Towards Automatized Studioless Audio Recording: A Smart Lecture Recorder

Gerald Friedland, Kristian Jantz, Lars Knipping
Institut für Informatik
Freie Universität Berlin
[fland|jantz|knipping]@inf.fu-berlin.de

September 2004

Abstract

Webcasting and recording of university lectures has become common practice. While much effort has been put into the development and improvement of formats and codecs, few computer scientist have studied how to improve the quality of the signal before it is digitized. A Lecture hall or a seminar room is not a professional recording studio. Good quality recordings require full-time technicians to setup and monitor the signals. Although often advertised, most current systems cannot yield professional quality recordings just by plugging a microphone into a sound card and starting the lecture. This paper describes a lecture broadcasting system that eases studioless voice recording by automatizing several tasks usually handled by professional audio technicians. The software described here measures the quality of the sound hardware used, monitors possible hardware malfunctions, prevents common user mistakes, and provides gain control and filter mechanisms.

Contents

List of Figures

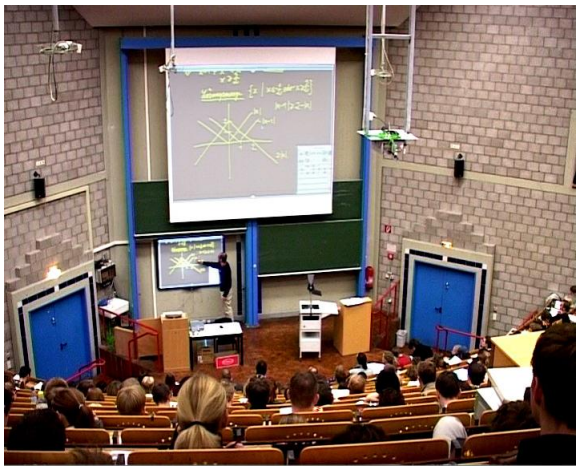

Figure 1: A lecturer using E-Chalk in a large lecture hall.

1 Introduction

The work presented in this paper has emerged from experiences developing a lecture recording system, called E-Chalk [6, 14, 13], and its accompanying user evaluation [7]. The system produces Web based learning modules as a by-product of regular classroom teaching, see Figure 1. The lecturer uses a pen sensitive display in place of the traditional chalkboard. In addition to drawings, the electronic chalkboard can handle several types of multimedia elements from the Internet. The system records all actions and provides both live transmission and on-demand replay of the lecture from the Web. Remote students follow the lecture looking at the dynamic board content and listening to the recorded voice of the instructor. To record audio, E-Chalk integrates a Java-based Internet audio broadcast system, called World Wide Radio2 [4]. An optional video of the lecturer can also be transmitted. The system has been designed with the aim of recording lectures without the presence of an audio technician [8]. The system is currently in use in several universities. Looking at the different usage scenarios of E-Chalk [5] and similar lecture recording systems, several practical problems that deteriorate the audio quality can be observed. Simply adopting an Internet audio streaming system does not yield satisfying results.

2 Audio Recording in Classrooms

Unfortunately, there are many possible audio distortion sources in lecture halls and classrooms. Since it is not possible to mention them all here, this paper will concentrate on those that have the greatest impact on the recording. For a more detailed discussion of these problems see for example [3, 11].

The room is filled with multiple sources of noise: Students are murmuring, doors slam, cellular phones ring. In bigger rooms there may also be reverberation that depends on the geometry of the room and of the amount of people in the seats. The speaker's voice has not always the same loudness. Even movements of the lecturer can create noise. Coughs and sneezes, both of the audience and the speaker, result in irritating sounds. The loudness and the volume of the recording depend on the distance between microphone and the speaker's head, which is usually changing all the time. Additional noise is introduced by the sound equipment: Hard disks and fans in the recording computer produce noise, long wires can cause electromagnetic interference that results in noise or humming. Feedback loops can also be a problem if the voice is amplified for the audience.

The lecturer's attention is entirely focused on the presentation and technical problems can be overlooked. For example, the lecturer can just forget to switch the microphone on. In many lectures weak batteries in the microphone cause a drastic reduction of the signal to noise ratio, without the speaker noticing it. Many people have also problems with the operating system's mixer. 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 minutes even to experiences users.

Another subject is equipment quality. Some sound cards cannot deliver high fidelity audio recordings. In fact, all popular sound cards focus on sound playback but not on sound recording. Game playing and multimedia replays are their most important applications. On-board sound cards, especially those in laptops, have often very restricted recording capabilities.

The quality loss introduced by modern software codecs is perceptually negligible compared to the described problems. Improving audio recording for lectures held in lecture halls means first and foremost improving the quality of the raw signal before it is processed by the codec. Lecture recording will not become popular in educational institutions until it is possible to produce satisfactory audio quality with standard hardware and without a technician necessarily present.

3 Related Work

Speech enhancement and remastering is a field where a wide range of research and commercial products exist.

Windows Media Encoder, RealProducer, and Quicktime are the most popular Internet streaming systems. None of them provides speech enhancement or automatized control mechanisms. A possible reason is, that in typical streaming use cases, a high quality audio signal is fed in by professional radio broadcasting stations. An audio technician is assumed to be present. Video conferencing tools, such as Polycom ViaVideo[1], do have basic filters for echo cancelling or

[1] www.polycom.com

feedback suppression. The audio quality needed for a video conference is much lower than what is required for a recording. It is well known that the same noise is less irritating for a listener when it is experienced during a live transmission.

Especially in the HAM amateur radio sector there are several specialized solutions for enhancing speech intelligibility, see for example [15]. Although these solution have been implemented as analog hardware, the underlying ideas are often effective and many of them can be realized in software.

Octiv, Inc[2] applies real time filters to increase 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.

The SpeechPro Denoiser[3] software takes given audio recordings and processes them using different filters. The software is able to reduce noise and filter-out some problems, such as pops and clicks. Experience in using audio filtering applications is needed to successfully denoise tricky files. Generic remastering packets like Steinberg WaveLab[4], Emagic Logic Pro[5], or Samplitude[6], even need an introductory seminar to be used.[7]

In academic research many projects seek to solve the Cocktail Party Problem [9]. Most approaches try to solve the problem with blind source separation and use extra hardware, such as multiple microphones.

Itoh and Mizushima [10] published an algorithm that identifies speech and non-speech parts of the signal before it uses noise reduction to eleminate 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.

The Virtual Director [12], developed at UC Berkeley, also helps automatizing the process of producing internet webcasts. It saves man power by enabling several webcasts to be run by a single technician. The system selects which streams to broadcast and controls other equipment such as moving cameras to track the speaker.

Davis is researching the automatization of media productions at UC Berkeley, see for example [2]. His work, however, aims more at automatizing video direction and editing.

4 Enhancing Audio Recordings

The software described in this paper focuses on the special case of lecture recording. Not all of the problems mentioned above can yet be solved only by software.

[2] www.octiv.com
[3] www.speechpro.com
[4] www.steinberg.net
[5] www.emagic.de
[6] www.samplitude.com
[7] Steinberg also provides easy to use software for special purposes like My MP3 Pro, but none is available for live recording.

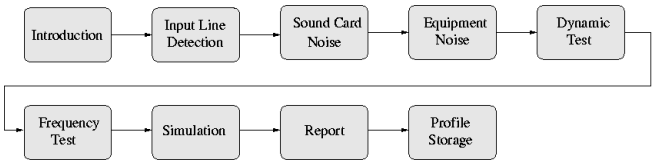

Figure 2: The steps of the audio profile wizard.

The system relies on the lecturer using some kind of directional microphone or a headset. Headsets provide good quality but they restrict the mobility of the speaker. They eliminate the influence of room geometry and of cocktail party noise.

A lecture recording system has the advantage that information about speaker and equipment are accessible in advance. Utilizing this divides the approach into two parts:

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

2. During recording, filters, hardware monitors, and automatic gain control work with the information collected by the expert system.

4.1 Setup

Before lectures are recorded, the user 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 2. This setup takes about three minutes and has to be done once per speaker and sound equipment. Each speaker uses his audio profile for all his recordings.

4.1.1 Detection of Sound Equipment

The setup screen asks the user to assemble the hardware as it is to be used in the lecture. The wizard detects the sound card and its mixing capabilities. 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 a certain threshold, 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

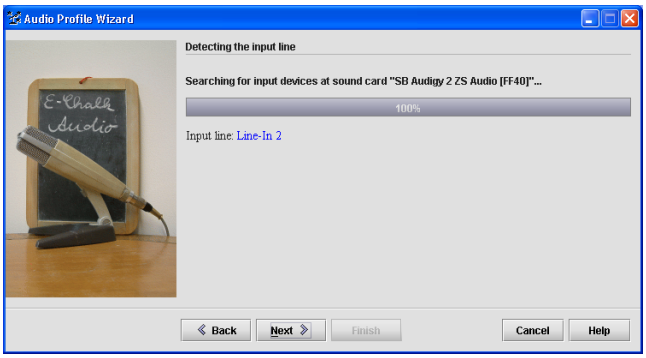

Figure 3: Soundcard ports are scanned for input devices.

be detected, for example if noise is constantly at zero decibel there is a short circuit.

The audio system analyzer takes control over the sound card mixer. There is no need for the user to deal with the operating system's mixer.

4.1.2 Recording of Background Noise

The next step is to record the sound card background noise. The user is asked to pull any input device out of the sound card[8]. A few seconds of noise are recorded. The signal is analyzed to detect possible hardware problems or handling errors. For example, overflows or critical noise levels result in descriptive warnings.

After recording sound card noise level, the user is asked to replug and switch-on the sound equipment. Again, several seconds of "silence" are recorded and analyzed. Comparing this signal to the previous recording exposes several handling and hardware errors. For example, a recording device plugged into the wrong input line is easily detected.

4.1.3 Dynamic and Frequency Tests

After having recorded background noise, the user is asked to record phrases with special properties. They are language dependent. A phrase containing many explosives[9] is used to determine the range of the gain. This measurement of the signal dynamics is used to adjust the automatic gain control. By adjusting the sound card mixer's input gain at the current port, the gain control levels out the signal. The average signal level should be maximized but overflows must be avoided. If too many overflows are detected, or if the average signal is too low, the user is informed about possible improvements.

[8] On notebook computers this is not always possible, because build-in microphones cannot always be switched off. The wizard then adjusts its analysis process.

[9] In English, repeating the word "Coffeepot" gives good results.

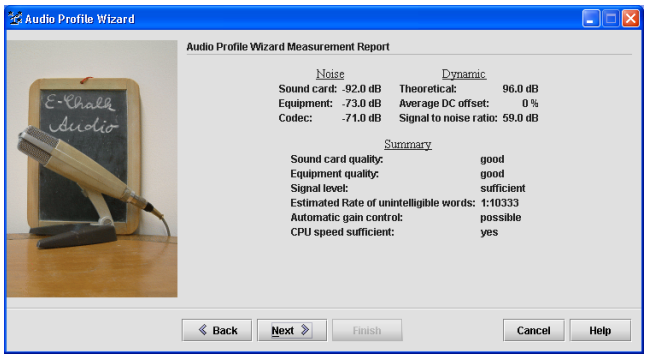

Figure 4: A report gives a rough estimation of the quality of the equipment. Word intelligibility is calculated according to IEC 268.

During the frequency test, a sentence containing several sibilants is recorded to figure out the upper bound frequency. The system looks at the frequency spectrum to warn the user about equipment anomalies.

4.1.4 Fine Tuning and Simulation

The final recording serves as the basis for a simulation and allows fine tuning. The user is asked to record a typical start of a lecture. The recording is filtered (as described in Section 4.2), compressed, and uncompressed again. The user can listen to his or her voice as it will sound recorded. If necessary, an equalizer (according to ISO R.266) allows experienced users to further fine tune the frequency spectrum. The time for filtering and compressing is measured. If this process takes too long, it is very likely that audio packets are being lost during real recording due to a slow computer.

4.1.5 Summary and Report

At the end of the simulation process a report is displayed, as shown in Figure 4. 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 is graded using background noise and the card's DC offset[10] calculated from the recordings. The grading of the equipment is based on the background noise recordings and the frequency shape. This is only a rough grading, assisting non expert users to judge the equipment and identify quality bottlenecks. Further testing would require the user to work with loop back cables, frequency generators, and/or measurement instruments.

[10] A high DC offset implies a low quality of the card's analog digital converters.

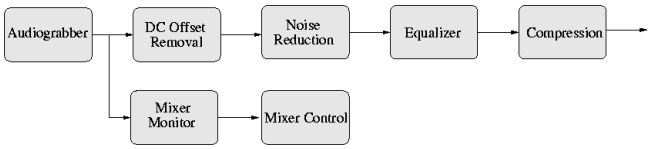

Figure 5: The system during recording.

Among other information, the created profile contains all mixer settings, the equalizer settings, the recordings, and the sound card's identification.

4.2 During Recording

For recording, the system relies on the profile of the equipment. If changes are detected, for example a different sound card, the system complains at start up. This section describes how the recording profile is used during the lecture. Figure 5 illustrates the signal's processing chain.

4.2.1 Mixer Monitor

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 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.

4.2.2 Mixer Control

The mixer control uses the values of the dynamic test to level out the input gain using the sound cards mixer. The analog preamplifiers of the mixer channels thus work like 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 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.

Mixer 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.

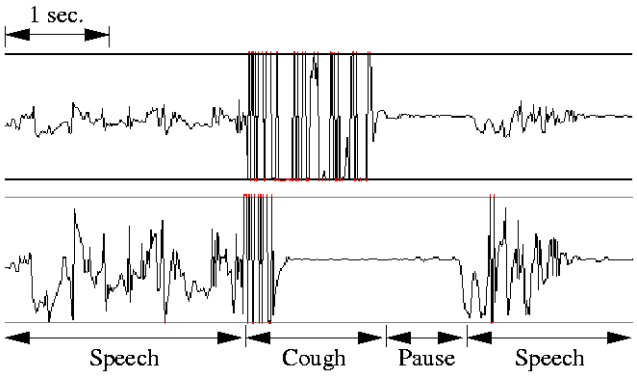

Figure 6: Without (above) and with (below) mixer control: The speech signal is expanded and the cough is leveled out.

4.2.3 Noise Reduction

The signal's DC offset is removed, the sound card background noise level recorded in the profile is used as threshold for a noise gate and the equipment noise as a noise fingerprint. 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 [11]. The best match is subtracted [1]. See Figure 7 for an example. It is not always possible to pre-record the 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 wished.

4.2.4 Final Processing

Equalizer settings are applied before the normalized signal is processed by the codec.

The filtering also results in a more efficient compression. Because noise and overflows are reduced, entropy also scales down and the compression can achieve better results.

Several codecs have been tested, for example, a modified version of the ADPCM[12] algorithm and codecs provided by the Windows Media Encoder.

[11] A typical situation that changes humming is when the light is turned on or off.
[12] See ITU-T recommendation G.726

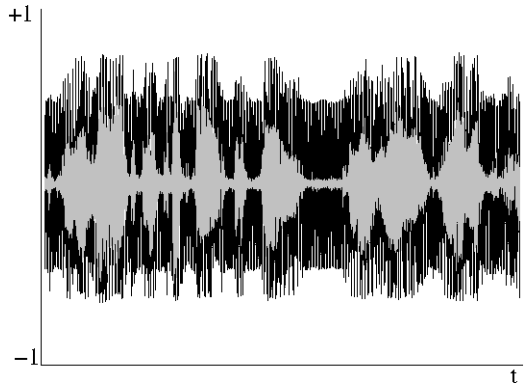

Figure 7: Three seconds of a speech signal with a 100 Hz sine-like humming before (black) and after filtering (gray).

5 Experiences

Since the summer of 2001 the lecture recording system E-Chalk has been in regular use. However, many of the lectures were produced without audio track, because setting up audio properly was too difficult for many users. The new automated audio system was tested during an algorithm design course in winter term 2004. A headset was used for recording and the system was tested both under Windows and Linux. Even though having a setup time of three minutes once was at first considered cumbersome, opinion changed when the lecturer was saved from picking up the wrong microphone. Common recording distortions were eliminated and the listeners of the course reported a more pleasent audio experience.

The system has now been integrated into the E-Chalk system for deployment and will be in wide use starting this winter term 2004/2005.

6 Summary and Perspective

The system presented in this paper improves the handling of lecture recording systems. An expert system presented via a GUI wizard guides the user through a systematic setup and test of the sound equipment. The quality analysis presented cannot substitute high-quality hardware measurements of the sound equipment but provides a rough guideline. Once initialized, the system monitors and controls important parts of the sound hardware. A range of handling errors and hardware failures are detected and reported. Classical recording studio equipments like graphical equalizers, noise gates, and compressors are simulated and automatically operated.

This 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 record-

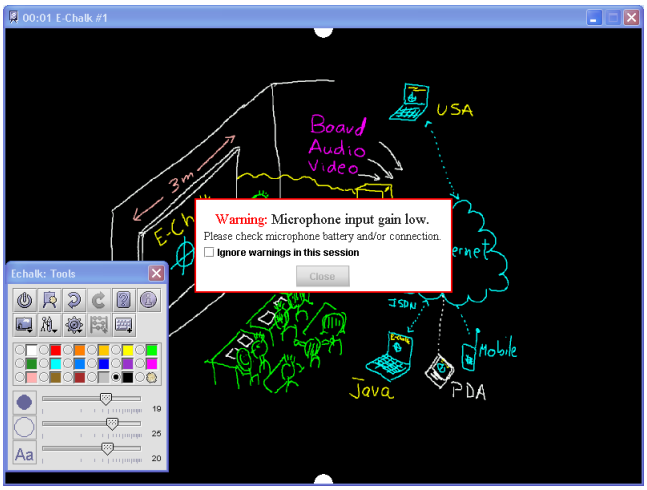

Figure 8: The microphone's floor noise level has sunk - batteries have to be changed.

ing, however, it eases the handling of lecture recording and improved the overall quality of the recordings without requiring cost-intensive technical staff.

In the future the system should be capable of handling multiple microphones and inputs to enable switching between classroom questions and lecturer's voice. A further advancement possible then is to integrate a fast blind source separation engine to reduce cocktail party noise eliminating the requirement of a directed microphone. One would also like to interface with external studio hardware, such as mixer desks, to enable auto operation. An in-deepth user evaluation should then be made for further improvement of the system and to gain more detailed feedback.

Freeing users from performing technical setups by automatization, as recently observable in digital photography, is still a challenge for audio recording.

7 Contributors

The system presented in this paper is being developed as part of the E-Chalk system that is developed at the Freie Universität Berlin, Institut für Informatik, ZDM (Center for Digital Media), led by Raúl Rojas. Gerald Friedland is member of the E-Chalk team and conceived and designed the audio system with the help of student member Kristian Jantz. Lars Knipping, who is, among other components, developing the board component gave very helpful advises when building the GUI of the wizard.

References

[1] S. Boll. Supression of acoustic noise in speech by spectral substraction. *IEEE Transactions, ASSP*, 28(2):113–120, 1979.

[2] M. Davis, J. Heer, and A. Ramirez. Active capture: automatic direction for automatic movies. In *Proceedings of the eleventh ACM international conference on Multimedia*, Berkeley, CA (USA), November 2003.

[3] M. Dickreiter. *Handbuch der Tonstudiotechnik*, volume 1. K.G. Saur, Munich, Germany, 6th edition, 1997.

[4] G. Friedland, L. Knipping, and R. Rojas. E-chalk technical description. Technical Report B-02-11, Fachbereich Mathematik und Informatik, Freie Universität Berlin, May 2002.

[5] G. Friedland, L. Knipping, and R. Rojas. Mapping the classroom into the web: Case studies from several institutions. In C. T. András Szüks, Erwin Wagner, editor, *The Quality Dialogue: Integrating Cultures in Flexible, Distance and eLearning*, pages 480–485, Rhodes, Greece, June 2003. 12th EDEN Annual Conference, European Distance Education Network.

[6] G. Friedland, L. Knipping, R. Rojas, and E. Tapia. Web based education as a result of ai supported classroom teaching. In *Knowledge-Based Intelligent Information and Engineering Systems: 7th International Conference, KES 2003 Oxford, UK, September 3-5, 2003 Proceedings, Part II*, volume 2774 of *Lecture Notes of Computer Sciences*, pages 290–296. Springer Verlag, Heidelberg, September 2003.

[7] G. Friedland, L. Knipping, J. Schulte, and E. T. a. E-chalk: A lecture recording system using the chalkboard metaphor. *International Journal of Interactive Technology and Smart Education*, 1(1), 2004.

[8] G. Friedland, L. Knipping, and E. Tapia. Web based lectures produced by ai supported classroom teaching. *International Journal of Artificial Intelligence Tools*, 13(2), 2004.

[9] S. Haykin. Cocktail party phenomenon: What is it, and how do we solve it? In *European Summer School on ICA*, Berlin, Germany, June 2003.

[10] K. Itoh and M. Mizushima. Environmental noise reduction based on speech/non-speech identification for hearing aids. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Munich, Germany, April 1997.

[11] B. Katz. *Mastering Audio: The Art and the Science*. Focal Press (Elsevier), Oxford, UK, 2002.

[12] E. Machnicki and L. Rowe. Virtual director: Automating a webcast. *SPIE Multimedia Computing and Networking*, January 2002.

[13] R. Rojas, L. Knipping, G. Friedland, and B. Frötschl. Ende der Kreidezeit - Die Zukunft des Mathematikunterrichts. *DMV Mitteilungen*, 2-2001:32–37, February 2001.

[14] R. Rojas, L. Knipping, W.-U. Raffel, and G. Friedland. Elektronische Kreide: Eine Java-Multimedia-Tafel für den Präsenz- und Fernunterricht. In *Tagungsband der Learntec*, volume 2, pages 533–539, October 2001.

[15] Universal Radio, Inc. MFJ-616 Speech Intelligibility Enhancer Instruction Manual. www.hy-gain.com/man/mfjpdf/MFJ-616.pdf.