

On the Use of FastMap for Audio Retrieval and Browsing

Pedro Cano, Martin Kaltenbrunner, Fabien Gouyon and Eloi Batlle

Music Technology Group
Universitat Pompeu Fabra
08002, Barcelona, Spain
34 93 542 22 02

{pcano, mkalten, fgouyon, eloi}@iua.upf.es

ABSTRACT

In this article, a heuristic version of Multidimensional Scaling (MDS) named *FastMap* is used for audio retrieval and browsing. *FastMap*, like MDS, maps objects into an Euclidean space, such that similarities are preserved. In addition of being more efficient than MDS it allows query-by-example type of query, which makes it suitable for a content-based retrieval purposes.

1. INTRODUCTION

The origin of this experiment is the research on a system for content-based audio identification. Details on the system are described in [1]. Basically the system decomposes songs into sequences of an alphabet of sounds, very much like speech can be decomposed into phonemes. Once having converted the audio into sequences of symbols, the identification problem results in finding subsequences in a superstring allowing errors, that is, approximate string matching. If we compare one sequence—corresponding to an original song in the database—to the whole database of sequences we retrieve a list of sequences sorted by similarity to the query. In the context of an identification system, this list reflects which songs the query—a distorted version of an original recording [1]—can be more easily confused with. Of course, studying this for each song is a tedious task and it is difficult to extract information on the matching results for the whole database against itself. Indeed, the resulting distances displayed in a matrix are not very informative at first sight. One possible way to explore these distances between songs by mere visual inspection is Multidimensional Scaling. MDS makes it possible to view a database of complex objects as points in an Euclidean space where the distances between points correspond approximately to the distances between objects. This plot helps to discover some structure in the data in order to study methods to accelerate the song matching search. It can also be used as a test environment to compare different audio parameterization as well as their corresponding intrinsic distances independently of the metrics. Finally, FastMap's indexing capabilities also provide an interesting tool for content-based browsing and retrieval of songs.

2. RELATED WORK

Research projects that offer visual interfaces for browsing are the *Sonic Browser* [2] and *Marsyas3D* [3]. The *Sonic Browser* uses sonic spatialization for navigating music or sound databases. In [2] melodies are represented as objects in a space. By adding direct sonification, the user can explore this space visually and aurally with a new kind of cursor function that creates an aura around the cursor. All melodies within the aura are played concurrently using spatialized sound. The authors present distances for melodic similarity but they acknowledge the difficulty to represent the melodic distances in an Euclidean space. *Marsyas3D* is a prototype audio browser and editor for large audio collections. It shares some concepts with the *Sonic Browser* and integrates them in an extended

audio editor. To solve the problem of reducing dimensionality and mapping objects into 2D or 3D spaces, Principal Component Analysis (PCA) is proposed. The drawback of this solution is that the object must be a vector of features and, consequently, it does not allow the use of e.g. the edit distance, the inclusion of metadata or other arbitrary distance metrics. In this article, the use of MDS, specifically *FastMap* is proposed to address this issue.

3. MAPPING COMPLEX OBJECTS IN EUCLIDEAN SPACES

3.1 Multidimensional Scaling

MDS [5] is used to discover the underlying (spatial) structure of a set of data from the similarity, or dissimilarity, information among them. It has been used for some years in e.g. social sciences, psychology, market research, physics. Basically the algorithm projects each object to a point in a k-dimensional space trying to minimize the *stress* function:

$$stress = \sqrt{\frac{\sum_{i,j} (\hat{d}_{ij} - d_{ij})^2}{\sum_{i,j} d_{ij}^2}}$$

where d_{ij} is the dissimilarity measure between the original objects and \hat{d}_{ij} is the Euclidean distance between the projections. The *stress* function gives the relative error that the distances in k-dimensional space suffer from, on average. The algorithm starts assigning each item to a point in the space, randomly or using some heuristics. Then, it examines each point, computes the distances from the other points and moves the point to minimize the discrepancy between the actual dissimilarities and the estimated distances in the Euclidean space. As described in [4], the MDS suffers from two drawbacks:

- It requires $O(N^2)$ time, where N is the number of items. It is therefore impractical for large datasets.
- If used in a 'query by example' search, each query item has to be mapped to a point in the k-dimensional space. MDS is not well-suited for this operation: Given that the MDS algorithm is $O(N^2)$, an incremental algorithm to search/add a new item in the database would be $O(N)$ at best.

3.2 FastMap

To overcome these drawbacks, Faloutsos and Lin [4] propose an alternative implementation of the MDS: *FastMap*. *FastMap* considers the objects as points of some unknown k-dimensional space. The points are iteratively projected to the hyperplanes perpendicular to an orthogonal set of k-lines passing through the most dissimilar objects. The algorithm is faster than MDS (being linear, as opposed to quadratic, w.r.t. the database), while it additionally allows indexing. They pursue fast searching in multimedia databases: mapping objects into points in k-dimensional spaces, they subsequently use highly fine-tuned spatial access methods (SAMs) to answer several types of queries, including the 'Query by Example' type. They aim at two benefits: efficient retrieval, in conjunction with a SAM, as discussed above, visualization and data-mining.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. © 2002 IRCAM - Centre Pompidou

4. RESULTS

To evaluate the performance of both least squares MDS and *FastMap*, we used a test bed consisting of 2 data collections. One collection consists in 1840 commercial songs and the second collection in 250 isolated instrument sounds (from IRCAM's Studio OnLine). Several dissimilarity matrices were calculated with different distance metrics. The results of these experiments are shown in detail in <http://www.iva.upf.es/mtg/SongSurfer>. In Figure 1 the representation of the song collection as points calculated with MDS and *FastMap* is shown. The MDS map takes considerably longer to calculate than the *FastMap*'s (894 vs 18.4 seconds) although several runs of *FastMap* are sometimes needed to achieve good visualizations. Although we did not objectively evaluate *FastMap* and MDS (objective evaluations of data representation techniques are discussed in [5]), MDS maps seem of higher quality. On the other hand, MDS presents a high computational cost and does not account for the indexing/retrieval capabilities of the *FastMap* approach.

5. CONCLUSIONS

We have presented the use of the existing *FastMap* method for improving a content-based audio identification system. The tool proves to be interesting, not only for audio fingerprinting research (visually exploring the representation space of audio data may reveal the possible weakness of a similarity measure), but also as a component of a search-enabled audio browser.

We first tested the tool with audio objects such as harmonic or percussive isolated sounds for which perceptually-derived distances exist. In this case the results are excellent. But songs have a more complex nature, they account for many aspects of interest. Not only good similarity measures are hard to design but also to extract automatically from low-level audio features. Song repositories are usually described with heterogeneous mixes of attributes, descriptors range from physical feature vectors (e.g. MFCCs), up to subjective labels defined by experts (e.g. the "genre").

The advantage of MDS and *FastMap* lies in their generality: they can combine any type of data attributes, from low-level attributes to metadata. We believe that this feature is relevant for improving browsing engines.

6. REFERENCES

- [1] P. Cano, E. Batlle, H. Mayer, and H. Neuschmied. Robust Sound Modeling for Song Detection in Broadcast Audio. In *Proceedings of the 112th Audio Engineering Society Convention*, Munich, 2002.
- [2] D. Ó. Maidin and M. Fernström. The Best Of Two Worlds: Retrieving and Browsing. In *Proceedings of the Conference on Digital Audio Effects*, 2000
- [3] G. Tzanetakis and P. Cook. Marsyas3D: A Prototype Audio Browser-Editor using a Large Scale Immersive Visual and Audio Display. In *Proceedings of the International Conference on Auditory Display*, 250-254. 2001.
- [4] C. Faloutsos and K. Lin. FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. In *Proceedings of the 1995 ACM SIGMOD*, 163-174. 1995.
- [5] W. Basalaj. *Proximity Visualization of Abstract Data*. Technical Report 509, University of Cambridge Computer Laboratory, January 2001.

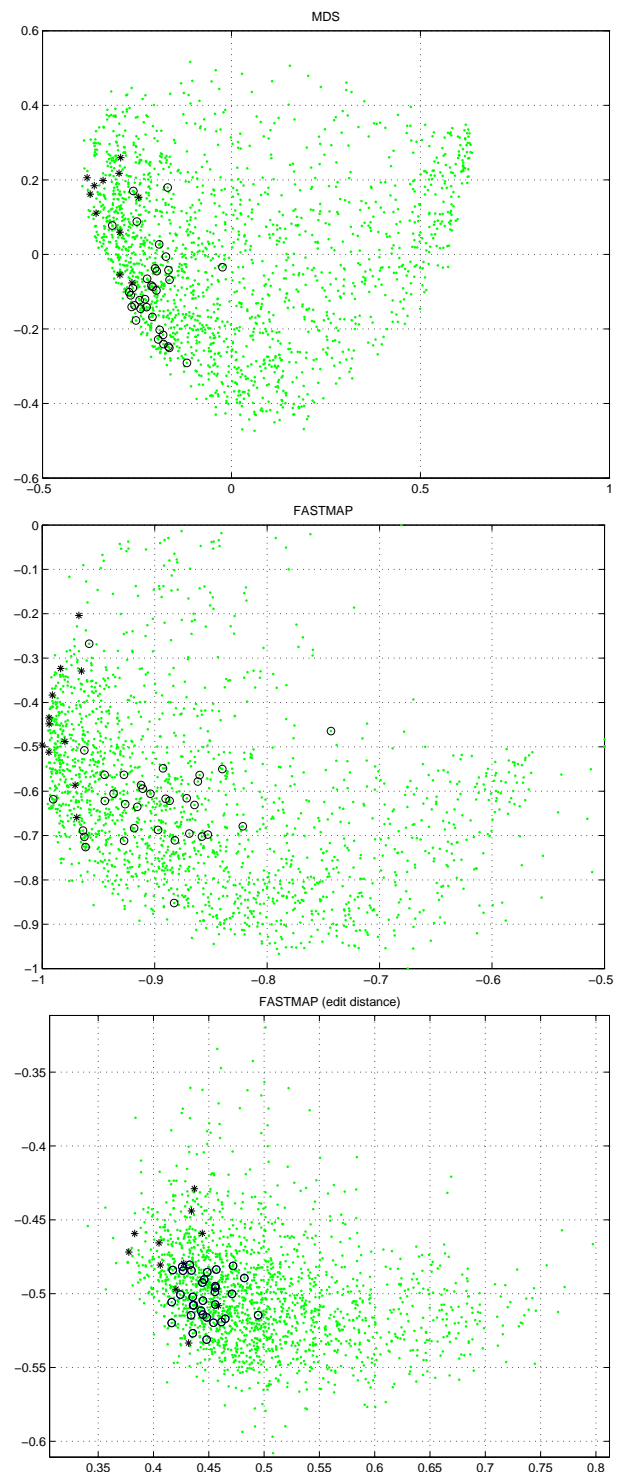


Figure 1: Representing 1840 songs as points in a 2-D space by MDS (top) and *FastMap* (middle and bottom). Asterisks and circle correspond to songs by Rex Gildo and Three Doors Down respectively. The *FastMap* plots correspond to two different similarity measures.