

RECOMMENDATION ITU-R BS.1693

Procedure for the performance test of automated query-by-humming systems

(Question ITU-R 8/6)

(2004)

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 query-by-humming systems enable a natural way to query audio databases;
- d) that different schemes for extraction of audio metadata are developed today;
- e) that Recommendation ITU-R BS.1657 – Procedure for the performance testing of automated audio identification systems, describes a procedure for the performance test of automated audio identification systems;
- f) that ISO/IEC JTC 1/SC 29 WG 11 is currently finalizing schemes for the coding of metadata for multimedia data;
- g) that no quality assessment procedures for audio metadata extraction schemes regarding melody recognition have been standardized until now,

recommends

- 1** that the procedure described in Annex 1 should be used to evaluate the performance of automated query-by-humming systems.

Annex 1**Procedure for the performance test of automated query-by-humming 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 these masses of data is also growing more urgent day by day. This desire is not only expressed 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 can be discerned here:

- The search for high-level metadata as a human listener would describe contents, e.g. melody, rhythm, timbre, instrumentation or genre. An example application for this would be a query-by-humming system, which can be used in recommendation engines.

- Extracting mid-level metadata for automatic identification of certain interpretations of musical contents. Descriptions of the technical features of the audio data (spectral contents, etc.) are distilled and compared to a database of known material, thus creating a link to relevant metadata such as artist, song name, etc.

For an overview of current state-of-the-art query-by-humming systems please refer to document, ISMIR 2002 (3rd International Conference on Music Information Retrieval, IRCAM – Centre Pompidou Paris, France, October, 2002).

2 Motivation

To meet the demands of the music business, the recognition rate of the applied query-by-humming technology must be high and withstand common alterations to the stored representations in the song database.

This problem is tackled by a number of different, often proprietary, solutions that have arisen recently ([Clarisse *et al.*, 2002], [Ghias *et al.*, 1995], [Haus and Pollastri, 2001], [Heinz and Brückmann, 2003]). All methods, however, face the same problems regarding their robustness to modifications of the original material. This leads to the proposition that automated query-by-humming systems should ideally be as precise and tolerant to signal modifications as human perception and recognition. Therefore, a sophisticated query-by-humming system has to be robust against different distortions regarding signal quality and variations from ideal melody inputs. Also a reliable handling of large song databases consisting of several thousands of songs should be provided.

Consequently, in order to assess the quality of an automated query-by-humming system, a test environment has to be defined that covers different types of signal modifications and describes how to determine other essential system parameters. To allow the objective evaluation of query-by-humming systems, a unified test procedure is needed.

3 Quality parameters

For the evaluation of query-by-humming systems the following quality parameters have to be considered:

Required audio input:

- Is it necessary to sing a certain part of the song or is it possible to sing any part?
- What is the minimal length of the input to provide a reliable result?

Size of data representation:

- How many data (bytes) per song have to be stored in the music database?

Size of the music database:

- How many songs can be held in the music database?