A Survey on Query-by-Example based Music Information Retrieval

Nastaran Borjian
Department of Electrical and
Computer Engineering,
Tarbiat Modares University,
Tehran, Iran

ABSTRACT
search in huge musical datasets using a query provided as a fragment of desired song while there exists no extra information is a particular concern in content-based music information retrieval (MIR), defined as query-by-example (QBE). A number of QBE based MIR systems have evolved in recent years, which search a desired song without any manual of its originality, such as title, composer, singer or etc., and return a list of songs ranked in descending order according to the similarity with the given query recorded by user on TV, in gym or so on. Although, too much attention has been paid to this topic by researchers and developers in several communities, such as information retrieval, data mining or multimedia browsing engines, but it still suffers from no existing a unique definition on structure, aim, similarity, performance and also output results. This paper focuses on providing a brief overview of available QBE based MIR systems to manifest variety, opportunities and challenges in this area.

General Terms
Music information retrieval, Pattern Recognition

Keywords
Music information retrieval; Query-by-example; Multimedia browsing engines; music recommendation

1. INTRODUCTION
Nowadays, multimedia content is growing rapidly and numerous contents are being created in each second. Parallel to this enormous growth; advancements in digital storage technology make it possible to store thousands of documents in a small storage device. Thus, the necessity for more accurate browsing and retrieval of documents such as text, image, audio, and video is unavoidable.

Actually, together with remarkable advances in image retrieval systems [1, 2], the audio retrieval systems, particularly in music, have been significantly developed in the last decade [3] and many music information retrieval systems (MIR) have been proposed e.g. [4-6]. In those systems, some musical-acoustical features are primitively extracted from a music dataset and stored. In first step of retrieval, the same features are extracted from a user-provided query, and then the music dataset is searched using those features. Depending on the purpose of retrieval, output is presented in the form of either "some retrieved contents similar to the query" or "a retrieved target song".

Common input query types in MIR systems are example [7-10], singing [11, 12] and humming [11, 13]. The popularity of multimedia devices (e.g. PDAs, tablets and smart phones) has led to a variety of applications, which one of the most common requests is music-play one, anywhere the consumer asks. Thus, music search engines, known also as music discover or music recommendation ones, to create a playlist of desired songs is the associated requirement in this area. In addition to mobile devices, which support different queries, more and more people still prefer to perform searches using services given on the Web by text based keyword queries or so on.

Now, one can use information related to the song, (such as song title, composer, singer etc.) to carry out a search for a desired song through a voice recognition algorithm. Discovering based on the content of music (for instance, melody) is another approach in requirements of song search. So, to perform a search, a content detection (for example, melody extraction) is first done. But, as is observed, people often don't know or cannot remind the song's related information. The query-by-humming (QBH) and query-by-singing (QBS) are two powerful methods to find music when user can remember only a part of song to hum/sing as input query. In another position, user may listen to a song enthusiastically in a café net, gym or restaurant and say what it sounds like, then if user is capable of recording of a part of that song, query-by-example (QBE) based MIR is an efficient way to search it. Thus, the key function for query by an example system is to carry out a comparison based searching derived by a similarity measure.

The main purpose of a QBE based MIR system is to search the desired song in offering dataset using the query given by user. The basic condition for QBE system is how it defines the similarity to find the music accurately and rapidly. Different musical query by example systems are roughly classified into two separate groups according to the similarity considered by user. The similarity can be expressed as "retrieval from the same group" that is mostly done in academic area or be expected as "retrieval of the target song" followed generally in commercial systems derived on mobile or Web.

In this paper, QBE based MIR systems that support music queries and cover typical designs for mobile and Web based audio search will be reviewed briefly. In the next section, a brief survey of existing works in this area in both commerce and academia will be presented. Then, in Section III, opportunities and challenges will be discussed. Finally, Section IV will conclude the paper.

2. LITERATURE REVIEW
Basic block diagram of a query by example system to retrieve the music information is shown in Fig. 1. The query is normally a few second fragment provided by user and recorded through a multimedia device. The signal processing is preferred for extracting the time and spectral features from the excerpts of dataset and query given by user, while some of

the existing works also use musical features such as pitch and rhythm. In the following, a similarity measure algorithm, referred also to pattern matching algorithm compares query with excerpts one by one and provides a list of recommended songs ranked according to similarity. There is a variety of query by example systems that some of them are developed to use as Web or mobile service and thus apply a number of advanced hardware and software modules to fulfill the marketing requirements. In literature, some of the QBE based MIR systems implemented by researchers or industrialized by engineers will be described briefly.

To date, many music retrieval systems have used pitch representation or aimed to extract melody of music. For example, Wei-Ho Tsai et al. introduced a query-by-example system to retrieve the cover versions of songs, which searches target songs with an identical tune while might be performed in another language or even by different signers and returns the songs having main melody similar to that of the query [8]. Although melody extraction and pitch contour detection yield good retrieval performance in query-by-humming or query-by singing systems, usually do not perform well in query-by-example music retrieval systems, in particular, with impure queries. Thus, some other solutions are proposed to solve this problem. For instance, Hel´en and Virtanen parameterized the signal by GMM and HMM models and utilized those parameters to search the database using different similarity measures in an audio query-by-example system [7, 14]. As well, Shazam as a query-by-example system detects intensity peaks of the spectrogram to produce a spare feature set in a frequency range and aims to retrieve the queries with length of up to 15s when even the offered music is transmitted over a mobile phone line from a noisy environment such as a nightclub. In another work, a QBE-MIR system based on sound source separation is proposed [15]. In this system, three groups of sound signals are separated from queries based on drums, guitar and vocal, and then processed by volume balance control. Next, a re-mix stage that is equal to musical genre-shift is performed on queries to find the retrieval results from some specific genres listed as Classical, Dance, Jazz and Rock.

In [16], the usage of traditional local features in a query-by-example based music information retrieval application are explored, while it proposes to add a stage of encoding with a pre-computed codebook to get compact vectorial representations. It experiments with different encoding methods, namely the LASSO, vector quantization (VQ) and cosine similarity (CS), whose results show that concise representations outperform. In addition to common signal processing based features used for typical QBE-MIR systems, musical characteristics are also paid attention in some recent studies, for example [17], which describes some algorithms for pitch content analysis of music signals and their application to music discovery. Balaji Thoshkahna and K.R.Ramakrishnan explain a QBE system entitled as ARMINION [18] that uses Mel frequency cepstral coefficients (MFCC) to compute the timbral similarity between any two songs in the database based on the Mahalanobis distance. Usage of features based on audio coding models such as MPEG4 is another approach evaluated in [18].

Moreover, along with advance in multimedia devices, commercial QBE music browsing engines have developed rapidly in a few past decades, e.g. freeDB [19] and MusicBrainz [20] are two web based query-by-example (QBE) music retrieval systems that are available online as well. The Shazam [21] and Musiwave [22] are also two well-known query-by-example music retrieval systems which have been advanced as Application Programing Interface (API) on cell phones. Those employ audio fingerprinting methods to search a user-desired song playing on the radio or in the environment. Neuros is another query-by-example system which lets a user plug into an online service to search the target song by a 30-second clip from it [23]. The length of the query for a QBE music retrieval system usually ranges from 10 seconds (e.g. for Shazam) to 30 seconds for Neuros.

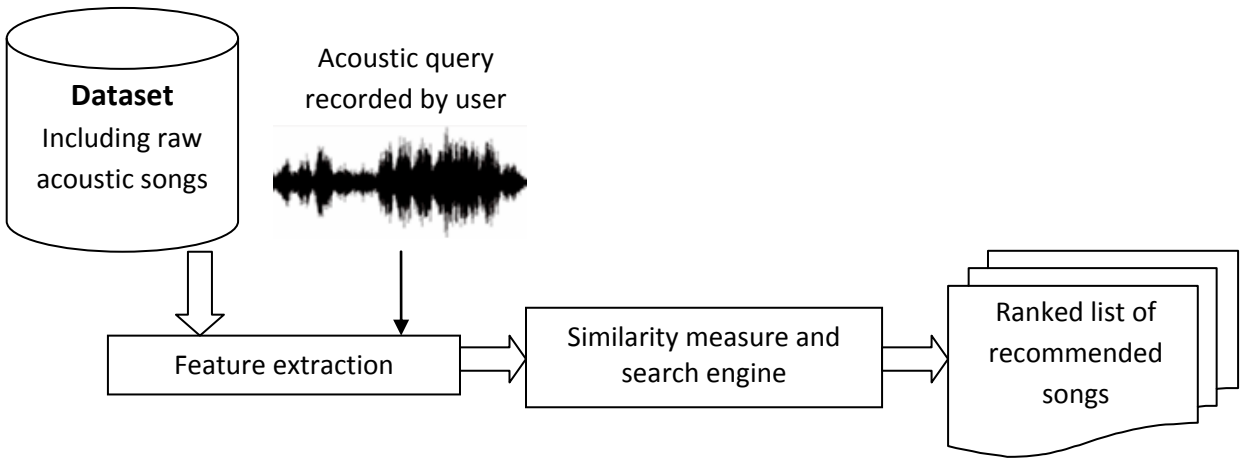

Fig 1: Block diagram of a typical query-by-example based music retrieval system

3. OPPORTUNITIES AND CHALLENGES

In recent years, music information retrieval (MIR) as an active and attractive research area has been grown rapidly. In addition, there has been too much interest to propose and develop new efficient methods in this field, originated from new accesses to the web-based and cell phone facilities as well as advances in digital storage technology. A topic of MIR enhanced by capabilities of online search to retrieve the desired songs is query-by-example (QBE) music retrieval. Many of us have experienced to listen enthusiastically to a piece of music and feel to like it very much, while we do not know anything about its title or artist. A query-by-example music retrieval system attempts to address this problem so that the user provides an example of his or her desired song

and asks the system to search it in a dataset and find the target song. Nowadays, query-by-example music retrieval systems have been a good-looking area of researches in the MIR, in particular, to develop the web-based music browsing engines and APIs. Therefore, terms of accuracy and retrieval time are two essential factors to evaluate those systems and to yield satisfactory results.

Though, some of people select an application heuristically, more and more people are interested in choosing a service that deals with the search accurately and rapidly and thus it is expected to perform a comprehensive evaluation on commercial applications. Recently, some users do limited assessments and report achieved comparisons, for example, an evaluation performed on Shazam and SoundHound [24] shows that the SoundHound has a retrieval accuracy of 95% with a delay of 21 seconds to find the target song while the Shazam has an accuracy of almost 85% with a 12-second delay for discovering time [25]. In competitive marketing research, it seems that an accuracy of 95% is adequate for a query-by-example music retrieval system, although an error rate of 5% is still high for verification [26]. The Shazam reveals the robustness against noise, while SoundHound has no claim to retrieve the query inputs recorded in noisy environments. The Shazam states that a query of 10-second duration is adequate to start the retrieval process, but users reported that the possibility for retrieval of the target song is satisfactory when a query with at least 30-second length is used for this application. However, there seems those reports are highly dependent on the number, duration and quality (noise and distortion) of queries as well as the determined device.

On the other hand, the comparison across the academic investigations in QBE based MIR area available in literature is difficult to accomplish because of variety of dataset, definition of similarity and achieved results. For instant, the audio query-by-example system described in [7, 14] defines the similarity as Top-k excerpts selected by k-nearest neighbor derived using a distance measure. Thus, accuracy in this system is considered as the number of output excerpts from the same group with query per k, such as used in [27] when groups are planned through genre concept and a similarity measure based on Kullback-Leibler divergence is performed. But similarity in [8] is according to song's melody confirmed by music specialists. Researchers in academic area mostly focus on output results and apply approaches with high computational cost to perform better, while do not report about time spent to those achievements.

Noise is extremely possible to exist in musical excerpts recorded in crowded environment such as gym, restaurant and café shop, while even a music piece played in car or on TV is likely affected from around noisy sounds if that is intended to be documented. Different distortions are another issue in QBE based MIR systems, for example, if user asks to record a song playing by an ancient player device, it can change original tempo or make some drop-off points during playing. Re-mix techniques also can shift fundamental frequencies of a song or pitches roughly 0.02%. Resolution and bit rate parameters are associated to the quality of an input query that might be differed from original song. However, a QBE based MIR system underperforms if input query includes noise or different distortions, and thus developers should notice to increase the robustness in such as systems.

Furthermore, the query and relevant songs in dataset can be recorded in different sampling frequency, and a stage for up sampling or down sampling might be necessary to be done before feature extraction in conditions that it makes no change on original tempo-pitch.

4. CONCLUSION

Retrieval of a music using recorded queries without any manual of its originality is an important and challenging issue in query-by-example (QBE) systems. In this paper, a number of QBE based music information retrieval (MIR) systems in both academic and commercial areas were introduced briefly to show variety in this field. In addition, some of the challenges and shortcomings in this field are presented, including comprehensive evaluation on commercial applications, difficulty of comparison in academic systems because of no existence of a unique definition, accuracy-time trade off and finally robustness against noise and other distortions. Obviously, any of mentioned challenges should be noticed in future works to develop QBE based MIR systems.

5. REFERENCES

[1] T. Dharani, and I. L. Aroquiaraj, "A survey on content based image retrieval," in International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME 2013) Salem, USA, 2013, pp. 485-490.

[2] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval," *IEEE Transaction on Pattern Analysis and Machine Intelligence* vol. 35, no. 12, pp. 2916-2929, 2013.

[3] J. S. Downie. "The International Society of Music Information Retrieval," http://www.ismir.net/.

[4] Z. W. Ras, and A. Wieczorkowska, *Advances in Music Information Retrieval*, 1 ed.: Springer-Verlag Berlin Heidelberg, 2010.

[5] M. Schedl, E. Gómez, and J. Urbano, "Music information retrieval: recent developments and applications," *Foundations and Trends in Information Retrieval*, vol. 8, no. 3, pp. 127-261, 2014.

[6] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney, "Content-based music information retrieval: current directions and future challenges," *Proceedings of the IEEE*, vol. 96, no. 4, pp. 668-696, 2008.

[7] M. Helén, and T. Virtanen, "Audio query by example using similarity measures between probability density functions of features," *EURASIP Journal on Audio, Speech, and Music Processing*, pp. 1-12, 2010.

[8] W.-H. Tsai, H.-M. Yu, and H.-M. Wang, "Query-by-example technique for retrieving cover versions of popular songs with similar melodies," in 6th International Conference on Music Information Retrieval, London, UK. September 11-15, 2005, pp. 183-190.

[9] I. S. H. Suyoto, A. L. Uitdenbogerd, and F. Scholer, "Effective retrieval of polyphonic audio with polyphonic symbolic queries," in MIR '07 Proceedings of the International Workshop on Multimedia Information Retrieval, 2007, pp. 105-114.

[10] J. Makhoul, F. Kubala, T. Leek, D. Liu, L. Nguyen, R. Schwartz, and A. Srivastava, "Speech and language technologies for audio indexing and retrieval," *Proceedings of the IEEE*, vol. 88, no. 8, pp. 1338-1353, 2000.

[11] W.-H. Tsai, Y.-M. Tu, and C.-H. Ma, "An fft-based fast melody comparison method for query-by-singing/humming systems," *Pattern Recognition Letters*, vol. 33, pp. 2285-2291, 2012.

[12] H.-M. Yu, W.-H. Tsai, and H.-M. Wang, "A query-by-singing system for retrieving karaoke music," *IEEE Transactions on Multimedia* vol. 10, no. 8, pp. 1626-1637, 2008.

[13] E. Unal, E. Chew, P. G. Georgiou, and S. S. Narayanan, "Challenging uncertainty in query by humming systems: a fingerprinting approach," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 2, pp. 359-371, 2008.

[14] M. Helén, and T. Virtanen, "A similarity measure for audio query by example based on perceptual coding and compression," in Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-2007), Bordeaux, France, September 10-15, 2007, pp. 1-4.

[15] K. Itoyama, M. Goto, K. Komatani, and T. Ogata, "Query-by-example music information retrieval by score informed source separation and remixing technologies," *EURASIP Journal on Advances in Signal Processing*, pp. 1-14, 2010.

[16] Y. Vaizman, B. McFee, and G. Lanckriet, "Codebook-based audio feature representation for music information retrieval," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 22, no. 10, pp. 1483-1493, 2014.

[17] J. Salamon, "Pitch analysis for active music discovery," in 33rd International Conference on Machine Learning, New York, 2016, pp. 1-3.

[18] B. Thoshkahna, and K. Ramakrishnan, "Arminion:a query by example system for audio retrieval," *Proceedings of Computer Music Modelling and Retrieval*, pp. 1-9, 2005.

[19] A. Schröder, and M. Keith. "Free database," http://www.freedb.org.

[20] R. Kaye. "the Open Music Encyclopedia," https://musicbrainz.org.

[21] C. Barton, P. Inghelbrecht, A. Wang, and D. Mukherjee. "Shazam Company," http://www.shazam.com/company.

[22] F. Chuffart. "Musiwave "; http://www.musiwave.net.

[23] J. Born. "Neuros "; www.neurostechnology.com.

[24] K. Mohajer, M. Emami, J. Hom, K. McMahon, T. Stonehocker, C. Lucanegro, K. Mohajer, A. Arbabi, and F. Shakeri. www.soundhound.com.

[25] M. Gowan. http://www.techhive.com/

[26] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*, 2 ed.: Morgan Kaufmann, 2007.

[27] H. Harb, and L. Chen, "A query by example music retrieval algorithm," in in Proceedings of the 4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '03), 2003, pp. 1-7.