Content-based Audio Retrieval via Hashing

Yi Yu

yuyi@ics.nara-wu.ac.jp
Graduate School of Humanity and Science
Nara Women's University

Nara Women's University

Seminar@PM3:00 2 July

What to Cover

- Research Background
- Short review -- ANN, LSH and E^2LSH
- Peer-to-peer network
- CBMR over peer-to-peer networks
- Challenges in scalable peer-to-peer environment
- Potential schemes of CBMR over P2P networks
- Current Work(Motivation,Methods,IBQBC Music Retrieval framework, Experiments and results)
- Conclusion and future work

Research Background

- A great number of multimedia contents appear on the Internet
  - These contents are shared and exchanged over P2P networks.
  - The choice of music on the major P2P networks is almost unlimited
  - Fast access to the Internet make music download (and upload) more convenient
- The actual search is often limited to the text tags (non-flexible)
- Content-based scalable music searching capabilities need to be exploited

What to Cover

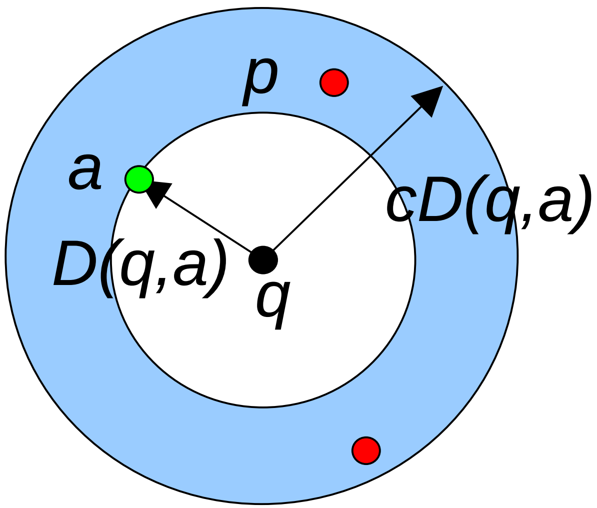
- Background and motivation
- Short review -- ANN, LSH and E^2LSH
- Peer-to-peer network
- CBMR over peer-to-peer networks
- Challenges in scalable peer-to-peer environment
- Potential schemes of CBMR over P2P networks
- Current Work(Motivation,Methods,IBQBC Music Retrieval framework, Experiments and results)
- Conclusion and future work

Approximate Nearest Neighbor(ANN)

- Given -- a set *P* of *n* points in R^d (*d* - dimension) and a slackness parameter $\varepsilon > 0$
- Goal -- with a query point *q* whose nearest neighbor in *P* is *a*, find one/all points *p* in *P*, satisfying

$$D(p,q) \le c\, D(q,a),\ c = 1 + \varepsilon$$

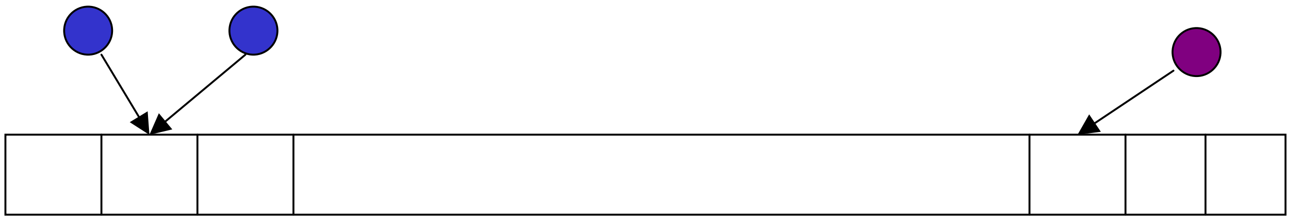
- Points in the shadowed ring are desired.



Nara Women's University

Locality-Sensitive Hashing (LSH)

- Hash function:
  - A pseudo random hash value is obtained
  - Hash value is nearly uniformly distributed.
- LSH: hash function is required to maintain the similarity. For any pair of points *p, q*,
  - Hash function *h*, generate *h(p)*, *h(q)*
  - *Pr[h(p)=h(q)]* is "high" if *p* is "close" to *q*
  - *Pr*[*h*(*p*)=*h*(*q*)] is "low" if *p* is"far" from *q*



Exact Euclidian LSH (E^2LSH)

- E^2LSH performs locality-sensitive dimension reduction by p-stable distribution
  - A distribution *D* over *R* is called *p*-stable, if

    (i) for any *n* real numbers $V = (v_1, v_2, \ldots, v_n)^T$

    (ii) i.i.d. random variables $X=(x_1, x_2, \ldots, x_n)$ and x with distribution *D*

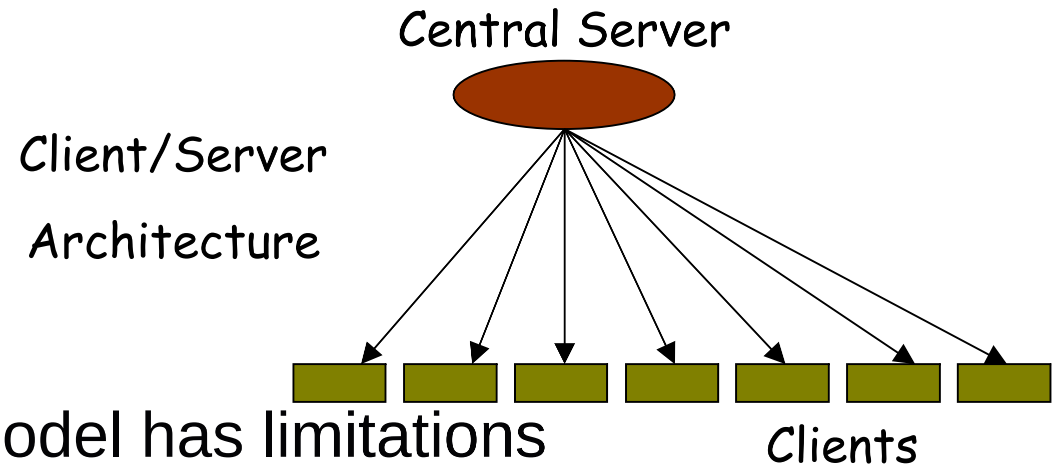
    (iii) there exists *p*, $y = \left(\sum_i |v_i|^p\right)^{1/p} x$ and $f_V(X) = \sum_{i=1}^{n} v_i x_i$ have the same distribution.
  - Dimension compression $X \rightarrow f_V(X)$

What to Cover

- Background and motivation
- Short review -- ANN, LSH and E^2LSH
- Peer-to-peer network
- CBMR over peer-to-peer networks
- Challenges in scalable peer-to-peer environment
- Potential schemes of CBMR over P2P networks
- Current Work(Motivation,Methods,IBQBC Music Retrieval framework, Experiments and results)
- Conclusion and future work

Traditional Client/Server Architecture

- A server is created to store the information that all nodes want to share
  - The server is the only data source
  - Clients request data from the server



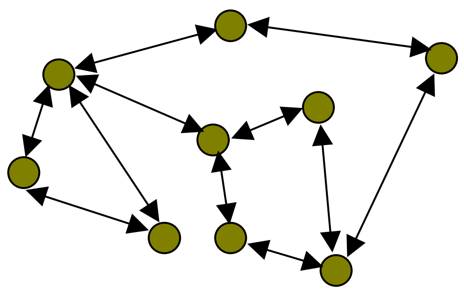
- Client/Server model has limitations
  - Hard to achieve scalability
  - Susceptible to a single point of failure (server break down)
  - Requires administration
  - Unused resource at the clients are wasted

Peer-to-Peer Concept

- Sharing of computer resources by *direct exchange* between systems (Such resource includes information, processing cycles, storage, etc.)
- Characteristics
  - Each node behaves as client, server, and router
  - Nodes are organized autonomously (there is no administrative authority)
  - Network topology is dynamic: nodes enter and leave the network frequently
  - Nodes collaborate directly with each other (not through well-known servers)
  - Nodes have widely varying capabilities

Typical P2P Architectures (1)

pure P2P



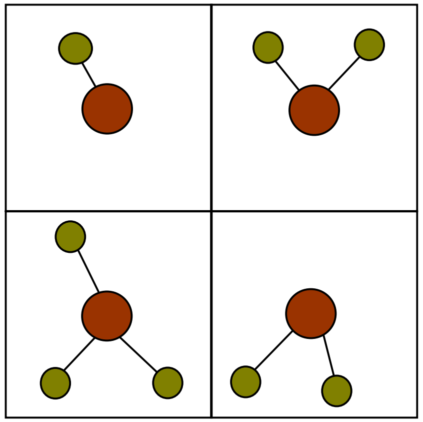
leaf node

- Pure P2P
  - Completely distributed, no central node
  - Robust—a single fault node does not affect others
  - Less efficient, the overhead may overload the network

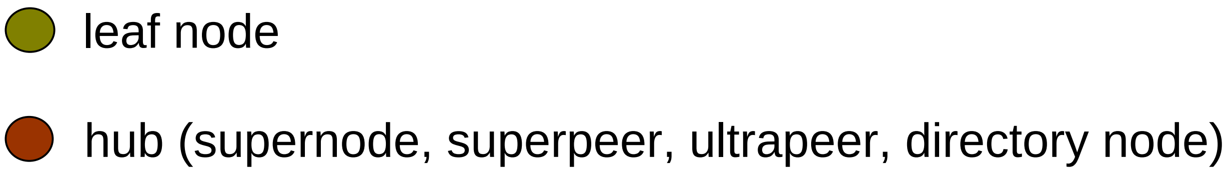
Nara Women's University

Typical P2P Architectures (2)

structured P2P



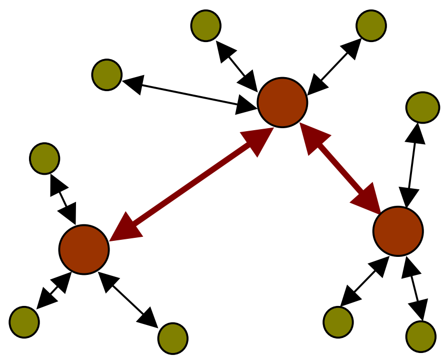
- Structured P2P
  - Most systems only support name-based retrieval
  - It is not straightforward to adopt more sophisticated retrieval models



Nara Women's University

Typical P2P Architectures (3)

hierarchical P2P



- Hierarchical P2P
  - Improves efficiency and scalability without sacrificing robustness
  - The special dedicated hubs can provide more sophisticated services to improve query routing efficiency as well as retrieval accuracy
  - It is straightforward to adopt various retrieval algorithms

leaf node

hub (supernode, superpeer, ultrapeer, directory node)

What to Cover

- Background and motivation
- Short review -- ANN, LSH and E^2LSH
- Peer-to-peer network
- CBMR over peer-to-peer networks
- Challenges in scalable peer-to-peer environment
- Potential schemes of CBMR over P2P networks
- Current Work(Motivation,Methods,IBQBC Music Retrieval framework, Experiments and results)
- Conclusion and future work

Information Retrieval System over Peer-to-Peer Networks

- What is it
  - The system performs the retrieval of documents to satisfy user's information requests in peer-to-peer networks
- What activities are involved
  - The querying node issues information requests
  - Other nodes respond to the requests with documents (document retrieval),
  - Or route requests further (query routing, resource selection)
- What architecture to use
  - Pure peer-to-peer architecture
  - Structured peer-to-peer architecture
  - Hierarchical peer-to-peer architecture
- What search mechanism to use
  - Name-based retrieval
  - Content-based retrieval

Why Content-based Retrieval

- Name-based retrieval only suffices known-item search
- Search across networks of digital libraries with more varied content requires content-based retrieval
  - Text documents usually don't have certain naming conventions and it is often difficult to describe a document in a few words
  - User usually does not know whether there are any relevant documents with respect to the information request

Nara Women's University

Why Music Retrieval over Peer-to-Peer Networks

- Existing music information retrieval models lack scalability, as a result, performance degrades when the database gets large.
- Fault-tolerance is easier to handle under a peer-to-peer architecture
- A peer-to-peer system can give access to a much larger database

What to Cover

- Background and motivation
- Short review -- ANN, LSH and E^2LSH
- Explain concept of peer-to-peer network
- CBMR over peer-to-peer networks
- Challenges in scalable peer-to-peer environment
- Potential schemes of CBMR over P2P networks
- Current Work(Motivation,Methods, IBQBC Music Retrieval framework, Experiments and results)
- Conclusion and future work

P2P Search

- Goal: Find documents with content of interest
- Search query is propagated over part of the network (from peer to a neighbor peer).
- Each search includes a query and a "*propagation rule*", which determines the search range (which neighbors the search is propagated to).
- When a peer receives a query
  - it checks if it can satisfy it
  - it decreases hop count
  - it forwards the query to a subset of its neighbors if the hop count is still greater than 0
- Overall performance of a P2P network highly depends on the efficiency and versatility of search

Challenges in Enhancing CBMR over P2P Networks

- More extensive semantics set for similarity retrieval is necessary
- Performance capability to the large database and the peer's number.
- P2P networks have many special properties, such as reliability, distributed computing and storage power,fault-tolerance, and low bandwidth.

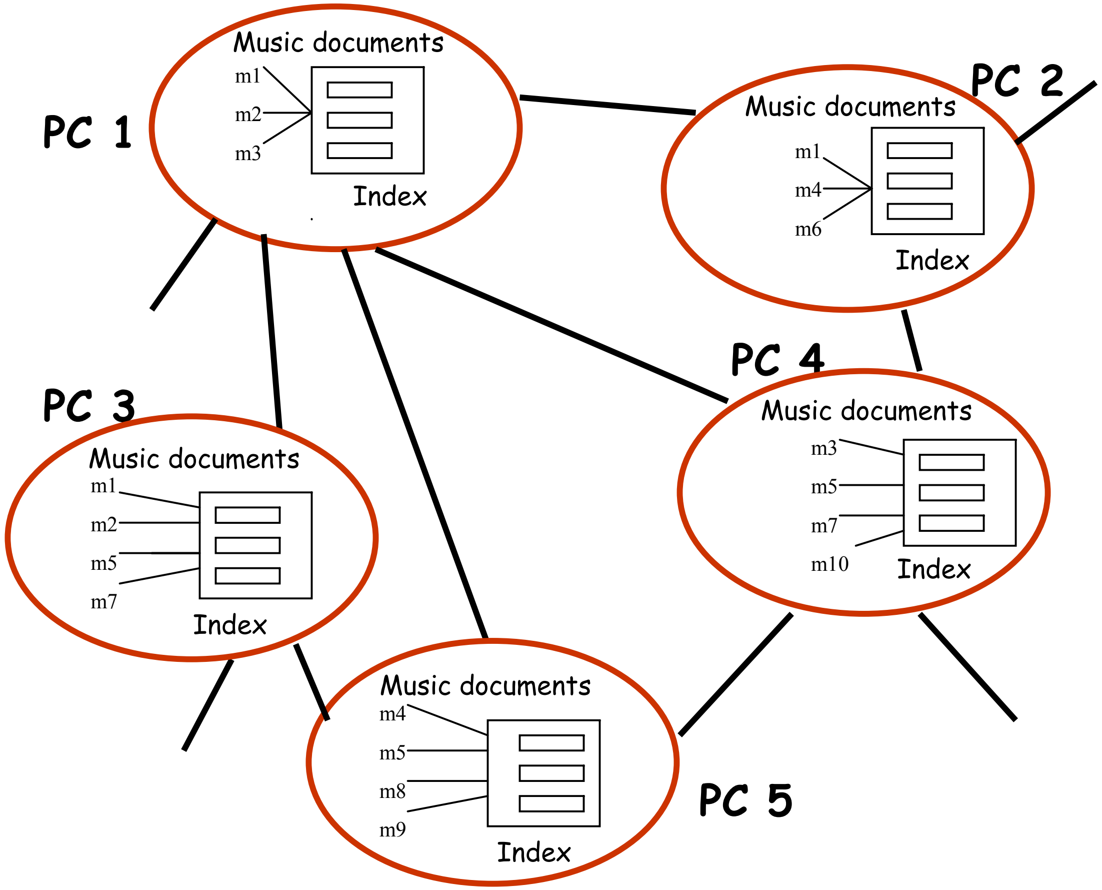
How to Improve Current Music Retrieval over P2P Networks

- We take the node or peer of a P2P system as a personal computer
- Two main aspects can be taken into account.
  - Facilitate content-based similarity retrieval by indexing audio music documents. For example, a hashing scheme--Locality Sensitive Hashing.
  - Load balance—distribute load in order to maximize throughput and minimize inconvenience to subscribers.

What to Cover

- Background and motivation
- Short review -- ANN, LSH and E^2LSH
- Peer-to-peer network
- CBMR over peer-to-peer networks
- Challenges in scalable peer-to-peer environment
- Potential schemes of CBMR over P2P networks
- Current Work(Motivation,Methods,IBQBC Music Retrieval framework, Experiments and results)
- Conclusion and future work

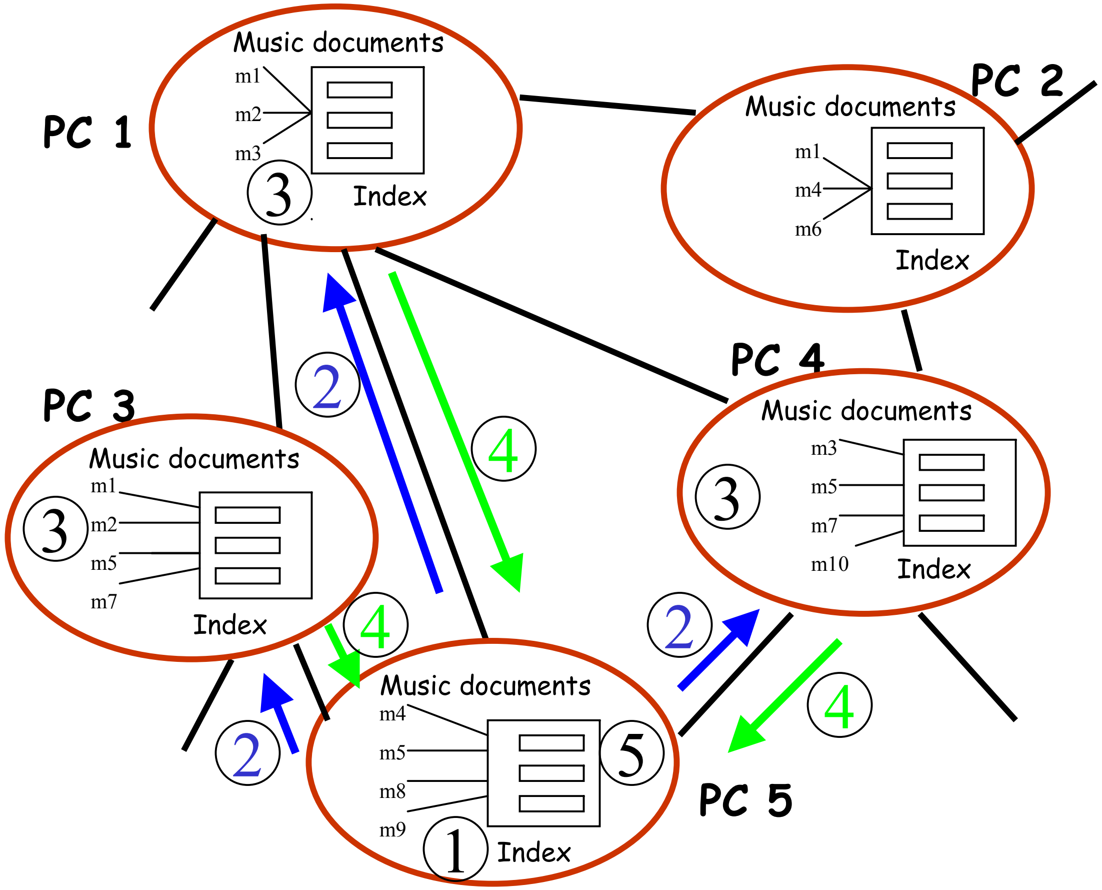
Peer-to-Peer Model for Content-based Music Retrieval



- All PCs are interconnected, each PC stores a collection of music documents.
- Analysis of acoustic data and conversion to characteristic sequences are done locally at each PC.
- While building the database, characteristic sequences for each music document are stored in multiple locally sensitive hashing instances.

Nara Women's University

Peer-to-Peer Model for Content-based Music Retrieval



Retrieval procedure

- 1 Query preparation
- 2 The query is propagated to neighbors over the network.
- 3 Each receiver processes the query and generates the matching results.
- 4 The query result is sent back to the query initiator
- 5 The query results are refined to generate the final result

Nara Women's University

What to Cover

- Background and motivation
- Short review -- ANN, LSH and E^2LSH
- Peer-to-peer network
- CBMR over peer-to-peer networks
- Challenges in scalable peer-to-peer environment
- Potential schemes of CBMR over P2P networks
- Current Work(Motivation,Methods, IBQBC Music Retrieval framework, Experiments and results)
- Conclusion and future work

Motivation

- Depend on:
  - Mapping features to integer values by heuristics
  - Reducing pairwise comparisons by hashing
- Challenges:
  - Characterize acoustic objects with relevant spectral features.
  - Represent audio features so that they can be indexed.
  - Locate desired music segments with a given query in the acceptable time.

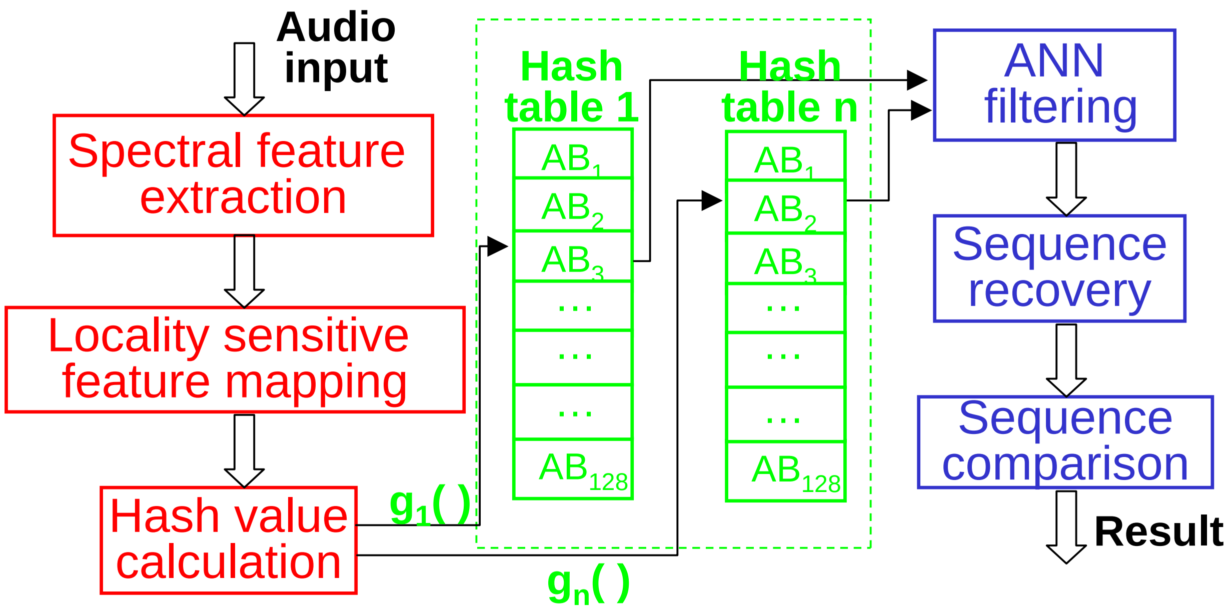
Problem Definition

- Match acoustic sequences without comparing a query to each object in the database.
  - A corpus of n musical reference pieces are represented by frames $R=\{r_{i,j} : r_{i,j} \in R_i, 1 \le i \le n, 1 \le j \le |R_i|\}$
    - $r_{i,j}$ -- j^{th} spectral feature of i^{th} reference melody in a high-dimension space
  - A query sequence $q_1, q_2, \ldots, q_Q$ filters some resemblances by E^{2}LSH/LSH-based ANN.
  - Resembled features are reorganized and compared by DP/Sparse DP.

Nara Women's University

Retrieval Framework

- Task:
  - Take a fragment of the query song as input
  - Perform a content-based similarity retrieval
  - Return melodies similar to this query fragment
- Major stages:
  - Metadata organization (*red* + *green*)
  - Querying (*red* + *blue*)



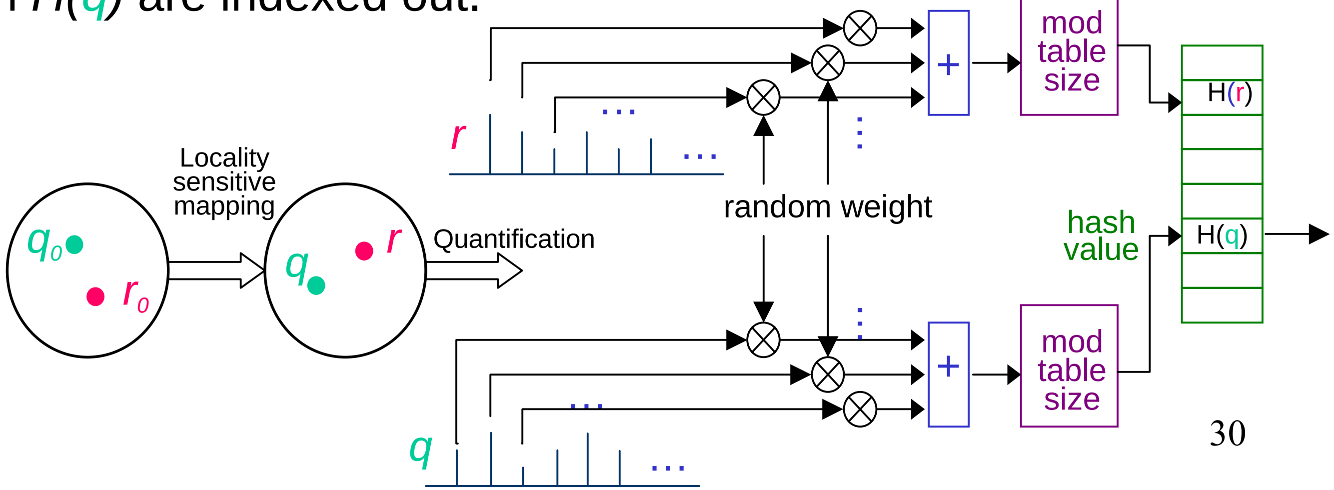
Nara Women's University

Metadata Organization

- Basic procedures:
  - Audio sequences are divided into small frames
    - STFT is calculated and used as the feature
  - Feature mapping and hash value are calculated
    - In LSH (*hash value is directly calculated from STFT*)
    - In E^2LSH (*STFT is first projected to a lower dimensional sub-feature, hash value is calculated*)
  - The features are stored in the bucket
- Results -- Convert audio features into "*indexable*" items.

Example: a Hash Instance

- Original feature (q_0, r_0), Locality sensitive mapping (q, r), Per-dimension quantification, Hash calculation [$H(r)$, $H(q)$]
- Random weight makes hash values of reference melodies almost uniformly distributed.
- If q and r have a short distance
  - They are quantified to same integer sequences
  - & generate same hash value ($H(r) = H(q)$)with a high probability.
  - Features in $H(q)$ are indexed out.



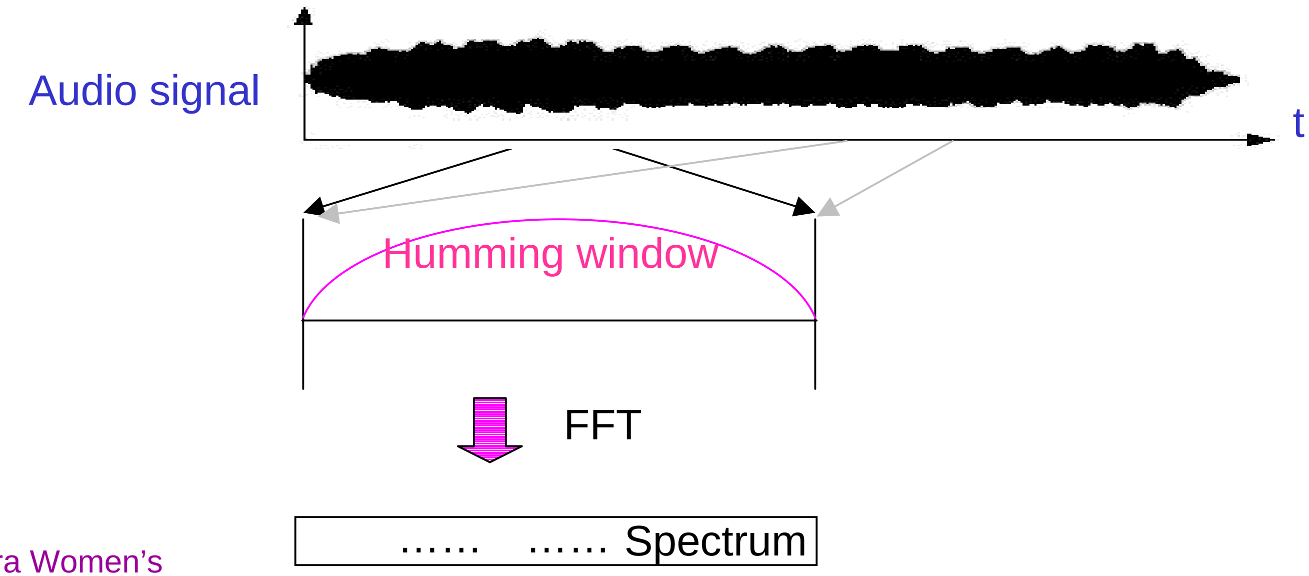
Nara Women's University

Parallel Hash Instances

- Necessary condition:
  - Each hash instance contains all the features.
  - Locality sensitive mapping generates different features & keep similarity
- Parallel lookup:
  - Construct *L* hash instances with random $g_1, g_2, \ldots g_L$
  - With a query feature *Q*, lookup buckets $g_1(Q), g_2(Q) \ldots g_L(Q)$
  - $g_1(Q)\ U\ g_2(Q)\ U \ldots U\ g_L(Q)$ gives total results

Query Stage I

- Feature extraction
  - Divide the query into overlapped frames
  - Calculate STFT for each frame

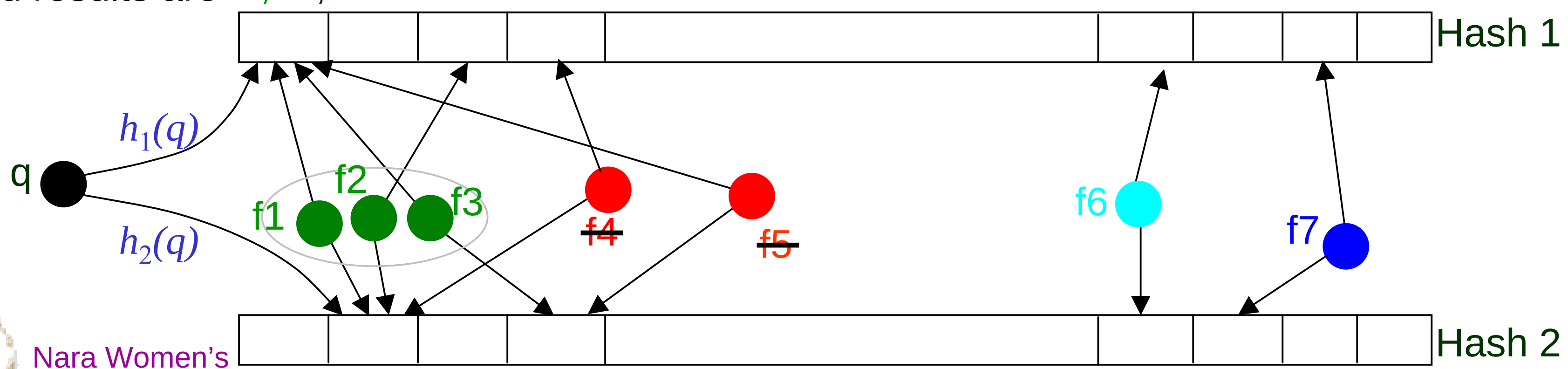


Nara Women's University

Query Stage II

- Hashing-based ANN:
  - Similar frames lie in the same bucket
  - However, dissimilar frames also exist (~~dissimilar frames~~)
  - Approximation allows a significant speedup of the calculation
- Example(Index with single feature):
  - Assume that q is similar to f1, f2, f3.
    - Lookup hash table 1, $h_1(q)$ gives query result f1, f3 and f5.
    - Lookup hash table 2, $h_2(q)$ gives query result f1, f2 and f4.
    - ~~f4~~ & ~~f5~~ are not similar to q and are removed by ANN.
    - Union of indexed results are f1, f2 and f3.

Indexed results are f1, f3, f5

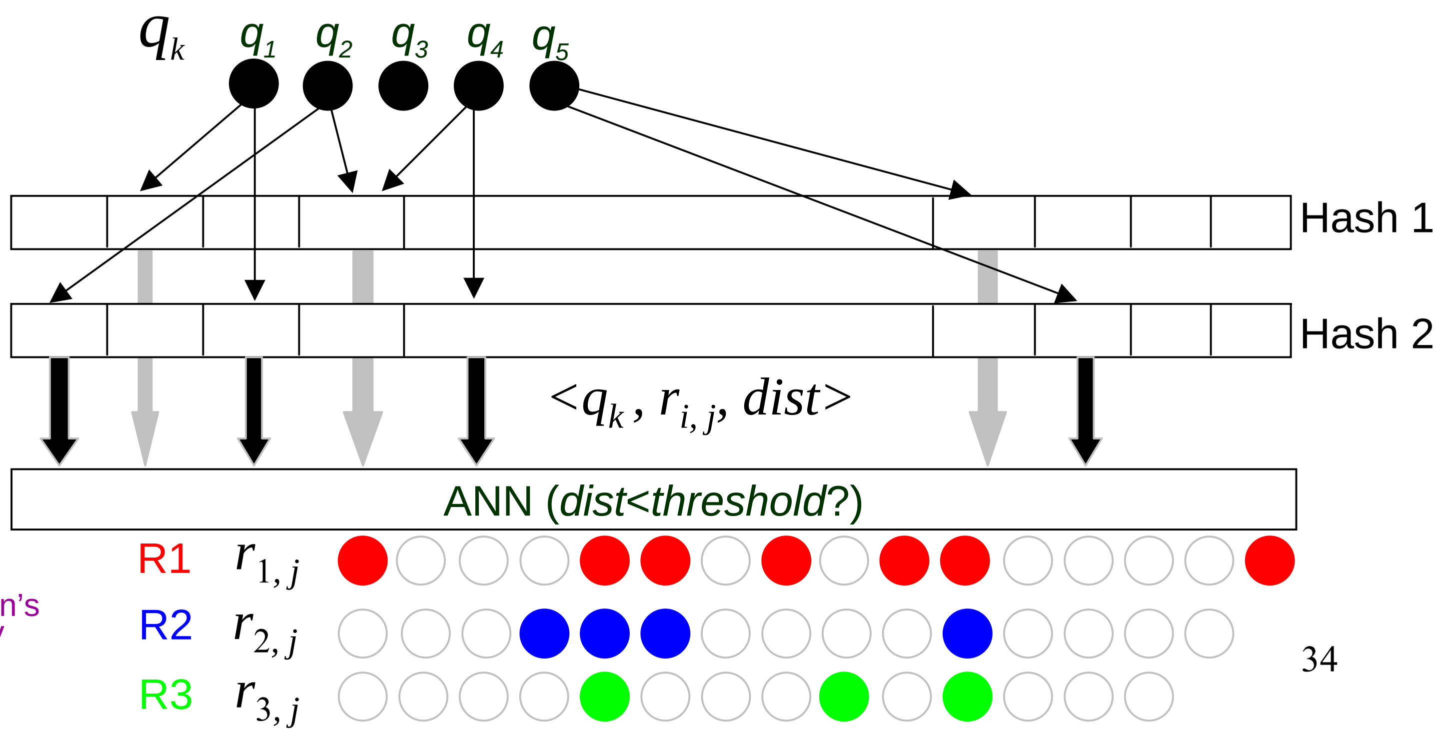


f1, f2, f4

Nara Women's University

Query Stage III

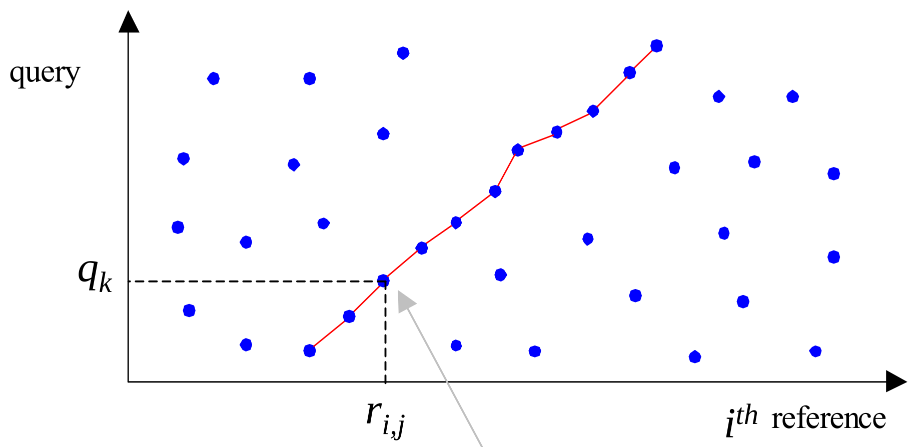
- Find desired target with a sequence of features
  - With query sequences (q_1, q_2, q_3, q_4, q_5) lookup parallel hash tables
    - Matched features belong to 3 reference melodies.
    - They are reorganized in time order.
      7 features in the 1st melody R_1, 4 features in the 2nd melody R_2,
      3 features in the 3rd melody R_3.
    - On this basis, the sequence comparison is performed



Nara Women's University

Query Stage IV

- Matched pairs are sparsely distributed over the Dynamic Time Warping (DTW) table.
  - The conventional Dynamic Programming (DP) is not efficient.
- Our sequence comparison scheme – Sparse DP (SDP)
  - Distance calculated in the filtering stage is converted into weights and filled into the DTW table
  - Melody generating the maximal weight path is the best candidate



Matched pair $<q_k, r_{i,j}, dist>$

weight=1/*dist*

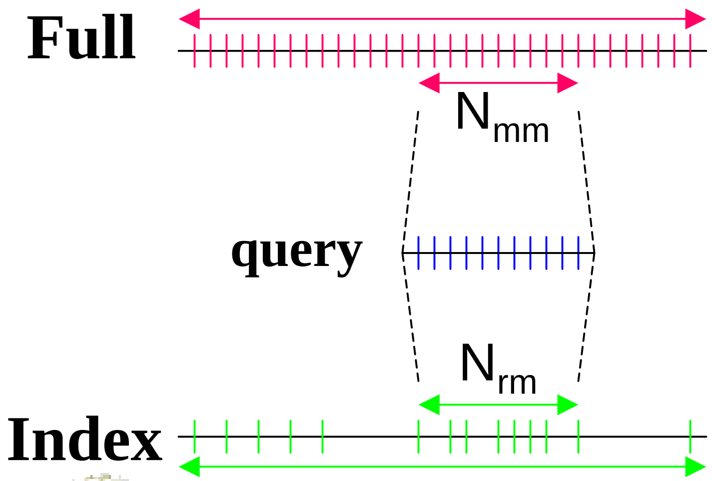
Nara Women's University

Experiment Setup

- System parameters
  - 462 reference melodies, each melody: 60s
  - A query piece: 8s
  - Sampling rate: 22.05KHz
  - Frame length: 1024, Frame overlap: 50%
  - Hash table size: 128
- Experiments goal:
  - Evaluate performance of avoiding full pairwise comparison
  - Compare LSH-DP, LSH-SDP, E^2LSH-DP, E^2LSH-SDP
- Evaluation metric:
  - Matched percentage
  - Computation time
  - Retrieval ratio

Experiments I -- Matched Percentage

- Focus on the accuracy of indexing
  - Ratio N_{rm}/N_{mm} is defined as Valid Match Percentage (VMP).
    - N_{mm}: Frames of the matched part under the conventional DP.
    - N_{rm}: Remaining frames of matched part after the filtering stage in LSH/E^{2}LSH
  - A good indexing scheme is to maximize VMP.



VMP under different filtering threshold (3 hash tables)

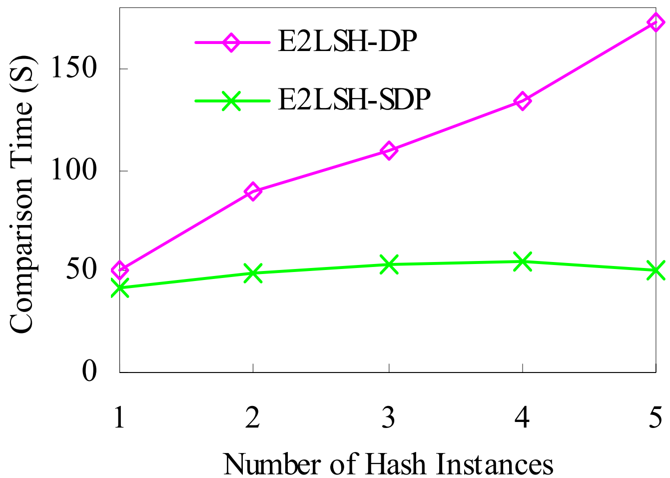
δ_{LSH}	0.01	0.02	0.03	0.04	0.05
VMP_{LSH}	0.133	0.255	0.400	0.537	0.669
δ_{E2LSH}	0.0025	0.005	0.0075	0.0100	0.0125
VMP_{E2LSH}	0.123	0.240	0.363	0.472	0.573

Increasing filtering threshold leads to a high VMP at the cost of more computation.

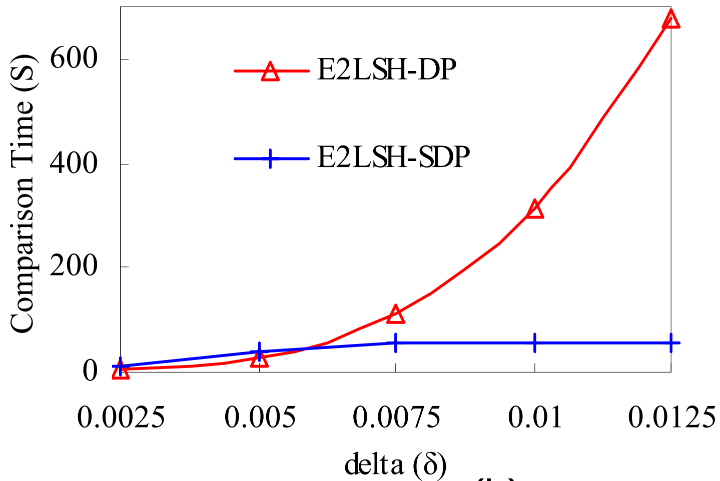
Nara Women's University

Experiments II -- Computation Time

- Computation is mainly considered in two aspects:
  - *Indexing the features by LSH/E^{2}LSH together with ANN*
  - *Comparing feature sequences*
- Short discussion
  - SDP has a very obvious superiority over DP
    - it avoids the calculation of feature distance
    - & its comparison time approaches a steady value, which guarantees worst retrieval time.
  - SDP outperforms DP



(a)



(b)

Nara Women's University

Comparison time in DP and SDP under different number of hash tables (δ_{E2LSH}=0.0075) or different filtering threshold δ (3 hash tables)

Experiments II -- Computation Time

- All the queries are performed under the different schemes
- Short discussion
  - Conventional DP without hashing takes the longest time
  - E2LSH-SDP accelerates retrieval speed by 42.7 times compared with conventional DP.

The total retrieval time consumed under different schemes

Scheme	LSH-DP	LSH-SDP	E2LSH-DP	E2LSH-SDP	DP
Time(s)	582.3	480.015	313.875	187.65	8014.95

Experiments III -- Retrieval Ratio

- A tradeoff is made between retrieval ratio and retrieval time
- With a suitable filtering threshold, the retrieval ratio is high enough while the computation time is controlled

Top-4 retrieval ratio of LSH/E^{2}LSH (3 hash tables) retrieval ratio under different filtering threshold δ

δ_{LSH}	0.01	0.02	0.03	0.04	0.05
LSH-DP	0.83	0.88	0.92	0.91	0.93
LSH-SDP	0.86	0.89	0.91	0.92	0.94
δ_{E2LSH}	0.0025	0.005	0.0075	0.01	0.0125
E^{2}LSH-DP	0.86	0.89	0.92	0.93	0.93
E^{2}LSH-SDP	0.88	0.91	0.93	0.93	0.94

δ_{LSH} =0.03 & δ_{E2LSH} =0.0075 are suitable thresholds since a smaller value decreases retrieval ratio while a larger value increases the computation cost.

Nara Women's University

What to Cover

- Background and motivation
- Short review -- ANN, LSH and E^2LSH
- Peer-to-peer network
- CBMR over peer-to-peer networks
- Challenges in scalable peer-to-peer environment
- Potential schemes of CBMR over P2P networks
- Current Work(Motivation,Methods, IBQBC Music Retrieval framework, Experiments and results)
- Conclusion and future work

Conclusion and Future Work

- Explain concept of peer-to-peer network
- Discuss CBMR over peer-to-peer networks
- Show some challenges in scalable peer-to-peer environment
- Introduce the potential schemes of CBMR over P2P networks
- Our contribution on current work
  - Established indexed framework for query-by-content audio retrieval
  - Effectiveness of proposed algorithms(E^2LSH-SDP, E^2LSH-DP,LSH-DP,LSH-DP)
- Future work
  - Evaluation of scalability of the proposed schemes with a larger database
  - Application of query-by-content audio retrieval over P2P network .

Nara Women's University

Thank You!