distortion in lecture halls.

MASI timing information according to a parameter defined in the invoking web page.

7.5 Practical Experiences

As explained above, Active Recording tries to eliminate common recording distortions using well-known filtering methods. The filters themselves have already established as standards and thus do not need to be evaluated. The performance of the entire system is difficult to evaluate empirically because if E-Chalk users are aware of audio recording problems and decent audio equipment is used, the filters are actually not needed. All in all, filtering results in a more efficient compression. Because noise and clipping is reduced, entropy also scales down and the codec is able to achieve better results. By compressing several experimental recordings before and after the filter chain using the default WWR3 format (see Appendix D), we have measured that the bandwidth reduction due to Active Recording is about 10 %.

The system has been integrated into the E-Chalk system since 2004. Both instructors and technicians that were regularly using E-Chalk considered a setup time of several minutes complicated at first. The opinion, however, usually changed when the recording was saved because a small usage error had been prevented, for example picking up the wrong microphone. Common recording distortions were eliminated and the listeners of the courses reported a more pleasent audio experience. The assessment of the sound card and equipment is meant to be rather strict and people often complained about their sound cards to be assessed "too badly". Having to go through a wizard before the first recording raises the awareness for potential recording issues. Using download logs and a registration procedure in the installer, we estimate about 800 E-Chalk installations at the time of writing this text. Before the introduction of the Active Recording system, we regularly had to answer support emails concerning the audio quality. With the system being part of E-Chalk, not a single question was emailed to us concerning audio-recording quality issues.

7.6 Limits of the Approach

Of course, the software system does not replace a generic recording studio, nor does it make audio technicians jobless. In the special case of on-the-fly lecture recording, however, it raises the awareness for audio recording issues and filters out several standard distortions. Sound quality enhancement provided by software is only one element of the signal chain. A heavily distorted audio signal can only seldom be restored by software. More problems can be solved if the sound equipment of the entire signal chain is known in advance. E-Chalk's Active Audio Recording works mainly by reacting to the results of a steady comparison between the sound intensity of the incoming signal with the prior recorded fingerprint of the equipment and the speaker. In order to get a better approximation to what an audio technician would do when monitoring the incoming audio signal to prevent quality deviations or malfunctions, it is necessary to interpret the sound information at a higher level than by basic operations on a set of samples in amplitude and frequency space. Speech recognition methods, for example, could provide the necessary operations to come closer to how a human being listens to the incoming signal.

7.7 Conclusion

None of the software presented in Chapter 2 takes into account any classroom-specific recording problems because they simply use commercial Internet audio-broadcasting systems. This does not yield satisfying results. Because audio encoding strategies have become better and transmission bandwidth has increased during the last years, quality bottlenecks are now primarily caused by human factors. However, freeing users from performing technical setups by automation, as recently observable in digital photography, is still a challenge for audio recording. The Active Audio Recording component of E-Chalk is a first small step towards making the creation of high-quality recordings easily possible for the layperson.