

# Development of multimedia learning software for MP3 Audio Coding with special regard to usability principles and multimedia learning problems

Daniel Pape  
ZAS BERLIN  
Berlin, Germany  
E-mail: pape@zas.gwz-berlin.de

Gerrit Kalkbrenner  
Universität Dortmund

**Abstract:** An e-learning program explaining “Perceptual Audio Coding” (famous representative: MP3) was specified, designed and implemented by the first author. The final software (750MB) is split into different modules. These include: (1) An auralization and explanation of the filterbank implemented in MP3 and (2) implementation of simulations of various classic psychoacoustic experiments (mainly masking) for three different music styles. Other implemented audio examples exhibit (1) a comparison of the sound quality of a Fraunhofer MP3 codec at different bitrates and (2) a comparison of today’s most important audio and speech codecs (like WindowsMediaEncoder and Real9) at different bitrates and (3) audio examples and explanation of typical error signals. Finally a structured explanation of the mode-of-operation of an MP3 encoder and technical papers with further references to publications on perceptive audio coding were included in the presented software. Special focus during the development was set at the network compatibility (internet and intranet) and the strict attention to software usability criteria and multimedia learning theories.

## 1. INTRODUCTION

Exploring the majority of multimedia learning software two biases are found. On the one hand exist plain adaptations of textbooks with no added value of multimedia against a traditional textbook. On the other hand there is a huge amount of multimedia software which excessively use multimedia means in a way that it leads to multimedia overkill. Often the multimedia developer seem to follow the principle “Multimedia as much as possible, not where it seems useful”. This has the effect that the learner gets lost in the program, is stressed and unsatisfied with the presentation of the content and the therefore with the learning process. Therefore good multimedia learning software should use the benefits of multimedia content to guide the learner actively and provide practical examples where needed to understand the underlying principles of the content to be learned. If one thinks of subjects for the development of e-learning software, Perceptual Audio Coding would be a thankful candidate. The practical consequences of this lossy audio compression call for a connection between theoretical

explanation (in technical papers) and practical explanation (consisting in audio examples explaining the underlying psychoacoustic phenomena). Some essentially important questions can only be explained practically - through the extensive use of audio examples, for example:

- Which bitrate should I choose to deteriorate my audio signal as little as necessary?
- How does it work exactly: I compress a file to one fifth of its original size and I cannot hear a difference between the original and the compressed version, even under professional monitoring conditions?
- Which codec works best with my kind of audio signals?
- Though masking is the most important process within coding, it remains to be an obscure phenomenon. What is the sound of masking?
- Which new error signals - formerly unknown to the sound engineer – can occur during the coding process?
- Are there differences in audio quality of different MP3 encoders?
- If yes: Are there also differences in audio quality of various MP3 decoders?

## 2. MOTIVATION

Answering these questions there are only few approaches in the multimedia domain:

- The AES produced the CD-ROM “Perceptual Audio Coding: What to listen for” [13] where typical error signals are generated and explanations are given.
- The IRT produced the CD-ROM “AIDA – Audio Internet Demonstration Aid” [14] where they provided a comparison of all important codecs at different bitrates with a wide choice of audio signal.

Both the AES and the IRT subjects must be a substantial part of any software concerning audio coding, but some basics must be introduced beforehand to understand the underlying principles of perceptual audio coding.

There is no approach for example to demonstrate in practice the important classic psychoacoustic experiments of Zwicker and others (like masking) [11] and to explain them non-academically. But since masking is the essential part of a perceptual audio codec, hence it is obvious to include explaining simulations of these classic psychoacoustic experiments in the e-learning software. In preliminary tests I realized that it was much more easy to explain the principles of MP3 when students made some aural experience beforehand, e.g. that a sine wave signal could be completely masked by a narrow band noise signal even if the level of the sine wave was 10 dB higher!

My software is aiming at two user groups: (1) sound engineers and audio professionals and (2) research and education. These groups would be interested in understanding the underlying psychoacoustic phenomena and audible decrease of sound quality and the sound of special error signals, introduced by perceptual audio coding. For their work they would appreciate sound examples out of a comparison of different bitrates and codecs.

A real advantage against normal textbook-learning could only be guaranteed if the attention was drawn to the different theories of multimedia learning [4] [9]. If video sequences and audio examples are redundant or interfere with the learning process then the advantage of a e-learning program is put at stake. There are numerous examples of weakly designed and implemented "multimedia learning software": For example there is software for foreign language learning where the user has just to fill in missing words and the only error correction the software provides is right-wrong counting. It is hard to see an advantage of this kind of e-learning against traditional foreign language education books.

With e-learning programs the self-explaining navigation through the program structure plays an important role. The learning process is strongly hampered if the navigation is a task of its own! If the navigation has to be learned first (an exception is: the navigation is already learned like the well known icons and navigation structure in Windows or Microsoft programs) the user would most probably reject the software. Addressing this problem is called "usability" which has rapidly developed in the last years through the interest of major enterprises in "user-friendly" websites [5] [6] [7] [8]. In the internet it is very easy for the customer to leave websites or webshops immediately. The potential customer always prefers the more user-friendly site. This results in financial losses of the "user-hostile" website or webshop.

Therefore the software should be developed in accordance with these main criteria:

- a comprehensive non-academic explanation of perceptual audio coding, including all the research done in the last years and providing a sufficient amount of practical audio examples to illustrate convincingly the theoretical explanations.

- Attention to the different multimedia learning theories and exploitation of the different media to speed up the learning process and strengthen the user. Attention to usability and navigation guidelines are certainly included in this criteria.

### 3. THEORETICAL BACKGROUND

#### 3.1. Perceptual Audio Coding

Explanation of perceptual audio coding is given in detail in the work of Prof. Brandenburg and others and in the standards. Hence, for the technical understanding of perceptual audio coding the references [1] [2] [3] should be consulted.

#### 3.2. Multimedia learning theories

The classic learning theories like the Behaviorism, the Cognitivism and the Constructivism were studied and adapted for the use with multimedia content. Taking these theories into account influences directly the way of designing multimedia learning environments. It is appropriate to adapt the theory of constructivism when designing multimedia software which includes experiments and simulations. According to this theory the learner has to construct the knowledge itself. The role of the teacher changes more to the role of a coach who supports the learner during the knowledge constructing process. The main element of the constructivism based learning is the experiment. Within this, the learner tries several assumptions and discovers the underlying concept. The aim of constructivism based learning is the design of an experimenting environment, which allows to follow weak or strong guided experiments.

In the presented learning environment about perceptual audio coding I designed some experimenting scenarios, where the learner is allowed to change several parameter of audio compression. Within this and during direct feedback with pre-recorded audio examples and explaining comments the learner is able to understand the concept of perceptual audio coding in a deeper sense.

A common mistake is (as described) the careful use of multimedia. It is obvious that research of the influence of multimedia on the learning process has to be regarded. The advices of Schnotz and others for the added-value-use of multimedia in learning environment [9] were followed during the design of the software to guarantee a maximum amount of learning success and to avoid common errors.

#### 3.3. Principles of Usability

Questions of software usability arose extensively in the last years. The reason was mainly that websites and e-shops were forced to develop user-friendly sites to keep the consumer and visitor. If they failed, the user went without hesitation to the next e-shop, where the navigation was more intuitive and

understandable [6]. Software developers profited from these usability guidelines to optimize their products.

In the presented program a main focus was set to the user friendliness and usability guidelines. That means that during the conception and development of the program the following common errors had to be avoided:

Usability error	Description
Missing message	If the screens are over-filled with content and super fluid design elements the user is not able to recognize his main task.
Lack of possibilities for orientation	In systems with non-linear navigation it is necessary to provide orientation information to the user. Otherwise it is probable that he will be lost in the structure while navigating.
Missing or erroneous menu structure	It is important to structure the content in a way that the user will understand immediately the structure and can navigate without cognitive load.
Graphic overkill	If the design of the graphics is like a baroque architecture the user could lose the focus and is not able to solve his task.
Extensive loading time (without process information)	During loading time without process information it is difficult for the user to decide whether the software is crashed or is loading.
Erroneous structure of content	Content which has to be read at the screen has to be structured differently than content for books and articles.

Table 1: common usability errors and their description following [5],[6] and [8]

#### 4. STRUCTURE AND TECHNICAL ASPECTS

To allow the user immediate access to his field of interest the program was designed a modular one. In the following the different modules will be explained in detail.

##### 4.1. The modules of the program

###### 4.1.1. Intro

To make the user sensitive for the topic “perceptual audio coding” I designed an introducing sequence where a synchronized decrease of graphic image quality and sound quality was composed: A high-quality image was shown together with the sound of a state-of-the-art pop recording. A message “Quality of the graphical image:100%” and “Quality of the sound: 100%” was shown on top of the screen. The quality of both the graphic image (achieved by the use of

graphic filters) and the sound (achieved by a decreasing bitrate of a MP3 encoder) was decreased stepwise every four beats of the music (the message on top of the screen displayed the decrease in percent values). At the end of the intro the structure of the image was completely fuzzy and the quality of the sound has the bitrate of 8kbts/s. The last four bars set the image quality and the sound quality with one step to 100% to illustrate the quality loss. The Intro can be skipped anytime with a “Skip intro” button.

###### 4.1.2. Simulation of psychoacoustic experiments

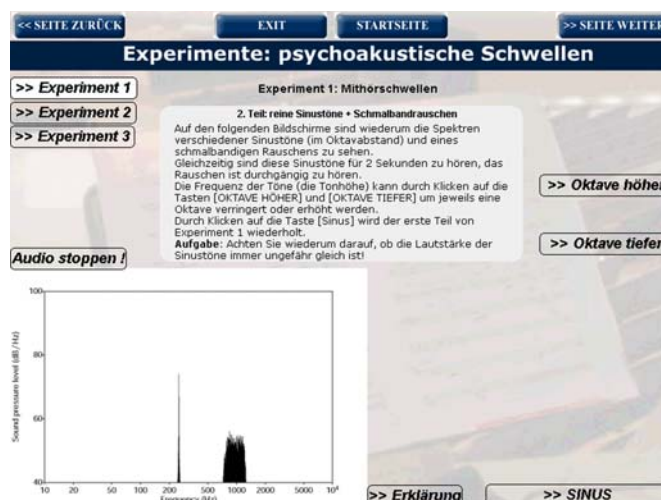


Figure 1: screenshot of the module “Psychoacoustic experiments and simulations”

Three classic psychoacoustic experiments important for the explanation of perceptual audio coding were designed:

In the first experiment the user could change the frequency of a sine wave in octave steps from 63Hz to 16kHz. The user should listen carefully if the loudness was the same for all presented discrete sine waves. In the second part of the experiment a narrow-band-noise (NBN) with the center frequency of 1kHz and a bandwidth of 150Hz was constantly audible. The user could listen to the same sine waves as in the first part of this experiment (frequency also changeable in octave steps). The amplitude level of the NBN was chosen in a way that the sine wave at 1kHz is completely masked by the NBN. The user should listen once more to the loudness of the sine waves and would find that all sine waves had the same loudness except the sine wave at 1kHz which is not audible. Figure 1 shows a screenshot of the described module.

In the second experiment the user could listen to three different music styles (rock, percussive samba, orchestral classic). He could rise and lower the level of an additional audible narrow-band-noise (the same as in the preceding experiments) in 3dB steps. He was invited to adjust the NBN

level until the noise was just noticeable. Then he should switch between “music with noise” and “only noise” to compare the perception of the loudness of the noise in both conditions. The comparison clearly shows that the level of the NBN in “only noise” is perceived considerably higher than in “music with noise”. This shows the effects of the masking of the music very clearly.

#### 4.1.3. Simulation of a MP3 filterbank

To auralize the split of the input audio signal into different frequency bands a simulation of a MP3 filterbank was implemented in the program. The conceptual background was that audio engineers and other audio professionals are used to the sound of filters and equalizing curves. But students and non-audio-professionals might have difficulties to understand practically that some frequency bands could be removed from the audio signal without audible notice (for example the high frequency region exceeding 8kHz for speech transmission). Therefore a filterbank equal to the MP3 filterbank was designed and the mentioned music examples were filtered. Clicking into one of the frequency bands the user will hear the correspondingly filtered signal of the music example. This enables him to detect the relevant frequency bands of a music example and - what’s more important – to decide which frequency bands do not contribute relevantly to the sound of the music. The latter decision determines which sub-band could be neglected without audible distortion. During user testing in the phase of design and implementation it appeared that this module was important for the understanding of the principles of perceptual coding even for audio semi-professionals because not everyone works with sharp high-order bandpass filters to understand these filtering principles.

#### 4.1.4. Explanation of the mode-of-operation of a MP3 encoder

A module which explains the mode of operation of a typical MP3 codec was implemented. A block diagram of this encoder was used as a navigation bar on top of the mode-of-operation module. Each part of the block diagram can be clicked on to allow the user a quick non-linear navigation through the module. The convenience consists in the fact that the block diagram on top of the screen is permanently shown. It delivers perfect assistance to the explanations on the bottom of every page of the module. Figure 2 shows a screenshot of the described module.

#### 4.1.5. Audio examples

First, a sound comparison of a typical MP3 encoder at different bitrates was implemented, aiming at sound engineers who are faced with the task to compress an audio file in the best possible sound quality.

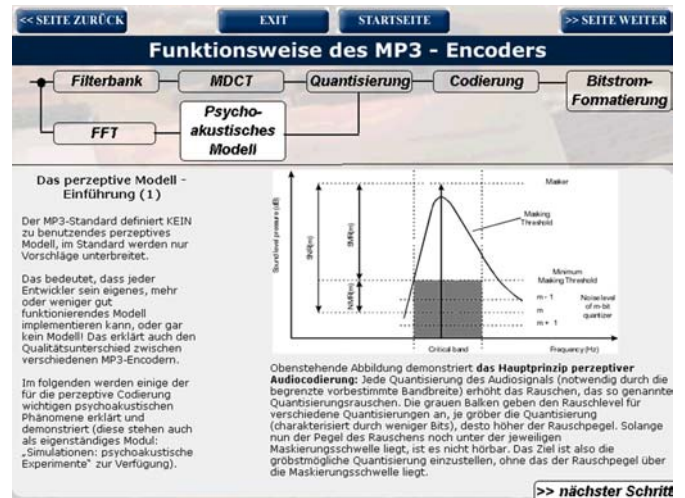


Figure 2: Screenshot of the module “Mode of operation of a MP3 encoder”

Secondly, error signals typical for perceptual audio coding were included. The conceptual background was the understanding that for decades audio professionals had been trained to hear hiss, hum and other error signals in audio. Perceptual coding introduces formerly unknown error signals like pre-echoes the sound character of which has to be learned by audio professionals from the very beginning. For every error signal an explanation page was written where the emergence of these error signals and the possibilities for their objective measurement are explained.

Thirdly, the program includes audio examples of a comparison of different commercially available audio codecs at the bitrate 56kb/s and 128kb/s. These consist of different sources, e.g. dry speech, speech embedded in music, jazz, hardly compressed pop music and other audio sources.

#### 4.2. Combining the modules: the navigation

To navigate between the modules and within the modules a combined linear and non-linear structure was chosen to satisfy the maximum amount of users. For the linear part it is possible to step throughout the complete program by using the “forward” and “backward” button.

The non-linear part consists of a tree structure where the user can jump to the start page at any time with one mouse click (where he can reach the other modules) or to the other parts of the module or he can exit the program. The aim of this start screen was to give access to all modules of the program. For demonstration purposes and non-linear navigation it seemed advisable to give a “sitemap” handle to users looking for one special module (for example only the comparison of different codecs). All navigation buttons were implemented as roll-over-images to ease the navigation. The momentary active navigation button is highlighted and cannot be clicked on.

### 4.3. The graphic layout – screendesign

The upper part of the screen was reserved for the navigation through the program. The module headline served as a division line between the navigation and the content of the program module. To stress the importance the headline was embedded in a dark blue frame. Intra-navigation in the modules was placed at the left part of the screen. The lower part of the screen always showed graphics and diagrams. In the modules “experiments” and “simulations” the task and explanations for the user were centered and framed in a different color. A background picture filled the complete screen. It was designed very softly to avoid reading difficulties.

### 4.4. Choosing the appropriate multimedia software

One condition was that the complete program should run without additional software. Secondly, synchronicity between audio and graphics should be guaranteed. In the development period of the software only Macromedia Director met these specifications and was therefore used. Moreover, Director includes a scripting language called Lingo which was necessary for the implementation of the programmed simulations and experiments.

### 4.5. Technical aspects: Formats, Platforms, Internet compatibility

All sound files used in the program were WAV files (16 bits, 44.1kHz) to guarantee playback stability for all kind of sound cards. All graphics were GIF for graphics with text (texture) and JPEG for photographs. To ensure the most possible graphic compatibility the complete program was designed in 8bit colors. The compiled program could be stored on one CD-ROM of 650MB, so program distribution was facilitated. An important point was the availability of the software for both PC and Macintosh computers (Macintosh computers are extensively used in the world of audio professionals). Director includes the option to compile a Director project for both Macintosh and PC.

The complete program was transferred to Macromedia Shockwave (a registered web player) with the option “no compression of audio” and can run in realtime over a normal fast internet connection.

## 5. CONCLUSION

Examining the multimedia programs explaining perceptual audio coding only few approaches exist. These approaches (see Motivation) do not provide the link between theory and practical application. Therefore the described software was designed and implemented to provide this link between theory and practical work for students and professionals working with audio and interested in the principles of perceptual audio coding.

Sound engineers and students working with the software were impressed by the provided complete explanation and

the given link to corresponding audio examples. Sound engineers who work every day with codecs even admitted that they finally understood how perceptual coding actually works, and also that this was the first time that they actually listened to some important psychoacoustic phenomena.

For the moment the program is only available in German. Further refinements of the software would include (1) the development of an English version, since for the moment the program is only available in German, (2) other modules like multichannel codecs (DolbyDigital, DTS) or objective testing of the sound quality of perceptual audio codecs.

Parts of the software containing the sine masking experiment can be downloaded from the program download region of the given URL in the references [12].

## 6. REFERENCES

- [1] Brandenburg, K. „An introduction to MPEG Layer 3”, EBU Technical Review, June 2000.
- [2] Brandenburg, K., “ISO-MPEG-1 Audio: A Generic Standard for Coding of High-Quality Digital Audio”, presented at the 92nd Convention of the Audio Engineering Society, Vienna, Austria, 1992, reprint J. Audio Eng. Soc., vol.42, pp.780-792 (1994 October).
- [3] ISO/IEC 13818-3: ISO 11172-3: Coding of Moving Pictures and associated audio for digital storage media at up to about 1.5 MBIT/s, Part 3 Audio, 1991.
- [4] Issing, L.J., “Instruktionsdesign für Multimedia” in Issing/Klimsa (Hrsg.): Information und Lernen mit Multimedia”, Weinheim, 1997.
- [5] Krug, S., „Don’t make me think! A common Sense Approach to Web Usability“, Peason Education Inc., 2000.
- [6] Manhartsberger, M. und Musil, S., „Web Usability - Das Prinzip des Vertrauens“, Galileo Design, 2001.
- [7] Petersen, M. G., „Towards Usability Evaluation of Multimedia Applications“, ACM, 1998.
- [8] Puscher, F., „Das Usability-Prinzip“, dpunkt-Verlag, 2001.
- [9] Schnotz, W. und Vosniadou, S. und Carretero, M. (Hrsg), „New Perspectives on Concept and Change (Advances in Learning and Instruction Series)“, Oxford: Elsevier, 1999.
- [10] “La red – Cómo cambiarán nuestras vidas los nuevos medios de comunicación“, Santillana, S.A. Taurus, Madrid, 1998.
- [11] Zwicker, E.: „Psychoacoustics - Facts and Models“, Springer Verlag, 1990.
- [12] URL: <http://www.audiocodierung.de/program.htm>
- [13] CD-ROM: “Perceptual Audio Coding: What to listen for“, Educational CD-ROM of the AES Technical Committee on Coding of Audio Signals, 2002
- [14] CD-ROM: “AIDA: Audio Internet Demonstration Aid”, Educational CD-ROM of the Institut für Rundfunktechnik (IRT), 2002