Setting Up an Audio Database for Music Information Retrieval Benchmarking

Perfecto Herrera-Boyer
Universitat Pompeu Fabra
Pg. Circumval·lació 8
08003 Barcelona, Spain
+34 935 422 806
perfecto.herrera@iua.upf.es

ABSTRACT
In this white paper we summarize some general requirements and issues to be clarified in order to set up a usable database for audio-based MIR benchmarking. A broad-minded approach is required in order to go beyond pure retrieval issues and include other problems that underlie in the core or in the surroundings of Music Information Retrieval. We conclude with a proposal for calling up a specific task-force that is capable of setting up the needed database in the next twelve months.

1. INTRODUCTION
In this paper we summarize some general requirements and issues to be clarified in order to set up a usable database for audio-based MIR benchmarking. Benchmark databases have become a powerful tool for other related areas to achieve the status of mature science. Besides the TREC database [9], the TIMIT [4] or SHATR [3] databases for speech recognition, and the UCI [1] database for machine learning can be cited here as examples (the interested reader may also check the NIST respository [5]). One of the main motivations for writing this position paper has been the realising that, among the Music IR test candidates listed by Byrd [2], no audio collection was present. We feel this document can be complementary to other White Papers such as those by Crawford and Brown [3], Melucci and Orio [6], and Pardo et al. [7].

2. CATEGORIES OF PROBLEMS
We need to address the following problems:

- Multi-level feature extraction (e.g. computing descriptors that can be used for solving some of the forthcoming problems)
- Segmentation (e.g. finding boundaries for homogeneous segments, that can be considered different than the others)
- Identification (e.g. telling if a given sound file corresponds to a given song "title" or if a given singer is singing in a piece of music)
- Classification (e.g. telling if a given sound file corresponds to a given genre, style or whichever enumerated-type based ontology have been defined for them)
- Structural description (e.g. telling if the music has parts that re-appear, if it has some mirroring or alternating organization of large blocks, etc.)
- Summarization (e.g. generating abridged versions of the original file that do not lose the essential music concepts that are present in the original)
- Retrieval by textual/numerical metadata (e.g. the user requests a sound from a given instrument taxonomy, or it requests files containing music with average tempos ranging from 130 to 160 b.p.m.'s)
- Retrieval by example (e.g. the user provides a file and requests to find a similar song, or the pervasive "query by humming/singing")

As you can see, even though the "retrieval" aspect is the one to consider as the most instrumental for having this meeting and discussion, there are other related problems that go "beyond retrieval" and that would be benefit from a coordinated strategy from ours. Broadening our concerns to accommodate them, and not only considering the song similarity retrieval or the query by humming/singing cases as those defining the area of music information retrieval can be a key point in opening our initiative to other scientific communities such as those of musicologists, signal processing engineers, or computer music scholars, which have shared interests with ours.

3. TYPES OF FILES
We can enumerate the following different types of files we need for a broad coverage of problems:

a) "sound samples" as isolated notes from acoustic or electronic instruments; this is the easiest type to be found; sounds from non-occidental music traditions should not be missed.

b) recordings from individual instruments ("phrases", "loops" or "solos"): phrases are used for studying the automatic computation of rhythm description features (tick, beat, tempo), of expressivity resources (vibrato, rubato, accentuation, etc.),

c) complete polyphonic music pieces covering a wide range of genres, styles and orchestrations

All recordings should be in a "bare-minimum" format of monophonic 16 bit resolution and sampled at 44.1 KHz. Stereo, and high-resolution, higher-sampling rate recordings are also welcome, but some agreement on how to deal with conversions should be matter for discussions.

A benchmarking audio database should be complemented with the corresponding MIDI files for at least the half of the included audio files in b) and c) categories.

4. FURTHER REQUIREMENTS: METADATA ANNOTATION

We need to set database-related "ground-truths" in order to be able to compare, for example, two algorithms capable of extracting beat information.

This means that:

a) in the case of having the corresponding MIDI files, part of those ground truths could be extracted from them

b) in the case of missing MIDI files, some agreement has to be required among a panel of "experts" in order to annotate the audio files

c) in any case, an agreement is needed regarding which metadata have to be included in the database

Format and availability for the metadata database is another issue to be discussed and summarizing the possible candidates would require several pages.

5. WHERE TO LOOK FOR AUDIO FILES?

The very first resource comes from the recording libraries of our own research institutions: I guess that most of us have decent recordings that would be good candidates to be included in the database, and probably could be contributed without any copyright infringement. If we pile them up on the table, it would emerge a modest interesting though heterogeneous collection to start with. There are also some collections that can be made available on specific agreements or under extremely favourable conditions, as it could be the case of the IRCAM's Studio Online Subset that was originally set up for usage in the MPEG-7 process, or the generic MPEG-7 collection (which contains some incidental music for TV).

Another source for getting files could be CD-ROMS bundled with magazines, or with some commercial software or hardware products. Future publishing magazines can be cited here as an example (look for ftp site).

One of the promising niches that is still pending of exploration and exploitation is that of commercial libraries. There are lots of music libraries companies that offer "royalty free" products. "Royalty free music" is mainly used by radio, TV, film, and video professionals. They pay a one-time license fee and the music can be used (publicly reproduced and broadcasted) repeatedly without further charge. A few royalty free public domain songs can also be found, but most royalty free music is original and has a very large variety in terms of genres, styles and instrumentation (though instrumental is quite far more frequent than vocal music). Typical license fees range from $35 to $150 or more for a 74 minute CD. Licensing terms vary, but once you purchase a license you can usually use the music any way you wish as long as you do not resell the music. As the file repository is not a commercial affair, there should be no problem with this type of collections. Anyway, it seems a better and safer strategy to specifically agree on our special conditions with the owners of the collections (wholesale prices could also be negotiated). One additional factor to be considered is that some of the largest companies offering royalty-free music could be interested in providing us with a moderate subset of their collections as an investment into their own commercial service improvement (as our work will help to devise better retrieval engines and interfaces). When possible, licensing should be made for using and distributing the sound and midi files for research activities. Provision should be made for including as research activities the following: user panel testing and reviewing and public demonstrations. Usual licensing conditions for TV music libraries are of this kind (though they usually impose the condition that the music should be "synchronized" with image and we cannot comply with that), so it would be a good idea to contact some of those companies in order to get some "tailored" CD collection to be used for our benchmarks.

6. PROPOSAL FOR A CALL FOR ACTION

It is clear for us that we need a collaborative and non-trivial effort in order to achieve a usable reference database for benchmarking. Therefore, we propose to set up a working group or task-force (12 people at most) that will take in charge the following goals to be achieved by the next MIR meeting:

1. Determine the composition of the audio files workbench
2. Determine the metadata that is needed as "ground-truth"
3. Call for contributions to the database from research institutions
4. Negotiate with commercial companies special conditions for getting a pack of their files, according to the needs established in point 1.
5. Effectively compile and organize the database, including some manual annotation (requirements and agreements on type of annotations and way of doing that should be further clarified)
6. Set up a remote repository with restricted access to its contents, and devise proper access policies (how and whom should access be granted? Only to ISMIR "declared" members?)
7. Periodically report activities to the ISMIR community

It is supposed that the financing resources for participating in the task-force will be up to each participating institution. There are research projects that can get an "added-value" if devoting a small but sufficient assignment to the task, provided a firm commitment by all the participants in the task-force. There are some research centers that . Finally, financing coming from some specific grant from European or American institutions is not out of question and should be explored and pursued.

7. SOME QUESTIONS FOR DEBATE AND CLARIFICATION

1. Are there other large scale problems that should be considered before setting up the database?
2. Do we need more/less/different audio files than those what we have outlined here?
3. Could we accept that having access to the DB could have a moderate financial cost? Where would be the cutoff point?
4. How to deal with different audio formats?

5. Is it possible to settle down to some list of metadata to be considered as ground-truth for the audio database? How?

8. REFERENCES

[1] Blake, C.L. and Merz, C.J. UCI: Repository of machine learning databases. University of California, Irvine, Dept. of Information and Computer Sciences. 1998. http://www.ics.uci.edu/~mlearn/MLRepository.html

[2] Byrd, D. Appendix A: Chart of Candidate Music IR Test Collections. Workshop on the Creation of Standardized Test Collections, Tasks, and Metrics for Music Information Retrieval (MIR) and Music Digital Library (MDL) Evaluation. Second Joint Conference on Digital Libraries. Portland, OR. 2002: pp. 49-50. http://music-ir.org/evaluation/wp1/wp1_chart.pdf

[3] Crawford, M., Brown, G. J., Cooke, M., and Green, P. Design, collection and analysis of a multi-simultaneous-speaker corpus. Proceedings of The Institute of Acoustics, Vol. 16, Part 5, The Institute of Acoustics, Windermere. 1994: pp. 183-190. http://www.dcs.shef.ac.uk/research/ilash/info/papers/ShATR/

[4] Fisher, W.M., Doddington, M., George, R., and Goudie-Marshell, K.M., The DARPA Speech Recognition Database: Specifications and Status. In Proceedings of the DARPA Speech Recognition Workshop, Report No. SAIC-86/1546, 1986. http://www.ldc.upenn.edu/readme_files/timit.readme.html

[5] NIST: Scientific and technical databases. http://www.nist.gov/srd/index.htm

[6] Melucci,M. and Orio, N. A Task-Oriented Approach for the Development of a Test Collection for Music Information Retrieval. Workshop on the Creation of Standardized Test Collections, Tasks, and Metrics for Music Information Retrieval (MIR) and Music Digital Library (MDL) Evaluation. Second Joint Conference on Digital Libraries. Portland, OR. 2002: pp. 33-35. http://music-ir.org/evaluation/wp1/wp1_melucci.pdf

[7] Pardo, B., Meek, C. and Birmingham, W. Comparing Aural Music-Information Retrieval Systems. Workshop on the Creation of Standardized Test Collections, Tasks, and Metrics for Music Information Retrieval (MIR) and Music Digital Library (MDL) Evaluation. Second Joint Conference on Digital Libraries. Portland, OR. 2002: pp. 39-41. http://music-ir.org/evaluation/wp1/wp1_pardo.pdf

[8] Reiss, J. and Sandler, M. Benchmarking Music Information Retrieval Systems Workshop on the Creation of Standardized Test Collections, Tasks, and Metrics for Music Information Retrieval (MIR) and Music Digital Library (MDL) Evaluation. Second Joint Conference on Digital Libraries. Portland, OR. 2002: pp. 43-48. http://music-ir.org/evaluation/wp1/wp1_reiss.pdf

[9] TREC data: http://trec.nist.gov/data.html

Annex 1. Some example websites which offer suitable music content, some of them with flexible licensing schemes. The list is by no means exhaustive.

http://www.sfx-gallery.co.uk/

http://www.samplenet.com

http://www.sonomic.com/

http://www.flyinghands.com

http://www.musicloops.com

http://freeplaymusic.com/

http://www.iamusic.com/

Annex 2. An example of probably suitable license:

FreePlay Music, Inc.'s musical compositions and recordings may be broadcast and/or copied for the purpose of being included in any programming or advertising being produced for transmission over any medium (including television, satellite, wireless or internet media) worldwide. FreePlay music may also be synchronized in programming for the purpose of retail videotape or DVD distribution, if such distribution follows broadcast or transmission of the synchronized FreePlay music in the programming over television, satellite, wireless or Internet media. These rights are also granted for personal, non-commercial uses. These reproduction rights (known as mechanical, synchronization and master recording rights) are absolutely FREE, and are granted in perpetuity.

FreePlay Music does, however, require a signed license if you are a manufacturer or distributor of electronic hardware, and you wish to include FreePlay music on hardware which you are selling or distributing to consumers or to other businesses. FreePlay also requires a signed license if you synchronize FreePlay music in programming which is produced solely for retail videotape or DVD distribution. Finally, FreePlay requires a signed license if you sell or license copies of FreePlay music to a third party.

FreePlay Music, Inc. reserves the right to collect license fees for public performances of these musical compositions and recordings, and authorizes ASCAP, BMI, SESAC and SoundExchange (and their international affiliates) to collect those license fees on its behalf. FreePlay Music is also willing to license public performances of its musical compositions and recordings directly, or at the source.