Chapter XIII
Subjective and Objective Quality Evaluation of Watermarked Audio

Michael Arnold
Thomson, Germany

Peter Baum
Thomson, Germany

Walter Voeßing
Thomson, Germany

ABSTRACT

Methods for evaluating the quality of watermarked objects are detailed in this chapter. It will provide an overview of subjective and objective methods usable in order to judge the influence of watermark embedding on the quality of audio tracks. The problem associated with the quality evaluation of watermarked audio data will be presented. This is followed by a presentation of subjective evaluation standards used in testing the transparency of marked audio tracks as well as the evaluation of marked items with intermediate quality. Since subjective listening tests are expensive and dependent on many not easily controllable parameters, objective quality measurement methods are discussed in the section, Objective Evaluation Standards. The section Implementation of a Quality Evaluation presents the whole process of testing the quality taking into account the methods discussed in this chapter. Special emphasis is devoted to a detailed description of the test setup, item selection, and the practical limitations. The last section summarizes the chapter.

Evaluation of Watermarked Audio Data

The primary goal for evaluation techniques is to conduct valid and reliable tests which provide data for research, development, and quality control during deployment of the developed algorithms. Quality assessment of watermarked audio tracks is a significant challenge. No single objective metric to quantify the quality of an audio track...
Subjective and Objective Quality Evaluation of Watermarked Audio

carrying a watermark is currently available (Cox, Miller & Bloom, 2002).

Nevertheless, in the case of audio data, the problem of evaluating watermarked audio tracks is very similar to the problem of evaluating the quality of perceptual codecs. Both signal processing applications introduce distortions in the output signal which should be inaudible by exploiting psychoacoustic phenomena. Due to the underlying algorithm and the used psychoacoustic model, the types of distortions are:

- Artefacts according to block boundaries, since nearly all audio watermarking algorithms perform block wise processing of the audio signal (Arnold, Schmucker, & Wolthusen, 2003).
- Pre-echoes resulting from the smearing of noise in temporal domain due to block wise transformation from frequency to time.
- Increased roughness and modulation artefacts of the signals due to a lack of masking noise in silent regions of the audio signal.
- Artificial signals by addition of nonexistent frequency components.
- Localization artefacts by changing delays in different channels.

Correspondingly, quality benchmarking principles and test methods used along the development of perceptual codecs can be applied in investigating the effects of audio watermarking encoding algorithms. Nevertheless, care has to be taken if one uses quality assessment methods for perceptual codecs (see the section, Implementation of a Quality Evaluation).

The different test procedures can be roughly classified into human subjective measurements and objective evaluation methods. A further distinction in the selection of the right test procedure is based on the definition of the quality evaluation problem:

- Testing the transparency of an audio watermarking codec.
- Comparing different audio watermarking algorithms regarding the quality of encoded tracks.
- Rating the quality of watermarked audio tracks.
- Testing watermarked tracks with intermediate quality.

Subjective Evaluation Standards and Methods

Performing subjective listening tests are still the ultimate evaluation procedures to judge the codec quality evaluation. Standardized test procedures have been developed to maximize the reliability of the results of subjective testing. The next two sections describe standards and general methods developed to perform subjective quality testing of coded audio data.

The subjective evaluation procedures will be generally distinguished regarding testing of the transparency of the watermarked items, rating of the quality of the processed items with respect to the reference signal, and testing of watermarked items with intermediate quality.

Testing Transparency

The ABX Double Blind Test

If the impairments introduced by the coding procedure are very small, one can assume transparency of the watermarked signal. In this case a subjective evaluation test for nontransparency can be conducted by the so-called ABX Double Blind Test or brief ABX Test. This is in contrast to an additional rating of the quality of watermarked tracks (see the section, The ITU-R BS.1116 Standard).

In the ABX test, the listener has access to three tracks labelled A, B, and X. The tracks A