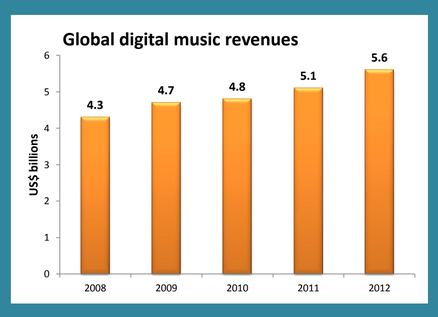
Sound Similarity: Adding a New Dimension to Music Discovery

Franz de Leon, Kirk Martinez, Robert Damper
Electronics and Computer Science

Background

"Digital downloads driving recorded music industry towards recovery" [1]



- Global recorded music revenues up 0.3%, boosted by downloads, subscription and other channels
- Digital revenues up 9%, with major music services now open in more than 100 markets
- Music is helping fuel the digital economy, but barriers to growth need to be addressed

With the growing appetite for digital music, there is a need for new applications for browsing, organising, discovering and generating playlists.

9 IN 10 MOST LIKED PEOPLE ON FACEBOOK ARE SINGERS

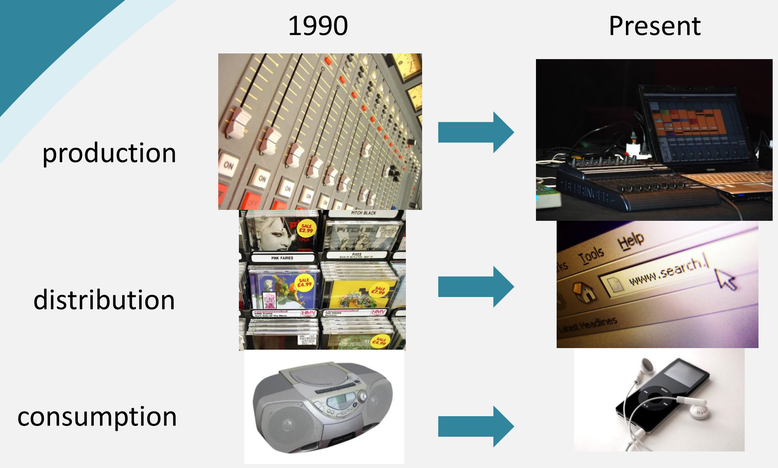
Searching similar music

- Collaborative filtering
- Social tag data
- Expert metadata
- Direct analysis of audio content

Research objectives

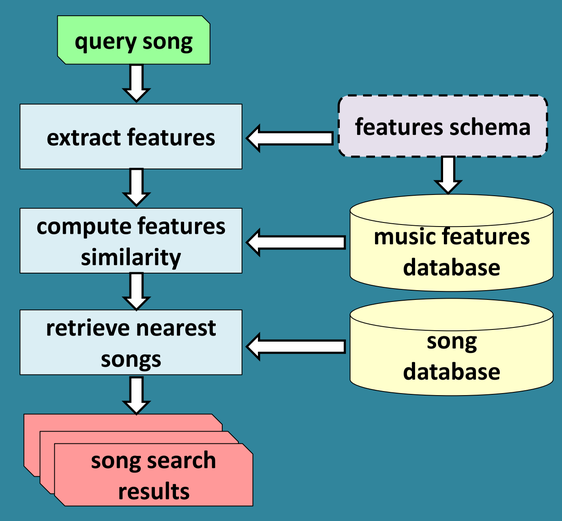
The research aims to investigate how content-based methods can be used to perform music similarity estimation. This work hopes to complement music search engines to provide an exciting journey of music discovery.

Digital Music Revolution



How do I search for music?

Music retrieval flowchart



Key issues

- What features are essential?
- Why are these features essential?
- How are these features extracted?
- How are these features summarized?
- How are the distances between features computed?
- What statistics are used to evaluate the algorithm?

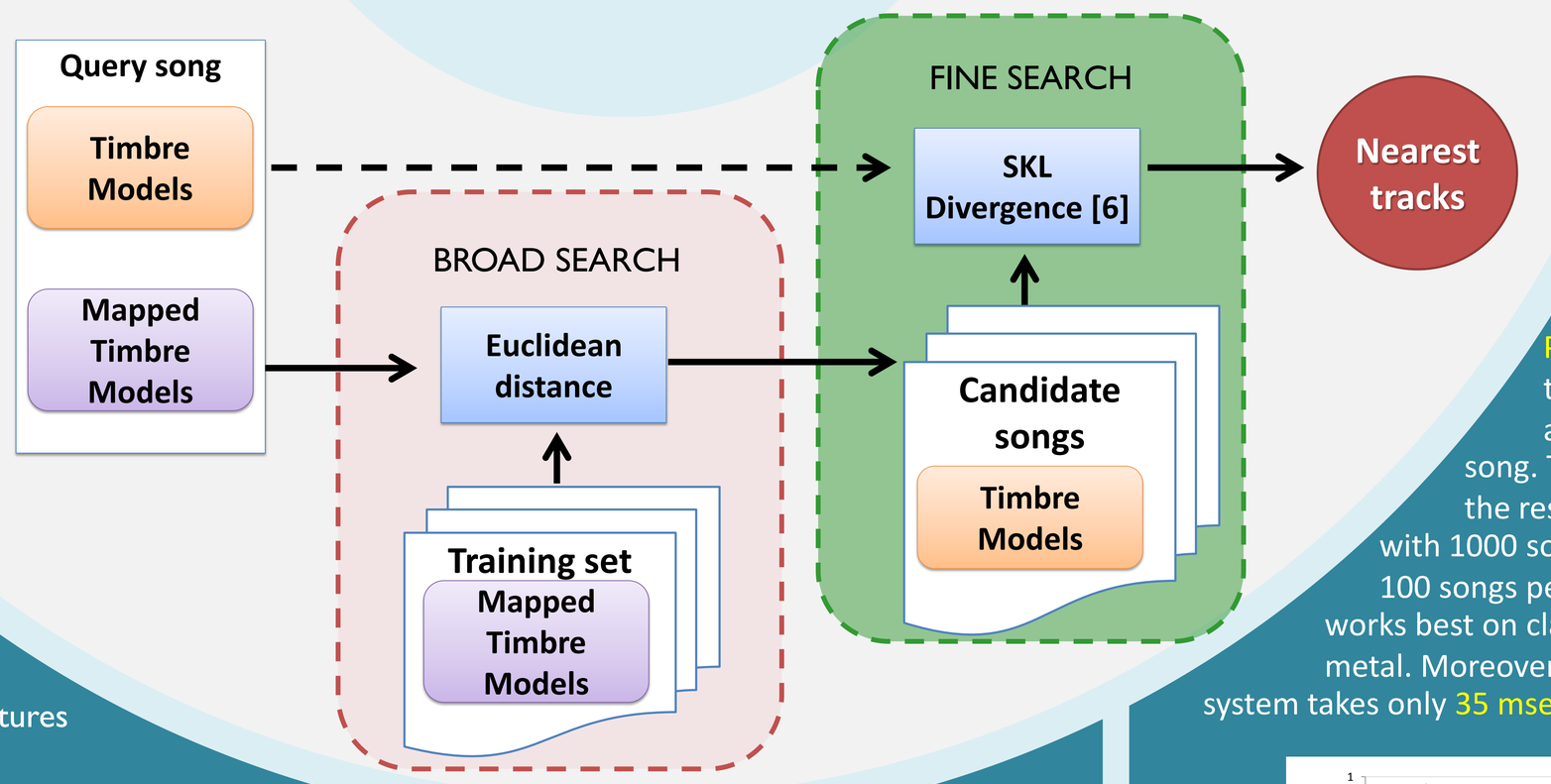
9 IN 10 OF THE MOST VIEWED VIDEOS IN YOUTUBE ARE MUSIC

Requirements

- Search database for similar tracks to a query song
- Algorithm should be fast and scalable
- System can handle wide variety of music genres

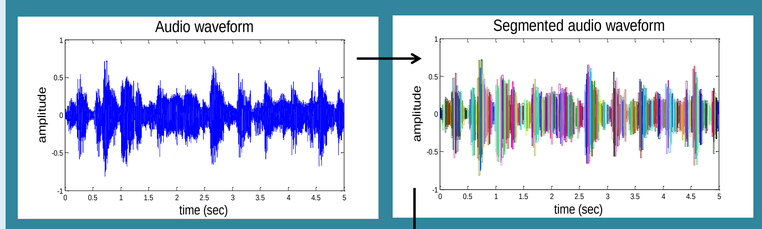
7 IN 10 MOST FOLLOWED TWITTER USERS ARE SINGERS

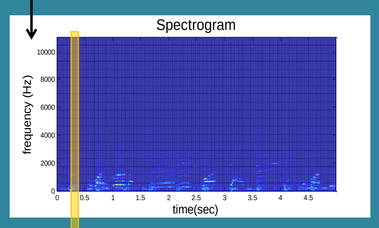
Proposed system

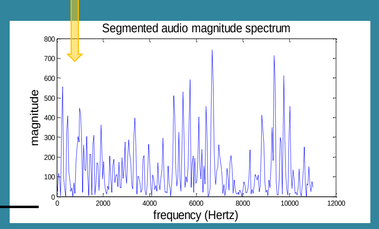


Feature extraction

Timbre is "that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are dissimilar." [2] Hence, it is crucial to develop a computational model that captures the salient features of timbre.







Extracted Features Per Segment	Description
Mel-frequency cepstral coefficients	spectral shape
Spectral contrast	difference between spectral peak and valley
Spectral flux	change in energy between successive frame
others	derivative features from the magnitude spectrum

Retrieval

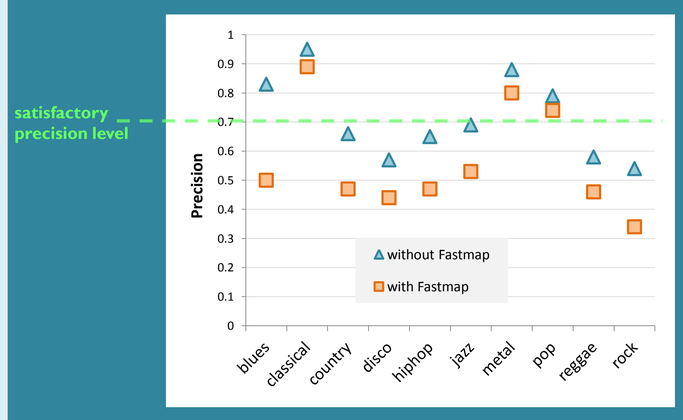
Music retrieval is done in two stages: broad search and fine search. Broad search quickly generates a list of candidate similar songs to a query song using a simple Euclidean distance between features. Fine search is performed on the candidate songs using SKL divergence [3].

Timbre Models
The extracted feature vectors per song are summarized by computing their mean and covariance matrix .

Mapped Timbre Models
The timbre features are mapped to Euclidean space using a modified Fastmap algorithm [4].

Initial results

Precision is the ratio of the songs retrieved that are similar to the query song. The plot below shows the results for a database [5] with 1000 songs from 10 genres, 100 songs per genre. The system works best on classical, pop and metal. Moreover, the proposed system takes only 35 msec. to retrieve tracks.



Acknowledgements
Franz de Leon is supported by the Engineering Research and Development for Technology Faculty Development Program of the University of the Philippines, and DOST

References
[1] Digital Music Report 2013, Feb 26, 2013 http://www.ifpi.org/content/section_resources/dmr2013.html
[2] American Standards Association 1960, 45
[3] W. Penny, "Kullback-Liebler Divergences of Normal, Gamma, Dirichlet and Wishart Densities," 2001.
[4] D. Schnitzer, A. Flexer, and G. Widmer, "A Filter-and-Refine Method for Fast Similarity Search in Millions of Tracks," in *ISMIR 2009, 2009, no. April, pp. 537-542.*
[5] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293-302, Jul. 2002.