Recognizing Classical Works in Historical Recordings

Tim Crawford
Goldsmiths, University of London, Centre for Cognition, Computation and Culture
t.crawford@gold.ac.uk

Matthias Mauch
Queen Mary, University of London, Centre for Digital Music
matthias.mauch@elec.qmul.ac.uk

Christophe Rhodes
Goldsmiths, University of London, Department of Computing
c.rhodes@gold.ac.uk

ABSTRACT

In collections of recordings of classical music, it is normal to find multiple performances, usually by different artists, of the same pieces of music. While there may be differences in many dimensions of musical similarity, such as timbre, pitch or structural detail, the underlying musical content is essentially and recognizably the same. The degree of divergence is generally less than that found between 'cover songs' in the domain of popular music, and much less than in typical performances of jazz standards. MIR methods, based around variants of the chroma representation, can be useful in tasks such as work identification especially where disco/bibliographical metadata is absent or incomplete as well as for access, curation and management of collections. We describe some initial experiments in work-recognition on a test-collection comprising c. 2000 digital transfers of historical recordings, and show that the use of *NNLS chroma*, a new, musically-informed chroma feature, dramatically improves recognition.

1. INTRODUCTION

As was pointed out by Richard Smiraglia in a paper at ISMIR 2001, "musical works (as opposed to musical documents, such as scores or recordings of musical works) form a key entity for music information retrieval. … However, in the [general] information retrieval domain, the work, as opposed to the document, has only recently received focused attention." [1] This largely remains true today; despite a steady advance in content-based MIR techniques, we have hardly begun to realize the potential power of using them to extract higher-level musical knowledge corresponding to what is embedded in bibliographical metadata, hitherto the exclusive domain of music-librarianship. In this paper we use the term 'work' simply to refer to the musical composition as represented by the notes in a musical score (though we acknowledge that the concept is much more complex than this naïve definition assumes). The importance of the work concept in classical music becomes immediately apparent when one is confronted with the kind of confused or inaccurate metadata that often results from the use of online CD-recognition systems which rely on the ID3 tagging scheme [2] used for identifying mp3 tracks, which is rarely applied correctly to classical music. Further problems arise when a track becomes isolated from its original media (e.g. by digital copying or 'ripping' from a CD). The situation is even more problematic when works are segmented differently in different recorded manifestations: there is, for example, no standard way to divide up the continuous music of an opera into CD tracks; although there exist musicological conventions about the navigation through numbered acts and scenes, even these can break down when, for example, it is not clear from the score whether an introductory recitative forms part of an aria or forms an independent number.

In the controlled environment of the digital music library these issues can be addressed by adopting cataloguing standards such as FRBR [3], which deals comprehensively with the musical work concept and its various manifestations in physical and recorded form. The correct identification of classical works (for example, on uncatalogued archive tapes), or fragments from them (as frequently encountered on movie or advertisement sound-tracks) remains a time-consuming task demanding considerable expertise. The solution to some of these problems may lie in a system built around content-based work-recognition, operating over the internet on well-documented 'authority' collections of recorded works whose metadata can be trusted.

For much of the mainstream classical repertory, however, the work concept is fairly straightforward. The collection investigated here can be claimed to be fairly representative of the taste of classical-music record buyers in the years before the Second World War. This paper deals with the particular case of historical recordings of classical music, much of which is still in the mainstream repertory, but in which the integrity of the work may be compromised by the restrictions of the recording process itself.

Music recorded before about 1960 almost exclusively exists in the form of 78-rpm gramophone recordings. Many of these, by famous artists from Caruso to Glenn Miller, are available in modern commercial transfers,

often painstakingly enhanced using a variety of digital technologies.[2] But there remains a vast heritage of recorded music performed by less well-known artists that is unlikely ever to prompt the investment necessary to make commercial release viable. Digitization initiatives in a number of countries are increasingly making these recordings available to scholars investigating the history of recording and of musical performance as well as to the general public.[3] Music information retrieval (MIR) techniques offer rich possibilities for the curation and management of, as well as the access to, such collections. Professional metadata standards used by music librarians for cataloguing mainstream classical music, whether in the form of scores or recordings, universally make use of the work concept, and it is natural to seek ways to aid them using MIR techniques such as those described in this paper.

The task we employ as a use-case in this paper, Classical Work Recognition, is described in Section 2. Early recordings present special problems and are not suitable for many MIR methods, which are usually developed with modern commercial recordings of popular music in mind. We discuss some of these special features of early recordings and our approach to them in section 2.1.

In section 3 we discuss the chroma features we use on the historical recordings, introducing a new feature, the *NNLS chroma* (Non-Negative Least Squares chroma), which proves to offer great advantages for this task. Here we also discuss certain aspects of the search method we adopt in using the OMRAS2 specialized audio search system, audioDB.

Section 4 gives further details of our test collection and some of its special features. We describe the retrieval experiment we carried out on the test, which demonstrates clearly the advantage of a musically-informed approach, as is the case with *NNLS chroma*; this is followed in Section 5 by a discussion of the results and mentions further work we shall be doing in the near future.

2. CLASSICAL WORK RECOGNITION

We situate the research described here as a step towards automatic metadata enrichment. The long-term aim, simply put, is to develop a system which can help to identify classical works in a collection of digital audio whose descriptive metadata is either incomplete or inaccurate; the more modest task reported here is the identification of classical works that appear more than once in a collection of digitized historical recordings. Such duplicates may range from identical repetitions of the same digital file, through multiple digitizations (with or without different parameter settings) of the same 78-rpm disc, different performances by the same or different artists, to re-scorings, arrangements, extracts and medleys, examples of all of which occur in our test collection.

In order to evaluate our method's performance on this task, we have to establish 'ground-truth' in the form of a list of duplicates and 'covers' within the test collection. In principle, we should be able to process the accompanying machine-readable metadata for this, but, for various reasons, this was not possible, so this has inevitably been a largely manual process (see 4.2, below). Since almost all commercial historical recordings carry clearly-printed labels, in general it should not be hard naively to identify the works performed on a 78-rpm disc or set of discs. However, once the music has been digitized and separated from this graphical information (as was the situation for us), the problem becomes potentially more complex. In general, for example, we cannot identify tracks with works in a one-to-one correspondence, as will be discussed below.

Furthermore, classical works often – perhaps usually – comprise more than one movement. In the experiment reported here we actually treat work-*movements* as if each was a separate 'work'; we make no attempt to categorise different movements as belonging to the same work, an exercise that would presuppose a degree of musical unity which cannot be said to apply universally. In a different use-case, matching music between different movements of a work may be of great interest to musicologists, as may close matches between different works, or even works by different composers. Similarly, we ignore multiple matches of musical sequences within a single track, although this is of central importance for musical structure analysis.

If classical work-recognition could be robustly achieved with historical recordings despite their technical drawbacks (discussed below) this would offer a useful tool for metadata enrichment when used online in conjunction with a standard reference collection of recordings with high-quality metadata.

2.1 Early Recordings

Some of the special features of early recordings which can cause problems in audio analysis, and thus in audio MIR, are: limited frequency range, surface noise, distortion, variability of pitch (both global and local) and the problem of side-breaks. We briefly mention some of these in this section, though space precludes a full discussion here.

The frequency range attainable in 78-rpm recordings ranged from 168–2000 Hz in early acoustic recordings to 100–5000 Hz in electrical recordings from 1925. However, this is complicated by the various degrees of equalization that were applied to compensate for the fact that mechanical recording systems respond much more

[2] For a detailed overview of the special features of early recordings that need to be borne in mind, see [4].

[3] Useful lists of URLs for online collections of historical recordings are [5, 6]; to these [7, 8, 9] should be added.

strongly to low-frequency sounds, leading to various kinds of distortion when global gain levels are adjusted to capture the higher frequencies [9]. In this research, we take on trust the work of the professional transfer engineers who carried out the digitizations.[4]

The most immediately obvious difference between a 78-rpm recording and a modern digital one is the amount of broadband background noise known as *'surface noise'*; this has various causes, usefully summarized in [10]. There is often other noise present, usually due to mechanical aspects of the recording process. Not all of this can be completely eliminated by digital techniques, especially when it has a more-or-less definite 'pitch'. Problems due to broadband noise can mostly be avoided by using chroma features such as *NNLS chroma*, designed to ignore the non-harmonic components from percussion instruments. Distortion is a common feature in early recordings, like noise due to a variety of causes, and it is a problem that cannot easily be sidestepped. We have observed that highly-modulated loud passages in certain recordings tend to be distorted and often behave anomalously in content-based matching. This will need to be the subject of future research.

As Daniel Leech-Wilkinson demonstrates[5], the pitch of early recordings is by no means reliable; in general we can neither be sure of the global pitch-standards used by the performers (e.g. A=440Hz) nor of the actual frequencies sounding in the studio during recording. We mention some strategies for overcoming this problem in Section 3, below. The problem of side-breaks is addressed in Section 4.

3. FEATURE SELECTION & SEARCH

The classical work-recognition problem is close, though not identical, to the well-known MIR 'cover song' problem. In fact, in some respects it is somewhat simpler, since cover songs vary from their original model in ways that are generally unpredictable and can occur in several directions simultaneously. In general, we can be fairly sure that sequences of pitch-based data will be more-or-less invariant between recorded instances of the same work. This is more likely to be true where the scoring and instrumentation are the same, and both performers are working from the same (or a similar) score; where more radical re-arrangement or rearranging of the music has taken place there will be less similarity. For this reason, we match sequences of chroma features, rather than whole-track features; unless the latter embody some notion of sequence (as might be the case in an n-gram model) the number of false positives is likely to be high, since many work-movements in the same key and using the same general harmonic language are likely to share similar overall pitch-class content. Furthermore, chromas are robust to variation in timbre, which allows us to match radically-differing instrumentations (subject to limits of noise caused by percussive sounds or distortion).

In this paper we compare the performance of two chroma features in our classical work-recognition task. These are: (a) 36-bin chroma features extracted using fftExtract [11] (FE); (b) the new *NNLS chroma* features described in the following section (NNLS).

3.1 NNLS Chroma

In this section we give a brief description of the new 12-dimensional *NNLS chroma* feature which has been developed for the purpose of chord transcription [12]. But first we explain our reasons for comparing this with the performance of 36-d chromas in the same task, as this may not be immediately obvious.

One commonly-observed feature of early recordings is that they were often recorded on machines operating at different speed than the standard 78 revolutions per minute that was normal. An additional complication here is that we cannot always be sure what pitch-standard was being used by the performers; a variety of pitch standards have co-existed across the world of music in the past, some flatter, some sharper than today's accepted standard of A=440Hz. While there is little we can do to reconcile these conflicting sources of error, we can allow some tolerance in matching covers recorded at different global pitch standards by using three (or more) bins per equal-temperament semitone bin in the chroma feature, rotating the query by plus or minus a single bin at query time, and choosing the best match from these three queries. We present results using non-rotated queries and also rotated by ± one semitone below.

Although our new *NNLS chroma* features have only 12-dimensions, corresponding to the 12 chromatic pitch classes of conventional music theory, they are derived from a spectrogram with three bins per semitone, with the intention of achieving a similar invariance to small pitch deviation; the most important practical difference is that a single exhaustive search of a collection of 12-dimensional features will inevitably be more efficient than three searches of one of 36-dimensional features.

NNLS chroma features are obtained using a prior NNLS-based approximate note transcription [12, 13]. We first calculate a log-frequency spectrogram (similar to a constant-Q transform), with a resolution of a three bins per semitone. We derive the tuning of the piece in a quartertone neighbourhood of 440 Hz and adjust the log-frequency spectrogram by linear interpolation such that the centre bin (of the three) of every note corresponds to the fundamental frequency of that note in equal temperament, as is frequently done in chord- and key-estimation [e.g. 14], we adjust the chromagram to compensate for differences in the tuning pitch. First, the tuning is estimated from the relative magnitude of the three bin classes. Using this estimate, the log-frequency

[4] http://www.charm.kcl.ac.uk/history/p20_4_4_1.html
[5] [4], chapter 3.1, heading 'Misrepresentations in early recordings'

spectrogram is updated by linear interpolation to ensure that the centre bin of every note corresponds to the fundamental frequency of that note in equal temperament. The spectrogram is then updated again to attenuate broadband noise and timbre. This is done using a kind of running standardization combining the removal of the background spectrum and a form of spectral whitening.

We assume a linear generative model in which every frame Y of the log-frequency spectrogram can be expressed approximately as the linear combination $Y \approx Ex$ of note profiles in the columns of a dictionary matrix E, weighted by the activation vector x. Finding the note activation pattern x that approximates Y best in the least-squares sense subject to $x \geqq 0$ is called the non-negative least squares problem (NNLS). We choose a semitone-spaced note dictionary with exponentially declining partials, and use the NNLS algorithm proposed by Lawson and Hanson [15] to solve the problem and obtain a unique activation vector. This vector is then mapped to the twelve pitch classes C,...,B by summing the values of the corresponding pitches.

In the work reported here, our feature-vectors are all averaged into one-second frames; future work will investigate the effect on retrieval of using finer granularity. Similarly, we do not consider here the effect of low-level DSP parameters such as FFT window-length, using default values in most cases.

3.2 Search

Searching was carried out using the audioDB software developed at Goldsmiths College in the OMRAS2 project [16]. Independent audioDB databases for each feature-set were searched for best matches by Euclidean distance between queries and items in the database specified as feature-vector sequences of a given length.

4. TEST COLLECTION & EXPERIMENT

The collection of audio files we used is a subset of one provided by the King's Sound Archive (KSA) which represented the set of their digitisations completed by February 2009. The current KSA is considerably bigger, numbering over 4,500 sides with highly-reliable metadata, and free download access to most of the collection is available via a metadata-searchable web interface.[6]

4.1 King's Sound Archive (KSA)

The King's Sound Archive is based on the BBC's donation of their holdings of duplicate 78-rpm records in 2001; KSA now holds over 150,000 discs including classical and popular music as well as spoken-word and sound-effect recordings from c. 1900 to c. 1960.[7]

Our test collection comprises digitizations (undertaken in the CHARM project [17]) of 2,017 78-rpm sides, mostly classical but including some jazz, spoken-word and sound-effect recordings. This number was arrived at by chance, being the number of sound files that we could process conveniently and reconcile with the metadata provided by the KSA. We received the sound-files before the detailed discographical data now on the CHARM web-site[8] was ready; we thus had to rely on the technical production metadata, which was not primarily concerned with work identification, although it did include catalogue numbers from the disc-labels as well as the 78-rpm matrix numbers. This necessitated a lot of manual metadata editing.

4.2 Relevance judgements

In the metadata editing process we have identified a 'cover list' of around 88 works for which duplicates or multiple performances exist in the test collection. This list forms the basis for relevance judgments in our experiments. We are aware that there are often other 'relevant' tracks for a given query, but since our experiments are comparative in nature we do not regard this as a problem; in fact their effect is likely to be detrimental to our precision results.

The existence of multi-side recordings of work-movements in the collection complicates the issue of establishing reliable relevance judgments. (The various possibilities for the disposition of work-movements and side-breaks is shown diagramatically in Figure 1.) While it is often the case that the same musical material is repeated or alluded to throughout a single movement of a classical work, we cannot be sure that such repetitions are distributed evenly so that each 78-rpm side over which a movement is spread contains a roughly-equal proportion of similar musical material. Furthermore, side-breaks do not always occur at the same point in the music.

There are two basic approaches that can be taken to solve this problem, both of which present some difficulty: post-processing of results or pre-manipulation of the data. Given a list of tracks in the database (each of which corresponds to a 78-rpm side), we can post-process the search results so as to regard as mutually-relevant all matches between a query and tracks that come from anywhere in the same movement. Alternatively, we can in a preprocessing stage digitally concatenate tracks that we suspect are from the same movement. Clearly the latter procedure does not fairly represent the case where we cannot rely on our metadata, and the exact correspondence between sides and movements is unclear. In our experiment we adjusted the lists to consider as relevant only sections from similar sections of a work-movement (so 'side 1' of a given recording of a work-movement, say, is not considered relevant to 'side 2' of another recording of the same movement); while we acknowledge the limitations of this approach it does not affect our comparative evaluation.

[6] www.charm.kcl.ac.uk/sound/sound_search.html
[7] www.kcl.ac.uk/schools/humanities/depts/music/res/ksahistory.html
[8] http://www.charm.kcl.ac.uk/discography/disco.html

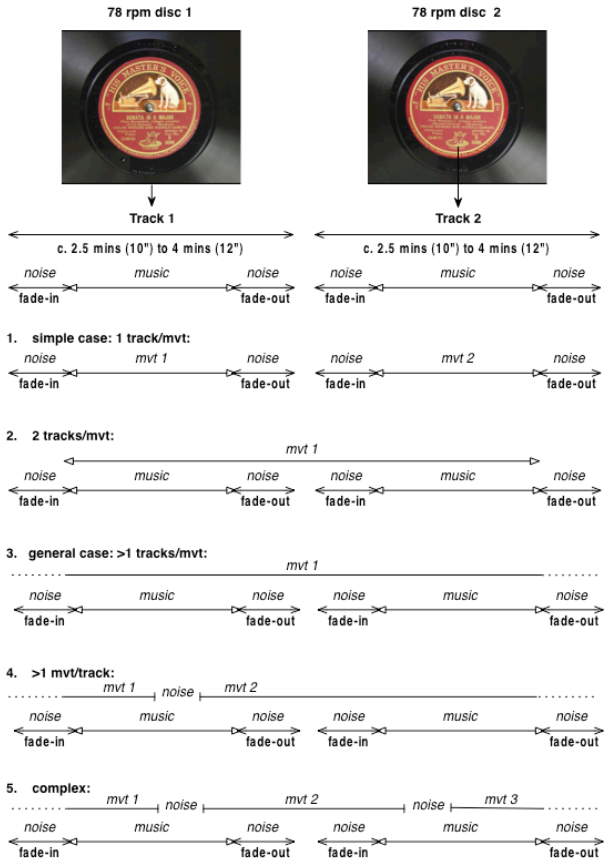

Figure 1. Disposition of work-movements on the sides of 78-rpm gramophone records.

One problem that is not solved either way is that many of our database tracks contain material from more than one movement (as in Figure 1, cases 4 and 5); furthermore, because all performances are not necessarily at the same tempo, or do not observe the same repeats, or are recorded on 78-rpm discs of different size (and consequent time-duration) the same pattern of side-breaks is generally not duplicated. This is one good reason why we use sequence-based matching, rather than using whole-track features in which material extracted from the whole of a track – even if some of it comes from a different movement in a different key – are consolidated into a single feature-vector.

4.3 Experiment

4.3.1 Method

We extracted one-second features as described in Section 3; we built an audioDB database with each set of features corresponding to the 2,017 tracks; with each pre-discovered 'cover' track in turn as query[9], we searched the database for best query/track matches of various-length sequences of feature vectors. We found that a sequence of 25 vectors consistently gave the best retrieval performance for this task across all tested features. We repeated the search with the queries rotated by up to a semitone flat and sharp (± 1 bin for NNLS; ± 1 and 2 bins for FE) taking the best result for each search.

4.3.2 Results

The 11-point recall/precision graph in Fig. 2 and the average precision values in Table 1 show a dramatic improvement (20%) in performance in this particular task brought about by the use of the *NNLS chroma* feature as opposed to the 'standard' chroma we used. While query-rotation for both features significantly improved retrieval performance, NNLS still did far better than FE. Bearing in mind the generally poor acoustic quality of the recordings in the collection, this is particularly encouraging and suggests that the new feature will be generally useful for classical work-recognition tasks on collections of higher recording quality, though as yet this remains to be tested.

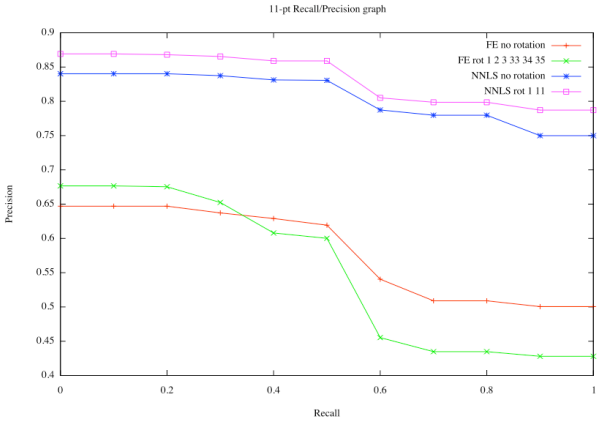

Figure 2. 11-point interpolated Recall/Precision graph for the classical work-recognition task.

	NNLS chroma	NNLS rotated	FE chroma	FE rotated
Average precision over all rel docs	0.80	**0.83**	0.57	0. 54

Table 1. Average precision (non-interpolated) for non-rotated and rotated queries over all 322 relevant tracks.

4.3.3 Discussion

The improvement in the retrieval performance of the system with the *NNLS chroma* feature (and no other changes) is striking; particularly since the feature was not designed with search or similarity judgments in mind. The model underlying it (described in section 3.1) does attempt to capture similar note content (in a way that generic chroma features attempt to capture similar acoustic pitch content) but there is potential to perform even better than our current results by tuning the NNLS features to better reflect perceptual musical similarity.

[9] The query track will, of course, be in the database at search-time; since the identity-match is always returned at the top of the ranked result-list we adjust the result-lists accordingly for our evaluation.

The fact that a query-sequence length of 25 seconds/vectors gave the best retrieval results with all features may be useful in distinguishing complete 'covers' of the kind we are dealing with here from works by classical composers which contain references to, or quotations of, other music. In general these are most likely to be short. However, this interesting topic needs to be the subject of further investigation.

5. CONCLUSIONS AND FUTURE WORK

We have demonstrated that using a chroma feature based on prior NNLS approximate transcription gives a 20% improvement in retrieval performance over conventional chroma for the work-recognition task that is the main focus of this paper. Since this copes with historical recordings of varying quality and a number of the special features of the collection (such as the arbitrary distribution of movements across 78-rpm sides), we are encouraged to hope that it will prove a particularly effective feature for general musical work recognition in other MIR contexts.

Amongst other work, then, we plan to characterize the details of the *NNLS chroma* feature in order to be able to align it better to human judgments of note-content musical similarity, as well as designing other audio features which reflect other aspects of musical sound such as timbre or rhythm.

6. ACKNOWLEDGMENTS

This work was supported by EPSRC grant EP/E02274X/1 and the NEMA (Networked Environment for Music Analysis) project funded by the Andrew S. Mellon Foundation. We are most grateful to Daniel Leech-Wilkinson, Andrew Hallifax and Martin Haskell for providing data from the King's Sound Archive. Thanks also to Ben Fields for help in various ways.

7. REFERENCES

[1] R. P. Smiraglia: "Musical works as Information Retrieval Entities: Epistemological Perspectives," Proceedings of the 2nd Annual Symposium on Music Information Retrieval (ISMIR 2001, Bloomington, Indiana, U.S.A.), 85-91

[2] Anon, "ID3", *Wikipedia*, website: http://en.wikipedia.org/wiki/ID3

[3] Anon, "Functional Requirements for Bibliographic Records", *Wikipedia*, website: http://en.wikipedia.org/wiki/Functional_Requirements_for_Bibliographic_Records

[4] D. Leech-Wilkinson: *The Changing Sound of Music: Approaches to Studying Recorded Musical Performance*, online publication, London, 2009: http://www.charm.kcl.ac.uk/studies/chapters/intro.html

[5] Thomas Edison National Historical Park, Links to Online Recordings, web-site: http://www.nps.gov/edis/photosmultimedia/links-to-online-recordings.htm

[6] Anon, 'Historic Sound Recordings Around the Web', web-site: http://www.phonozoic.net/listening.html

[7] DISMARC web-site: http://www.dismarc.eu/

[8] British Library, 'Archival Sound Recordings', web-site: http://sounds.bl.uk/

[9] Anon, "Gramophone record", *Wikipedia*, website: http://en.wikipedia.org/wiki/Gramophone_record

[10] R. Wilmut, "Reproduction of 78rpm records", website: http://home.clara.net/rfwilmut/repro78/repro.html#s/n

[11] fftExtract website: http://omras2.doc.gold.ac.uk/software/fftextract/

[12] M. Mauch and S. Dixon: "Approximate Note Transcription for the Improved Identification of Rare Chords," Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010, Utrecht, The Netherlands)

[13] M. Mauch: "Automatic Chord Transcription from Audio Using Computational Models of Musical Context", unpublished PhD dissertation (Queen Mary University of London, 2010)

[14] C. Harte and M. Sandler: "Automatic Chord Identification using a Quantised Chromagram," Proceedings of 118th Convention of the Audio Engineering Society, 2005

[15] C. L. Lawson and R. J. Hanson: *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1974

[16] audioDB website: http://omras2.doc.gold.ac.uk/software/audiodb/

[17] AHRC Centre for the History and Analysis of Recorded Music (CHARM); website: http://www.charm.kcl.ac.uk/