AN AUTOMATIC EQUALIZATION ALGORITHM FOR AUDIO

Steven van de Par, Jeroen Breebaart, Martin McKinney
Philips Research, Digital Signal Processing Group,
High Tech Campus 36, 5656 AE, Eindhoven, The Netherlands

ABSTRACT

We present a method for automatically equalizing music and audio signals based on information extracted from the signal. The main goal is to enhance the perceived audio quality especially if the sound quality of the audio signal is unsatisfactory. This method is based only on information contained within the audio signal, thus no further information is needed about recording conditions and equipment. The method adapts the equalizing in such a way that it reduces the self-masking within an audio signal, in this way increasing the audibility of the information contained in the audio signal. Assuming that audio quality increases with the audibility of information within the audio signal, this approach will lead to an improved audio quality.

1 INTRODUCTION

Audio travels through a considerable chain of transductions and devices starting from its acoustical or electronic generation until it becomes encoded on a Compact Disk or in some kind of digital audio format. At various points in this chain, linear filtering operations occur that can potentially reduce the audio quality. One of the most important points at which this filtering can occur is when the audio is recorded with a microphone; the transfer function of the microphone and recording room can have a significant effect on the timbre of the resulting reverberated audio. Also in the recording studio, the tonal balance may be adjusted using equalizers. Although the latter type of equalizing is intended to improve audio quality, the equalizer setting that is chosen by the recording engineer will be partially dependent on the reproduction equipment and studio; i.e. the equalizing can sometimes componensate for the particularities of the studio and loudspeakers. Additional modifications of the audio timbre may occur due to other stages in the recording chain such as e.g. amplifiers, but these modifications are usually rather small.

It is often difficult to determine how to compensate for timbre changes in the recording chain such as those due to the room responses, because in part these changes are integral part of what is captured in the recording. For recordings of classical orchestras, the concert hall acoustics contribute significantly and positively to the auditory impression that should be conveyed to the listener.

Nevertheless, when listening to recorded audio, as a listener, one can sometimes have a clear impression that the tonal balance is not optimal and depending on the recording, one can prefer to adjust the tonal balance. In this contribution, we propose an approach to automatically adjust the equalizer setting to improve the overall perceived audio quality. The method is based purely on an analysis of the audio signal itself without any additional information about the recording chain or content type (e.g. speech, rock, jazz, or classical).

2 AUTOMATIC EQUALIZING BASED ON REDUCING MASKING

The approach that we want to follow is based on some insights from the work of Janssen and Blommaert [1, 2] on image quality. According to these insights, perceived quality is a compromise between naturalness and usefulness. In the context of audio, optimal naturalness would mean that the audio, including the room acoustics of the recording room, would be represented without any further adjustments in the equalizing. Usefulness is more difficult to define in the context of audio, but one sensible interpretation would be that an audio track is optimally useful when the maximum amount details present in the track can be perceived by the listener. Since the optima for naturalness and usefulness do not necessarily coincide, one equalizer setting may optimize naturalness while another setting may optimize usefulness. Therefore, it is not a-priori clear what setting is optimal for audio quality. However, following the insight of Janssen and Blommaert [1, 2], we should make a compromise between both optima in order to achieve optimal perceived quality.

Although the aforementioned expose may seem impractical, we will propose a practical method that attempts to optimize audio quality using the insights outlined before.

In auditory perception, masking is a well known phenomenon. It refers to the situation where a certain sound that would be perfectly audible when presented in quiet, is not audible due to presence of some other, masking sound. In a complex audio signal, it may well happen that certain frequency components of the audio are not audible because they are masked by other frequency components in the audio that are present simultaneously. By amplifying these masked components, they could become audible.

When the tonal balance of a certain recording is rather poor, typically some regions in the spectrum are either too dominant or alternatively, too weak to be clearly audible. When certain spectral regions are very dominant they will tend to mask neighboring spectral regions, while spectral regions that are too weak will also tend to be masked by neighboring regions. As a result, part of the spectrum will not be audible for the listener. This reduces the amount of information that is conveyed to the listener, which makes the audio less useful. By modifying the spectrum such that these masked regions become audible again, the usefulness of the audio is improved.

When such equalizing, directed towards improving usefulness, is applied without any reserve, the timbre of instruments might be severely modified and the perceived naturalness may suffer significantly. Therefore, the equalizing should be more moderate than what is needed for optimal usefulness.

In the following section, an algorithm will be outlined that attempts to minimize the mutual masking effect of simultaneous frequency components in an audio signal.

3 AN AUTOMATIC EQUALIZING ALGORITHM

The algorithm that was developed for automatically equalizing is shown in Fig. 1. The input signal is filtered with an equalizer (B) resulting in the output signal. In the initial state of the algorithm, the equalizer (B) will be set to a flat response.

The input signal is also sent into the control unit that has the function of adjusting the equalizer settings. In the control unit the signal is fed through a second equalizer (A) that always has a filtering setting that is double the setting of equalizer (B). The output of equalizer (A) is split into short analysis frames which are subjected to two spectral analyses. In the first analysis, the so-called *excitation pattern* is derived (cf. [3]), which represents the activity pattern that can be observed on the basilar membrane in the inner ear due to the equalized input signal. The second spectral analysis determines the energy per critical band of the signal spectrum. By dividing the first by the second representation, a 'masking strength' estimate is obtained regarding the degree to which the signal in each critical band is masked. The larger the excitation at one place on the basilar membrane (first spectral representation), the more masking can be expected, however, this is counteracted by the signal energy present in the corresponding critical band itself (second spectral representation).

The equalizers are now adapted to reduce the amount of masking in each critical band. For this purpose the estimated masking strength is used to create an incremental adjustment of the equalizer settings. A frequency band that receives a relatively large amount of masking will be increased in level using the equalizers, making it better audible, while the reverse happens for bands with relatively little masking. The process of incremental adjustments is done once for each frame, but continues across frames until it converges after some time. When this process converges, the equalizer (A) has been adjusted such that all spectral regions are, on average, equally well audible, resulting in optimal usefulness. For optimal naturalness we will simply assume that a flat equalizer setting would be optimal due to lack of any better information. The compromise between optimal naturalness and optimal usefulness is obtained by taking an equalizer setting that is the average of the settings for both criteria; i.e. only half of the total equalizing effect that is derived for optimal usefulness should be used. This is achieved by setting equalizer (B) to half of the filtering effect of equalizer (A).

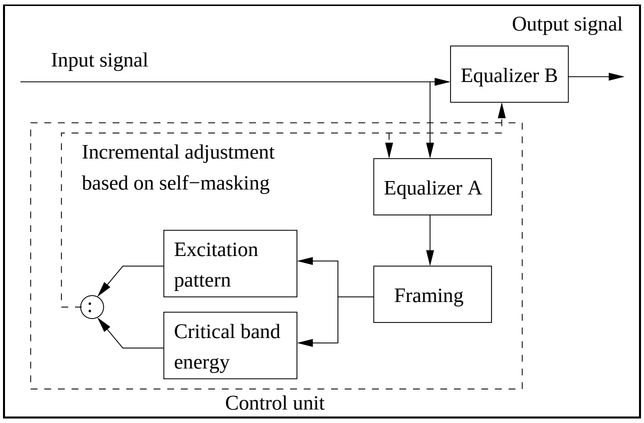

Figure 1. Outline of the automatic equalizing method. A control unit adapts the equalizer settings in order to reduce the self-masking that is determined by the ratio between the excitation pattern and the signal energy within each critical band.

4 CONCLUSIONS

We presented an approach to automatic equalizing with the purpose to improve audio quality using only information that is extracted from the input audio signal. The approach is based on reducing the amount of self-masking of the audio signal. Informal listening tests revealed that specifically for audio signals of poor audio quality we obtain an improvement in quality as a result of our algorithm, while for good quality audio, the algorithm applied just moderate equalizing adjustments.

5 REFERENCES

[1] Janssen, T., and Blommaert, F. "Image quality semantics", Journal of Imaging Science and Technology, **41**, pp. 555–560, 1997.

[2] Janssen, T.J.W.M. *Computational image quality*. Ph.D. thesis, Technische Universiteit Eindhoven, 2000.

[3] Moore, B.C.J., "An Introduction to the Psychology of Hearing," London: Academic Press. (2001)