

Sample Identification in Hip Hop Music

Jan Van Balen¹, Joan Serrà², and Martín Haro³ *

¹ Dept of Information and Computing Sciences, Utrecht University, the Netherlands

² Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Barcelona, Spain

³ Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
j.m.h.vanbalen@uu.nl, jserra@iiiia.csic.es, martin.haro@upf.edu

Abstract. Sampling is a creative tool in composition that is widespread in popular music production and composition since the 1980's. However, the concept of sampling has for a long time been unaddressed in Music Information Retrieval. We argue that information on the origin of samples has a great musicological value and can be used to organise and disclose large music collections. In this paper we introduce the problem of automatic sample identification and present a first approach for the case of hip hop music. In particular, we modify and optimize an existing fingerprinting approach to meet the necessary requirements of a real-world sample identification task. The obtained results show the viability of such an approach, and open new avenues for research, especially with regard to inferring artist influences and detecting musical reuse.

Keywords: Digital Sampling, Sample Recognition, Musical Influence, Content-based Music Retrieval

1 Introduction

Digital sampling is a creative tool in composition and music production. It can be defined as the use of a fragment of another artist's recording in a new work. The practice of digital sampling has been ongoing for well over two decades. It has become widespread amongst mainstream artists and genres, including hip hop, electronic, dance, pop, and rock [20]. Information on the origin of samples holds valuable insights into the inspirations and musical resources of an artist. Furthermore, such information could be used to enrich music collections, e.g. for music recommendation purposes. However, in the context of music processing and retrieval, the topic of automatic sample identification has been largely unaddressed [8, 21].

In this contribution we introduce automatic sample identification as a new line of research in Music Information Retrieval (MIR), and present a first approach to detecting whether a query song samples another song inside a given

* This research was done between Jan. and Sept. 2011 at the Music Technology Group of Universitat Pompeu Fabra in Barcelona, Spain. The authors would like to thank Perfecto Herrera and Xavier Serra for their advice and support. JS acknowledges JAEDOC069/2010 from Consejo Superior de Investigaciones Científicas and 2009-SGR-1434 from Generalitat de Catalunya. MH acknowledges FP7-ICT-2011.1.5-287711.

music collection. The next section of this article will state some of the motivations for developing a sample identification system and lists the requirements that such a system should meet, in relation to existing research in content-based music retrieval. Audio fingerprinting techniques are shown to be a good basis for experiments. In the third and fourth sections, experiments are presented and discussed.

1.1 Sampling in Popular Music

The Oxford Music Dictionary defines sampling as “the process in which a sound is taken directly from a recorded medium and transposed onto a new recording” [14]. As a tool for composition, it first appeared when *musique concrète* artists such as Pierre Schaeffer started assembling tapes of found field-recordings and in musical collages. Examples of the use of previously released music recordings are James Tenney’s repurposing of Elvis Presley’s *Blue Suede Shoes* in *Collage #1* (1961), and the Terry Riley composition *Bird of Paradise* (1965), which uses the song *Shotgun* by Junior Walker and his All-Stars as direct source material [15, 16].

The phenomenon of sampling reappeared in the 1970’s when New York DJs such as Kool DJ Herc started using their vinyl players to repeat and mix parts of popular recordings, to provide a continuous stream of music for the dancing crowd. The breakthrough of sampling followed the invention of the digital sampler around 1980. It allowed producers to isolate, manipulate, and combine portions of others’ recordings to obtain entirely new sonic creations [10, 23]. The possibilities that the sampler brought to the studio have played a role in the appearance of several new genres in electronic music, including hip hop, house music (from which a large part of electronic dance music originates), jungle (a precursor of drum&bass music), dub, and trip hop [22]. The first famous sample-based single was Sugarhill Gang’s *Rapper’s Delight* (1979), containing a looped sample taken from *Good Times* by Chic (1979) [14]. A famous example of sampling in rock music is the song *Bittersweet Symphony* by The Verve (1997), which looped a pattern sampled from a 1966 instrumental string arrangement of The Rolling Stones’ *The Last Time* (1965) [14].

1.2 Motivations for Computational Research on Sampling

A first motivation to undertake the automatic identification of samples originates in the belief that the musicological study of popular music would be incomplete without the study of samples and their origins. Knowledge on the origin of samples provides a direct insight into the inspirations and musical resources of an artist, and reveals some details about his or her composition methods and production choices. At the level of popular music history, there has been an emerging interest to study how some of the particular musical properties of contemporary popular music may be traced back to the influence of the technology that has been used to produce it, such as song structures, groove (the activating quality

of music), and rhythm [5, 9, 18]⁴. To our knowledge, an extensive study of how harmony, timbre, rhythm, groove and other qualities of sampled music may have descended from past cultural activity has yet to be performed.

Samples also hold valuable information on the level of genres and communities, revealing cultural influences and dependence. Many researchers have studied the way that hip hop has often sampled 60's and 70's African-American artists [10, 18] and, more recently, Bryan and Wang [4] analysed musical influence networks in sample-based music, inferred from a unique dataset provided by the WhoSampled web project⁵. Such annotated collections exist, but they are assembled through hours of manual introduction in a collaborative effort of amateur enthusiasts. It is clear that an automated approach could both widen and deepen the body of information on sample networks, while at the same time bringing the freedom to individuals to perform their own analysis, so that they do not need to rely on the collaborative efforts of others, or the platforms that store them.

Equally interesting opportunities lie alongside recent advances in folk song [27] and version identification [25] research, where sample identification research can be applied within a larger effort to trace specific musical ideas and observe musical reuse in the recorded history of the last century.

As the amount of accessible multimedia and the size of personal collections continue to grow, sample identification from raw audio also provides a new way to bring structure to the organisation of large music databases, complementing a great amount of existing research in this direction [8, 21]. Finally, sample identification could serve legal purposes. Copyright considerations have always been an important motivation to understand sampling as a cultural phenomenon; a large part of the academic research on sampling is focused on copyright and law [20].

1.3 Requirements for a Sample Identification System

The challenges of automatic sample recognition can be directly related to the way samples have been manipulated by producers. Typical parameters controlling playback in popular hardware and software samplers (e.g. AKAI, Yamaha, Ableton, Native Instruments) include filtering parameters, playback speed, and level envelope controls (attack, sustain, decay, release). Filtering can be used by producers to maintain only the most interesting part of a sample. Playback speed may be changed to optimise the tempo (time-stretching), pitch (transposition), and mood of samples. Naturally, each of these operations complicates the automatic recognition of affected samples. In addition, samples may be as short as one second or less, and do not necessarily contain tonal information, i.e. they may consist of only percussive sounds. Finally, given that it is not unusual

⁴ Talking Heads singer David Byrne also devotes two insightful chapters of his recent book *How Music Works* (2012) to the influences of technology on the artistic process.

⁵ www.whosampled.com.

for two or more samples to appear at the same time in a mix, the sample’s energy may be low compared to that of the musical elements that obscure it. This further complicates recognition.

In the light of these observations, three important requirements for any sample recognition system should be that:

1. The system is able to identify heavily manipulated query audio in a given music collection. This includes samples that are filtered, time-stretched, transposed, very short, tonal and non-tonal, processed with audio effects, and/or appear underneath a thick layer of other musical elements.
2. The system is able to perform this task for large collections.
3. The system is able to perform the task in a reasonable amount of time.

These requirements allow us to introduce and situate the problem of sample identification in the field of content-based music retrieval. Like in other fields of information retrieval, performance in music retrieval is typically expressed in terms of precision (how good the retrieved results are) and recall (how many good results are retrieved). It should be noted that, in the context of the applications proposed above, a good recall is the more important requirement: a musicologist trying to identify source material in a composition can assess system output fairly easily. Regardless, we argue that the unseen difficulty of facing these three challenges at once is in itself an excellent motivation for the proposed research.

1.4 Content-based Music Retrieval

Research in content-based music retrieval can be characterised in terms of the *specificity* [8] and *granularity* [17] of the task. Specificity refers to the degree of similarity between query and match. Tasks with a high specificity intend to retrieve almost identical documents; low specificity tasks look for more loosely associated matches that are similar with respect to some musical properties. Granularity refers to the difference between fragment-level and document-level retrieval: audio fingerprinting is an example of a fragment-level (low granularity) task, while version detection requires a more document-level (high granularity) approach. Automatic sample recognition has mid-specificity and very low granularity (i.e. very short-time matches that are similar with respect to some musical properties). Given these characteristics, it relates to audio fingerprinting.

Audio fingerprinting systems attempt to identify unlabeled audio by matching a compact, content-based representation of it, the fingerprint, against a database of labeled fingerprints [6]. Just like fingerprinting systems, sample recognition systems should be designed to be robust to additive noise and several transformations. However, the deliberate transformations possible in sample-based music production, especially changes in pitch and tempo, suggest that the problem of sample recognition is in fact a significantly more challenging task.

Audio matching and version identification systems are typical mid-specificity problems [17]. Version identification systems assess if two complete musical recordings are different renditions of the same musical piece, usually taking

changes in key, tempo and structure into account [25]. Audio matching works on a more granular level and includes remix recognition, amongst other tasks [7, 17]. Many of these systems use chroma features [8, 21]. These descriptions of the pitch content of audio require the audio to be tonal and are generally not robust with respect to changes in tonality, as may occur with the addition of other musical layers. This may be problematic in the case of sampling. We therefore claim that sample recognition should be cast as a new problem with unique requirements, for which the right tools are still to be made.

Potential sample identification tools have been proposed by a small number of authors. A system capable of fingerprinting pitch-shifted audio is described by Fenet [13]. The system relies on an existing fingerprinting technique and, as we propose in [26], uses a logarithmic frequency representation to facilitate the search for pitch-shifted audio. It performs well on the recognition of radio broadcast audio, as intended in its design. As a result however, it does not provide robustness to time-stretching or to pitch shifts up to a semitone and more. More recently, Dittmar et. al. [11] explicitly addressed sample recognition in the context of plagiarism assessment. Two approaches are proposed: a brute-force comparison of spectrograms (again with logarithmically spaced frequencies), and a method based on non-negative matrix factorization (NMF). The techniques are claimed to be robust against time-stretching and pitch-shifting respectively, but have not been formally evaluated yet.

2 Experiments

2.1 Evaluation Methodology

We now present our first approach to the automatic identification of samples [26]. Given a query song in raw audio format, the experiments aim to retrieve a ranked list of candidate files with the sampled songs first. To narrow down the scope of experiments, only samples used in hip hop music were considered, as hip hop is the first and most widely known genre to be built on samples [10]. Regarding the origins of samples, there were no genre restrictions.

An evaluation collection was established using data from specialized internet sites⁶. The set consists of 76 query tracks and 68 sampled tracks [26]. It includes 104 sample relations (expert-confirmed cases of sampling). Additionally, 320 ‘noise’ files, very similar to the candidates in genre and length, were added to challenge the system as suggested in [24]. This makes a total of 388 candidates. All examples are real-world cases of sampling. Aiming at representativeness, the ground truth was chosen to include both short and long samples, tonal and percussive samples, and isolated samples (the only layer in the mix) as well as background samples. So-called ‘interpolations’, i.e. samples that have been re-recorded in the studio, were not included, nor were non-musical samples (e.g. film dialogue).

⁶ WhoSampled (www.whosampled.com), accessed 02/2011 and Hip Hop is Read (www.hiphopisread.com, accessed 02/2011)

Figure 1 shows a visualisation of sample relations between the artists appearing in the music collection established for the evaluation of our experiments. The orange nodes represent sampled artists, the blue nodes represent the artists that sampled them. The diagram shows how the emerging links between artists quickly give rise to a complex network of influence relations.

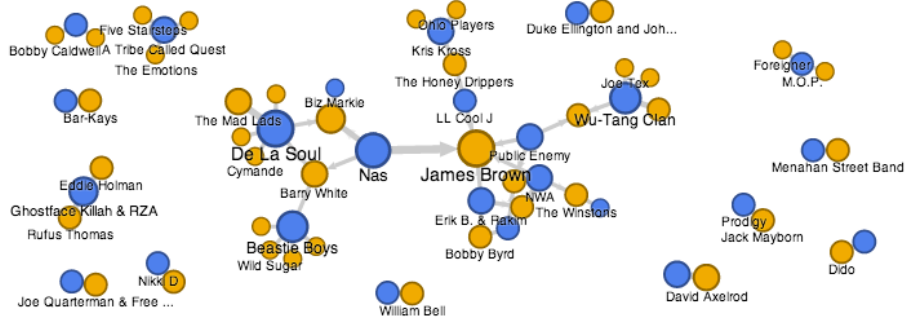


Fig. 1. Network visualisation of the connections between some of the artists in the music collection established for the evaluation of experiments. The light nodes (orange) represent sampled artists, the dark nodes (blue) represent the artists that sampled them.

Table 1. Example: two audio tracks and a sample relation as they are represented in the ground truth dataset. The sample relation S019 is identified by the associated query Q, the candidate C, and the times TC and TQ at which the sample occurs the first time. N counts the number of times the sample occurs in Q.

	Artist	Title	Year	Genre
T034	Pete Rock & C.L. Smooth	Straighten it Out	1992	Hip-hop
T035	Ernie Hines	Our Generation	1972	R&B/Soul

	C	Q	TC	TQ	N	Comments
S019	T035	T034	0:40	0:10	48	Vocals

The mean average precision (MAP) was chosen as the experiment’s evaluation metric [19]. The MAP is a common measure in information retrieval and has been used in several related tasks [12]. With the collection used in this study, a random baseline of 0.017 was found over 100 iterations, with a standard deviation of 0.007.

2.2 Optimisation of a State-of-the-Art Audio Fingerprinting System

In a first experiment, a state-of-the-art fingerprinting system was optimised to perform our task. We chose to work with the spectral peak-based audio fingerprinting system designed by Wang [28]. A fingerprinting system was chosen because of the chroma argument in Section 1.4. The landmark-based system was chosen because of its robustness to noise and distortions and the alleged ‘transparency’ of the spectral peak-based representation: Wang reports that, even with a large database, the system is able to correctly identify each of several tracks mixed together. The same system was used by Fenet et al. [13] and has previously also been adapted for use in version recognition [1]. Table 2 lists some of the system’s strengths and weaknesses with respect to the current task.

Table 2. Strengths and weaknesses of spectral peak-based fingerprints in the context of sample identification.

Strengths	Weaknesses
<ul style="list-style-type: none"> – High proven robustness to noise and distortions. – Ability to identify music from only a very short audio segment. – ‘Transparent’ fingerprints: ability to identify multiple fragments played at once. – Does not explicitly require tonal content. 	<ul style="list-style-type: none"> – Not designed for transposed or time-stretched audio. – Designed to identify tonal content in a noisy context, fingerprinting drum samples requires the opposite. – Percussive recordings may not be representable using spectral peak locations only.

As in most other fingerprinting systems, the landmark-based system consists of an extraction and a matching component [28]. Briefly summarized, the extraction component takes the short time Fourier transform (STFT) of audio segments and selects from the obtained spectrogram a uniform constellation of prominent spectral peaks. The time-frequency tuples with peak locations are paired in 4-dimensional ‘landmarks’, which are then indexed as a start time stored under a certain hash code for efficient lookup by the matching component. The matching component retrieves for all candidate files the landmarks that are identical to those extracted from the query. Query and candidate audio segments match if these corresponding landmarks show consistent start times [28].

An implementation of this algorithm was made available by Ellis⁷. It works by the same principles as [28], and features a range of parameters to control the implementation-level operation of the system. Important STFT parameters are the audio sample rate and the FFT size, determining the frequency and time resolution of the spectral analysis. Another important quantity is the number of

⁷ <http://labrosa.ee.columbia.edu/matlab/fingerprint/>

spectral peaks to consider. It is governed by the Peak Density parameter (controlling the density of peaks in the time domain) and a Peak Spacing parameter (in the frequency domain). The number of resulting landmarks is governed by three more parameters: the peak pairing horizons in the frequency and time domain, and the maximum number of formed pairs per spectral peak.

A wrapper was written to slice the query audio into short fixed-length chunks, overlapping with a hop size of 1s and a length around the expected length of the longest samples. The same wrapper then feeds these chunks to the fingerprinting system as implemented by Ellis, and uses a distance function to sort the results into a ranked list. The distance that was used is

$$d_a = \frac{1}{m+1} \quad (1)$$

a function of the number of matching landmarks m .

As a complete optimisation of the system would have been too time-consuming, we have performed a large number of tests to optimise the most influential parameters. Table 3 summarizes the optimisation process, of which more details can be found in [26]. The resulting MAP was 0.218. Interestingly, better performance was achieved for lower sample rates. The optimal density of peaks and number of pairs per peak are also significantly larger than required in a standard fingerprinting context, resulting in many more extracted landmarks per second. This requires more computation time for both extraction and matching, and requires for a higher number of extracted landmarks to be stored in the system’s memory.

The MAP of around 0.23 is low for a retrieval task but promising as a first result: it is well beyond the random baseline and the system retrieves a correct best match (top 1) for around 15 of the 76 queries. These matches include both percussive and tonal samples. However, due to the lowering of the sample rate, some frequency resolution is lost, and much of the spectral information remains unused. This may affect the scalability of the system: a sufficient number of frequency bins is needed to ensure that the landmarks allow differentiation between a high number of almost identical spectra.

2.3 Constant Q Fingerprints

As a second experiment, we did a number of tests using a constant Q transform (CQT) [2] instead of a Fourier transform (FT). We would like to consider all frequencies up to the default 8,000 Hz but make the lower frequencies more important, as they contributed more to the best performance so far. The constant Q representation, in which frequency bins are logarithmically spaced, allows us to do so. The CQT also suits the logarithmic representation of frequency in the human auditory system.

We integrated the fingerprinting system with Ellis’ implementation⁸ of Brown’s fast algorithm to compute the CQT [3]. A brief optimisation of the

⁸ See <http://www.ee.columbia.edu/~dpwe/resources/matlab/sgram/> and <http://labrosa.ee.columbia.edu/matlab/sgram/logfsgram.m>

Table 3. Some of the intermediate results in the optimisation of the audio fingerprinting system by Wang as implemented by Ellis [26]. The first row shows default settings with its resulting performance.

pairs/pk	pk density (s^{-1})	pk spacing (bins)	sample rate (Hz)	FFT size (ms)	MAP
3	10	30	8,000	64	0.114
10	10	30	8,000	64	0.117
10	36	30	8,000	64	0.118
10	36	30	4,000	64	0.193
10	36	30	2,000	64	0.176
10	36	30	2,000	128	0.228
10	36	30	2,000	256	0.144

new parameters returns an optimal MAP of 0.21 at a sample rate of 8,000 Hz, showing that similar precision as before can be obtained. With this sample rate and with an optimised resolution of 32 bins per octave, the information in the frequency domain is now restored: at 8000 Hz sample rate and 32 bins per octave, the CQT spans 224 bins, whereas a 128ms FT at 2000 Hz results in only 128. The use of the constant Q will also prove convenient for reasons explained in the next section.

2.4 Repitching Fingerprints

In a third and last experiment, a first attempt was made to deal with repitched samples. As laid out in section 1.3, artists often time-stretch and pitch-shift samples. This is typically done by changing the samples' playback speed. As a result, the samples' pitch and tempo rescale with the same factor. Algorithms for independent pitch-shifting and time-stretching without audible artifacts have only been around for less than a decade, after phase coherence and transient processing problems were overcome. Even now, repitching is still popular practice amongst producers, as inspection of the used music collection confirms.

The most straightforward, brute-force method to deal with repitching is to repitch query audio several times and perform a search for each of the copies. Alternatively, however, the extracted landmarks themselves can also be repitched through the appropriate scaling of time and frequency components. This way the extraction needs to be done only once. We have performed three tests in which both methods are combined: all query audio is resampled several times, to obtain N copies, all pitched ΔR semitones apart. For each copy of the query audio, landmarks are then extracted, duplicated and rescaled to include all possible landmarks repitched between $r = 0.5$ semitones up and down. This is feasible because of the CQT's finite resolution in time and frequency. Note that the use of the constant Q transform also provides us with a convenient advantage at this point: the difference between two peaks' logarithmic frequency is invariant to pitch-shifting. In more detail: one of the landmark's dimensions is rewritten as the logarithmic frequency difference between the peaks to exploit this property.

The invariance of this component reduces the amount of landmarks needed to cover a range of repitch values.

The results for repitching experiments are shown in Table 4. We have obtained a best performance of MAP equal to 0.390 for the experiment with $N = 9$ repitched queries, all $\Delta R = 0.5$ semitones apart. This results in a total searched pitch range of 2.5 semitones up and down. A MAP of 0.390 is rather low, yet it is in the range of some early version identification systems, or even slightly better [24]. A total of 29 out of 76 queries now retrieve a correct song as their best match, examples now including several repitched samples, both percussive and tonal.

Table 4. Results of experiments using repitching of both the query audio and its extracted landmarks to search for repitched samples.

sample rate (Hz)	bins/octave	min. freq.	N	ΔR (st)	r (st)	MAP
8000	32	32	1	-	0.0	0.228
8000	32	32	1	-	0.5	0.288
8000	32	32	5	1.0	0.5	0.334
8000	32	32	9	5.0	0.5	0.390

3 Discussion

We have introduced and detailed the first research to fully address the problem of automatic sample identification. The problem has been defined and situated in the broader context of sampling as a musical phenomenon and the requirements that a sample identification system should meet have been listed. A state-of-the-art fingerprinting system has been adapted, optimised, and modified to address the task. Many challenges have to be dealt with and not all of them have been met, but the obtained performance of $\text{MAP} = 0.39$ is promising and unmistakably better than the performance obtained without taking repitching into account. Overall, our approach is a substantial first step in the considered task.

A more detailed characterisation of the unrecognised samples is rather time-consuming but will make a very informative next step in future work. Furthermore, we suggest performing tests with a more extensively annotated dataset, in order to assess what types of samples are most challenging to identify, and perhaps a larger number of ground truth relations. This will allow us to relate performance and the established requirements more closely and lead to better results.

Advances in the presented research will eventually pave the road for reliable fingerprinting of percussive audio, sample recognition based on perceptual models, or the analysis of typical features of sampled audio. These can in turn support many of the proposed applications presented in this contribution, such as

musicological research, music understanding and recommendation and perhaps even tools for providing musicians with new and inspiring resources. All this will hopefully allow a greater understanding of sampling as an artistic phenomenon and help musicologists make sense of the popular music of today.

References

1. Bertin-Mahieux, T., Ellis, D. P. W.: Large-scale Cover Song Recognition Using Hashed Chroma Landmarks, In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 117-120 (2011)
2. Brown, J. C.: Calculation of a Constant Q Spectral Transform. *Journal of the Acoustical Society of America*, vol. 89, no. 1, 425-434 (1991)
3. Brown, J. C., Puckette, M.: An efficient algorithm for the calculation of a constant Q transform. *Journal of the Acoustical Society of America*, vol. 92, no. 5, 2696-2701 (1992)
4. Bryan, N. J., Wang, G.: Musical Influence Network Analysis and Rank of Sample-Based Music, In: 12th International Society for Music Information Retrieval Conference, pp. 329-334. (2011)
5. Butler, M. J.: *Unlocking the groove: Rhythm, meter, and musical design in electronic dance music*. Indiana University Press, Bloomington (2006)
6. Cano, P., Battle, E., Kalker, T., Haitisma, J.: A Review of Audio Fingerprinting. *Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology*, vol. 41, no. 3, 271-284 (2005)
7. Casey, M., Slaney, M.: Fast Recognition of Remixed Music Audio. In: IEEE International Conference on Acoustics Speech and Signal Processing, vol. 4, no. 12, 300-304 (2007)
8. Casey, M., R. Veltkamp, Goto, M., Leman, M., Rhodes, C., Slaney, M.: Content-Based Music Information Retrieval: Current Directions and Future Challenges. In: *Proceedings of the IEEE*, vol. 96, no. 4, 668-696 (2008)
9. Danielsen, A. (ed.), *Musical Rhythm in the Age of Digital Reproduction*. Ashgate, Farnham, Surrey (2010)
10. Demers, J.: Sampling the 1970s in Hip-Hop. *Popular Music*, vol. 22, no. 1, 41-56 (2003)
11. Dittmar, C., Hildebrand, K. F., Gaertner, D., Wings, M., Müller, F., Aichroth, P.: Audio Forensics Meets Music Information Retrieval, A Toolbox for Inspection of Music Plagiarism. 20th European Signal Processing Conference, Bucharest, Romania, 1249-1253 (2012)
12. Downie, J. S.: The Music Information Retrieval Evaluation EXchange (2005-2007): a Window into Music Information Retrieval Research. *Acoustical Science and Technology*, vol. 29, no. 4, 247-255 (2008)
13. Fenet, S., Richard, G., Grenier Y.: A Scalable Audio Fingerprint Method with Robustness to Pitch-Shifting. In: 12th International Society for Music Information Retrieval Conference, Miami, USA, 121-126 (2011)
14. Fulford-Jones, W.: Sampling. In: *Grove Music Online*. Oxford Music Online <http://www.oxfordmusiconline.com/subscriber/article/grove/music/47228> (2011)
15. Grimshaw, J.: Bird of Paradise, Tape Collage. In: *Allmusic*, <http://www.allmusic.com/composition/bird-of-paradise-tape-collage-mc0002598825> (2013)
16. Grimshaw, J.: Collage # 1 ("Blue Suede"). In: *Allmusic*, <http://www.allmusic.com/composition/collage-1-blue-suede-mc0002499059> (2013)

17. Grosche, P., Müller, M., Serrà, J. Audio Content-Based Music Retrieval. In: Multimodal Music Processing, Müller, M., Goto, M., Schedl, M. (eds.) Dagstuhl Publishing, Wadern, Germany (2012)
18. Katz, M.: Music in 1s and 0s: The Art and Politics of Digital Sampling. In: Capturing Sound: How Technology has Changed Music, University of California Press, Berkeley, 137-157 (2004)
19. Manning, C. D., Prabhakar, R., Schütze, H.: An Introduction to Information Retrieval. Cambridge University Press, Cambridge (2008).
20. McKenna, T.: Where Digital Music Technology and Law Collide - Contemporary Issues of Digital Sampling, Appropriation and Copyright Law. *Journal of Information Law and Technology*, vol. 1 (2000)
21. Müller, M., Ellis, D., Klapuri, A., Richard, G.: Signal Processing for Music Analysis. *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6 (2011)
22. Peel, I.: Dance Music. In: Grove Music Online. Oxford Music Online <http://www.oxfordmusiconline.com/subscriber/article/grove/music/47215> (2011)
23. Self, H.: Digital Sampling: A Cultural Perspective. *UCLA Ent. L. Rev.*, vol. 9, no. 2, 347-359 (2001)
24. Serrà, J., Gómez, E., Herrera P.: Audio Cover Song Identification and Similarity: Background, Approaches, Evaluation and Beyond, In: Zbigniew W. R., Alicja, A. W. (eds.) *Advances in Music Information Retrieval, Studies in Computational Intelligence*, vol. 274, 307-332. Springer, Heidelberg (2010)
25. Serra, J.: Identification of Versions of the Same Musical Composition by Processing Audio Descriptions. PhD Thesis, Universitat Pompeu Fabra, Barcelona, Spain (2011)
26. Van Balen, J.: Automatic Recognition of Samples in Musical Audio. Master thesis, Universitat Pompeu Fabra, Spain, <http://mtg.upf.edu/node/2342> (2011)
27. Wiering, F., Veltkamp, R., Garbers, J., Volk, A., Kranenburg, P.: Modelling Folk-song Melodies. *Interdisciplinary Science Reviews*, vol. 34, no. 2-3, 154-171 (2009)
28. Wang, A.: An Industrial Strength Audio Search Algorithm, In: 4th International Society for Music Information Retrieval Conference, 7-13 (2003)