

RECOMMENDATION ITU-R BS.1657

Procedure for the performance testing of automated audio identification systems

(Question ITU-R 8/6)

(2003)

The ITU Radiocommunication Assembly,

considering

- a) that metadata will be accompanying most audio broadcast transmissions in the future;
- b) that the automatic generation of metadata will be necessary to offer a complete cost-efficient service in future;
- c) that automatic identification of audio items enables tracking of transmitted programme material;
- d) that different schemes for extraction of audio metadata are developed today;
- e) that ISO/IEC JTC 1/SC 29/WG 11 is currently finalizing schemes for the coding of metadata for multimedia data;
- f) that no quality assessment procedures for audio metadata extraction schemes have been standardized until now,

recommends

- 1 that the procedure described in Annex 1 should be used to evaluate the performance of automated audio identification systems.

Annex 1**Procedure for the performance testing of automated audio identification systems****1 Introduction**

In a time of ever-increasing databases filled with musical content, be it genuine audio material or associated metadata (“data about data”), the demand for tools to maintain this mass of data is also growing more urgent day by day. This desire is not only voiced by professionals, but also by the common Internet user and music-lover, who searches the web on numerous errands for her or his preferred musical style. In order to facilitate the retrieval of the desired material two different levels of abstraction are here discerned:

- The search for metadata that can more or less be extracted automatically from the audio content, such as instrumentation, melodic theme, rhythm, etc. An example application for this would be a query-by-humming system or the classification into genres, which is commonly used in recommendation engines.
- Automatic identification of titles, where only insufficient, unreliable or no metadata at all is available. An “essence” of the audio data is distilled and compared to a database of known material, thus creating a link to relevant metadata such as artist, song name, etc.

While the first mentioned class contributes mainly to the human interaction interface, the second topic finds its application also in the protection of rights by tracking radio programmes and Internet transactions. It is foremost in this latter context that algorithms fitting that profile are referred to as “fingerprinting” techniques.

2 Motivation

To meet the demands of the music business, the recognition rate of the applied fingerprinting technology must be high and withstand common alterations and modifications of the original audio content. For this purpose, the music business has acknowledged the need of quality assurance for audio identification systems by recently formulating a request for information on audio fingerprinting technologies.

The severity and urgency of this problem is also underlined by the fact that a number of different, often proprietary, solutions have appeared recently. All methods, however, face the same problems regarding their robustness to modification and deterioration of the original material. Although the original material may have changed by a number of processing steps or degradations, it nevertheless shall be recognized as the intellectual property of the artist and composer. This leads to the proposition that automated music identification should ideally be as precise and tolerant to signal modification as human perception and recognition. Beyond robustness to signal alterations, a good fingerprinting system should exhibit a small fingerprint size (considering that certain applications might require the storage of millions of fingerprints), fast fingerprint extraction and recognition and further desirable properties. It should be noted that robustness to signal alterations and compactness of fingerprint representation are two conflicting requirements which have to be reconciled by systems.

Consequently, in order to assess the quality of an automated audio identification system, a test environment has to be defined that covers different types of signal degradation in multiple degrees of severity and describes how to determine other essential system parameters. To allow the objective evaluation of identification systems, a unified test procedure is needed.

3 Quality parameters

For audio identification systems the following quality parameters have to be considered:

- Segment size of the audio material to be identified.
 - What portion of an item is necessary for the identification?
- Size of the fingerprint.
 - How many data (bytes) per item have to be stored in the database?
 - Is the size of the fingerprint constant or variable (with respect to the length of the item)?
- Size of the database.
 - How many items can be handled simultaneously by the system?