Music Information Acquisition, Preservation, and Retrieval (MUMT-611)

Audio Fingerprinting Summary

Denis Lebel – 110125645 – denis.lebel@mail.mcgill.ca

Introduction

The human ability to recognize a song from a very short excerpt is an astonishing process that researchers attempt to mimic for some years now. The technique developed to perform this task is called audio fingerprinting and roughly consists in creating unique and compact identifiers for audio signals that can be used for efficient recognition.

An audio fingerprint is a content-based signature summarizing an audio recording and allowing its identification. In order to be unique, the fingerprint is based on perceptually and acoustically relevant characteristics of the audio signal. Audio fingerprinting systems are composed of a collection of known fingerprints along with a query system.

The remainder of this document presents properties that are desirable in fingerprinting systems, usage modes and applications of audio fingerprints, and a general framework for fingerprinting systems as described in (Cano et al. 2005).

Properties

In order to adequately compare fingerprinting systems, a set of properties is used. The most important ones are briefly described here.

- **Accuracy**: Function of correct, missed, and wrong identifications.
- **Reliability**: Correct identification method.
- **Robustness**: Ability to *accurately* identify a signal.
- **Granularity**: Ability to identify signal from short excerpt.
- **Security**: Vulnerability to cracking.
- **Versatility**: Ability to identify a signal regardless of audio format.
- **Scalability**: Performance with very large databases.
- **Complexity**: Computational costs the system (lower = better).
- **Fragility**: Integrity verification (detection of changes in content).

A fingerprinting system should be designed taking into consideration these interrelated properties. Emphasis may be put on certain properties that are more relevant to the purpose of the system. In general the fingerprint should be a compact perceptual summary of the audio recording, invariant to distortions, and easily computable.

Usage Modes and Applications

The main purpose of audio fingerprints is to identify the content of an audio signal. It can also be used to verify the integrity of a signal to ensure its content has not been altered. Fingerprinting can join forces with watermarking to enhance the protection of the audio content. Moreover, some audio features can be extracted solely based on a fingerprint since it is itself meant to be a summary of the signal.

Applications of fingerprinting systems include monitoring and tracking of audio content at the distributor and/or consumer end. The transmission channel can be directly monitored to obtain statistics on the content distributed. Fingerprints can also help to automatically provide additional information on a musical work, such as the artist name, year of composition, or album cover.

General Framework

Audio fingerprinting systems can be described through a general framework composed of a fingerprint extraction component and a matching algorithm as shown in Figure 1.

The fingerprint extraction contains a front-end module used to extract features from the audio signal that are then used by the fingerprint-modeling block to create a unique and compact fingerprint. The fingerprint model is designed so that redundancies are reduced to a minimum. Figure 2 summarizes the general sub-steps necessary for the fingerprint extraction. For more details, please refer to (Cano et al. 2005).

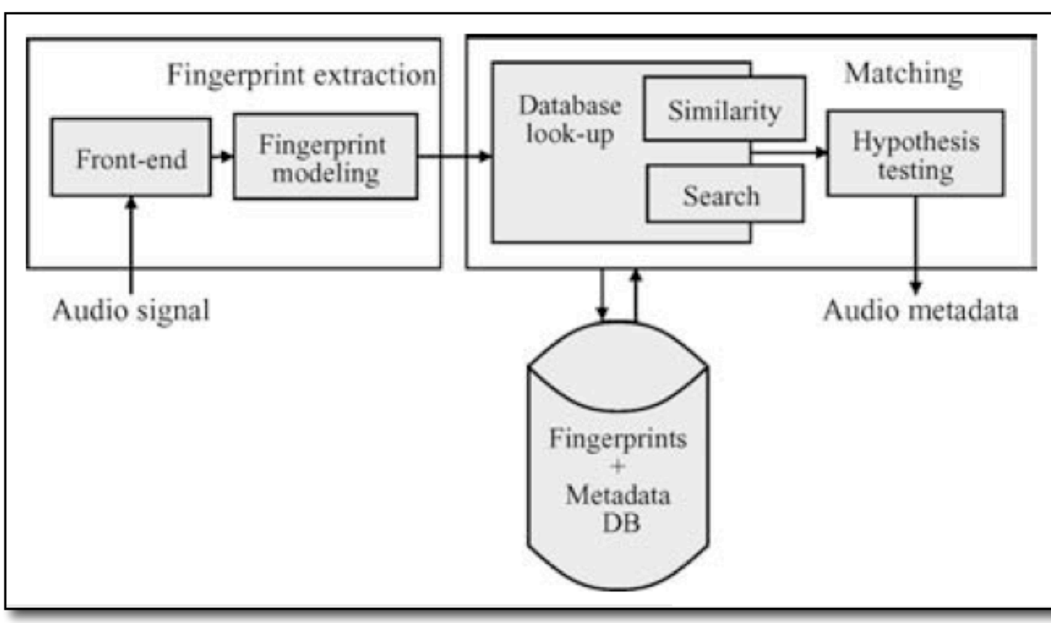

Figure 1: Content-based audio identification framework. (Cano et al. 2005)

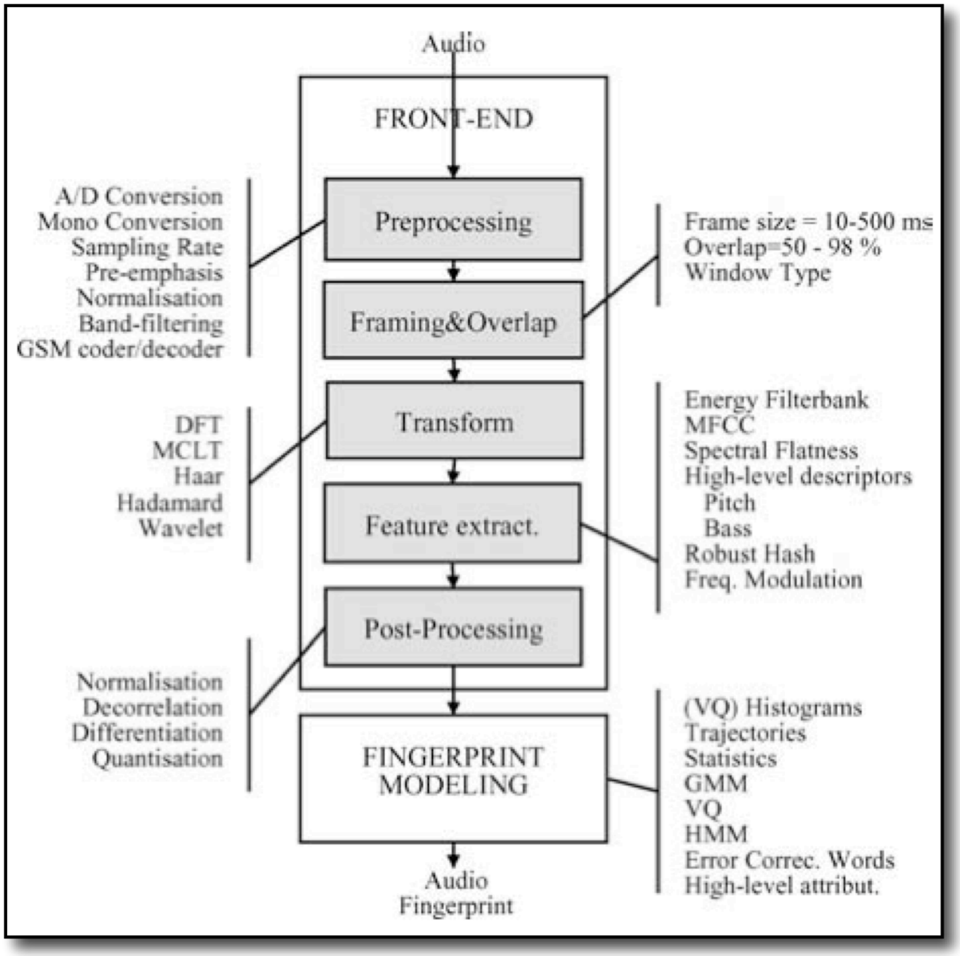

Figure 2: Fingerprint Extraction Framework. (Cano et al. 2005)

In order to identify an audio signal, its fingerprint needs to be compared to the known fingerprints contained in a database. The matching algorithm uses similarity measures to perform this task. These measures are strongly related to the fingerprint model chosen and use a correlation metric, such as the Euclidean distance, to determine the degree of similarity of two fingerprints.

Having a way to compare fingerprints, the system also needs an efficient search method to find a query in the database. For very large databases, the brute-force method is inappropriate and thus, optimization of the search algorithm is crucial. A first optimization example is to pre-calculate the distances used for the similarity measures for all entries of the database and build a data structure that reduces the number of evaluations needed during a query. Another common optimization is to use candidate pruning which consists in keeping only the best solution encountered so far when performing the search. This approach is often used with a tree-like data structure. Many more search optimizations exist and the reader is encouraged to consult the audio fingerprinting review (Cano et al. 2005) for pointers to other techniques.

In the previous steps, a search was performed using some similarity measures and the closest fingerprint found in the database was returned. However, the search result may or may not be the fingerprint requested. The last section of the system is called hypothesis testing and consists in determining whether the query is present in the repository. In order to take this decision, a threshold must be used against which the difference between the search result and the query is compared. The choice of this threshold is not trivial and influences the overall quality of the system. The threshold depends on several factors including the fingerprint model chosen, the similarity of the fingerprints contained in the database, the database size, and discriminative information of the query. In general, the larger the database is, the higher the probability of wrong match is (i.e., False Positive Rate).

Conclusion

Audio fingerprinting is a very useful technique that can be used in several contexts. Most existing fingerprinting systems can be represented using the general framework described. However, one remaining challenge is to make those systems efficient and accurate for very large collections.

References

Cano, P., E. Batlle, T. Kalker, and J. Haitsma. 2005. A review of audio fingerprinting. *The Journal of VLSI Signal Processing* 41: 271–84.

Haitsma, J., and T. Kalker. 2002. A highly robust audio fingerprinting system. *Proceedings of the International Symposium on Music Information Retrieval*. 107–15.

Kalker, T., D. Epema, P. Hartel, R. Langendijk, and M. Van Steen. 2004. Music2Share: Copyright-compliant music sharing in P2P systems. *Proceedings of the IEEE* 92 (6): 961–70.

Allamanche, E., J. Herre, O. Hellmuth, B. Froba, and M. Cremer. 2001. AudioID: Towards content-based identification of audio material. *Presented at the 110th Audio Engineering Society Convention*, Amsterdam, The Netherlands.

Doets, P., and R. Lajendijk. 2005. Extracting quality parameters for compressed audio from fingerprints. *Proceedings of the International Symposium on Music Information Retrieval*. 498–503.

Doets, P., and R. Lajendijk. 2004. Stochastic model of a robust audio fingerprinting system. *Proceedings of the International Symposium on Music Information Retrieval*. 349–52.

Burges, C., J. Platt, and S. Jana. 2003. Distortion discriminant analysis for audio fingerprinting. *IEEE Transactions on Speech and Audio Processing* 11 (3): 165–74.