音樂搜尋與分析

陳良弼

國立政治大學

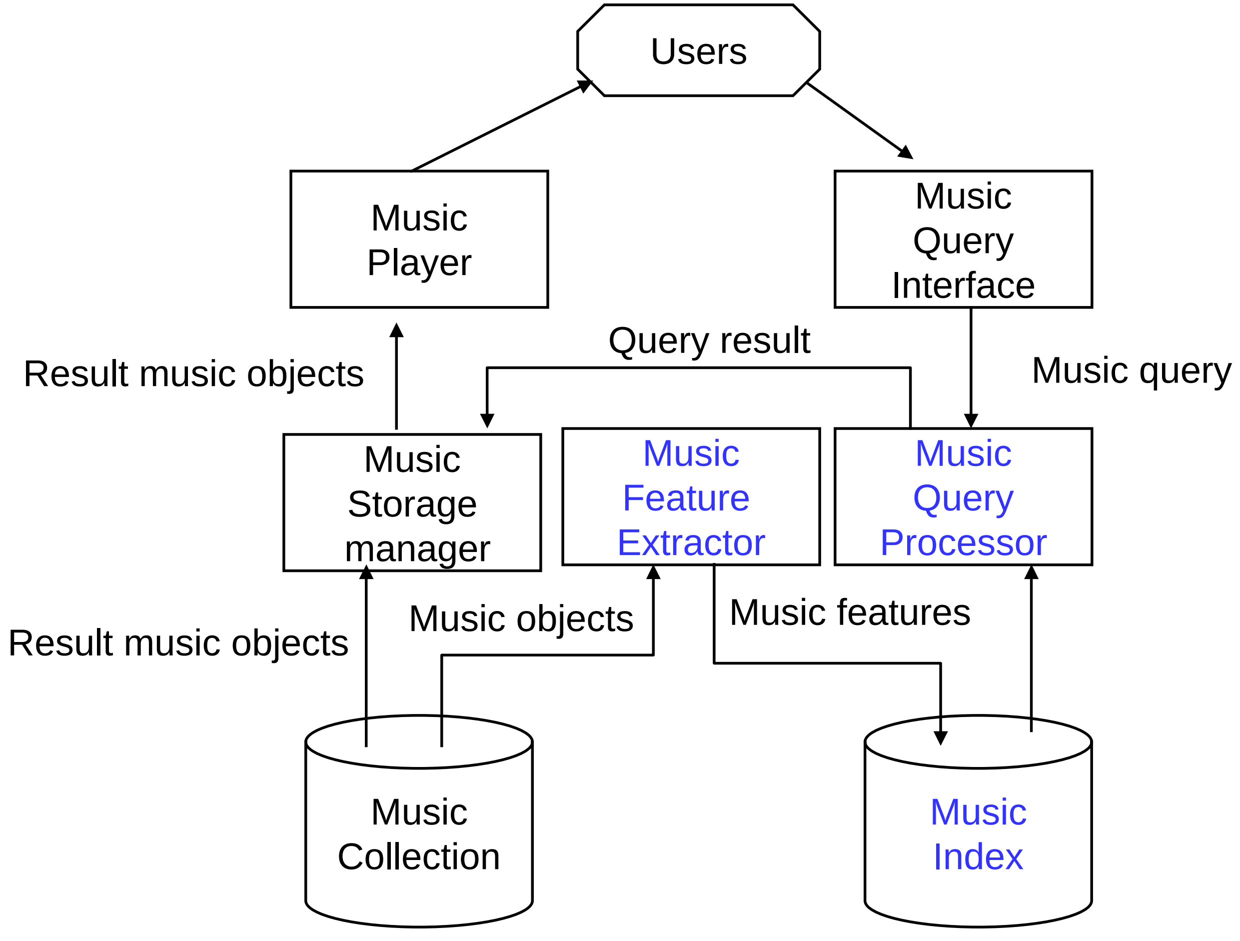
第五屆數位典藏技術研討會

August 31, 2006

Outline

- Music Retrieval
  - Architecture for music retrieval
  - Music representations
  - Query processing
  - Indexing
  - System Evaluation
- Music Analysis
  - Repeating pattern discovery
  - Phrase extraction
  - Music classification
  - Music recommendation
- Future Research Directions

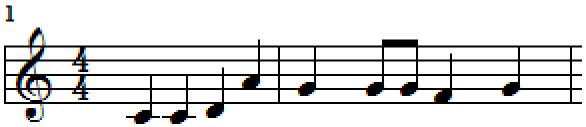
Architecture for Music Retrieval



Styles of Music Composition

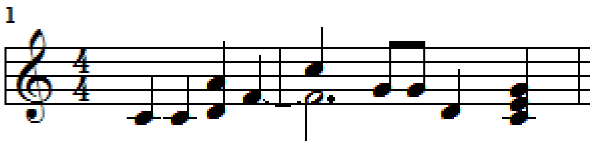
- Monophony
  - Monophonic music has at most one note playing at any given time; before a new note starts the previous note must have ended
- Polyphony
  - Polyphonic music has no such restrictions. Any note or set of notes may begin before any previous note or set of notes has ended

Monophony Representations



- Absolute measure
  - Absolute pitch
    - C5 C5 D5 A5 G5 G5 G5 F5 G5
  - Absolute duration
    - 1 1 1 1 1 0.5 0.5 1 1
  - Absolute pitch and duration
    - (C5,1)(C5,1)(D5,1)(A5,1)(G5,1)(G5,0.5)(G5,0.5)(F5,1)(G5,1)
- Relative measure
  - Contour (in semitones)
    - 0 +2 +7 -2 0 0 -2 +2
  - IOI (Inter onset interval) ratio
    - 1 1 1 1 0.5 1 2 1
  - Contour and IOI ratio
    - (0,1)(+2,1)(+7,1)(-2,1)(0,0.5)(0,1)(-2,2)(+2,1)

Polyphony Representations



- All information preservation
  - Keep all information of absolute pitch and duration (start_time, pitch, duration)
    - (1,C5,1)(2,C5,1)(3,D5,1)(3,A5,1)(4,F5,4)(5,C6,1)(6,G5,0.5)(6.5,G5,0.5)…
- Relative note representation
  - Record difference of start times and contour (ignore duration)
    - (1,0)(1,+2)(0,+7)(1,-4)…
- Monophonic reduction
  - Select one note at every time step (main melody selection)
    - (C5,1)(C5,1)(A5,1)(F5,1)(C6,1)...

Music Representation - Theme

- Theme
  - A short tune that is repeated or developed in a piece of music
- A small part of a musical work
  - Efficient retrieval
- A highly semantic representation
  - Effective retrieval
- Automatic theme extraction
  - Exact repeating patterns
  - Approximate repeating patterns

Query Processing

- String Matching Algorithms
  - Exact string matching
    - Brute-force method
    - KMP algorithm
    - Boyer-Moore algorithm
    - Shift-Or algorithm
  - Partial string matching
    - Shift-Or algorithm
  - Approximate string matching
    - Edit distance
    - Dynamic programming
    - Candidate pruning [Liu, Wu, Chen, ACM MSJ]

Indexing

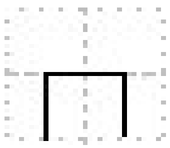
- Tree-based index (Suffix tree)
- List-based index (1D-list)
- N-gram index

Tree-Based Index

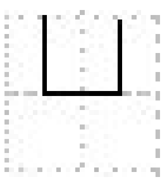
- [Chen, et al., ICME'00]
- Music objects are coded as strings of music segments
  - Four segment types to model music contour
  - Pitch and duration are considered
- Index structures
  - Augmented suffix tree
- Both incipit/partial and exact/approximate matching can be handled

Tree-Based Index (Cont.)

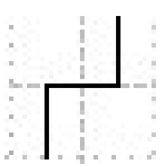
Four segment types



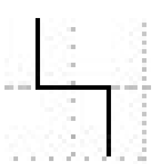
type A



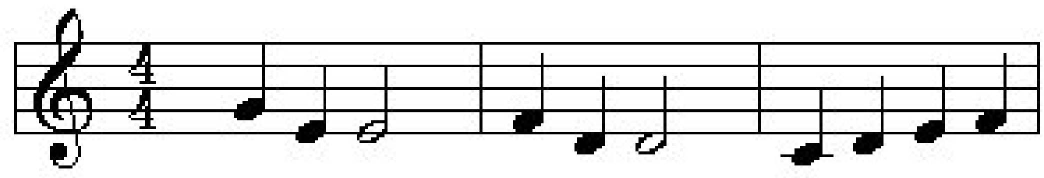
type B

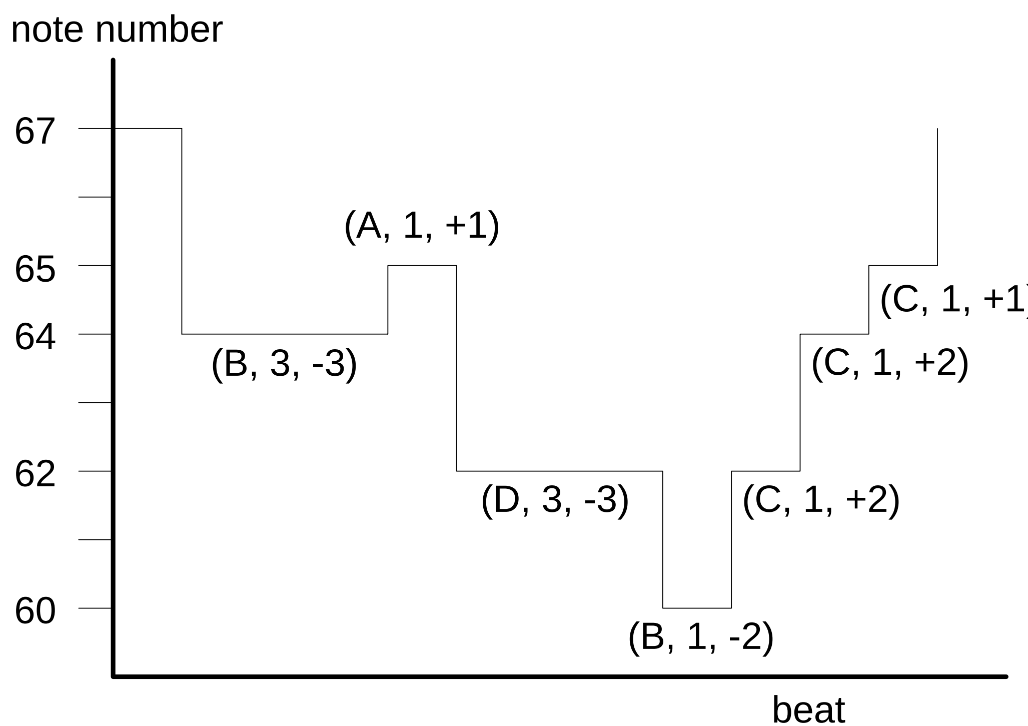


type C

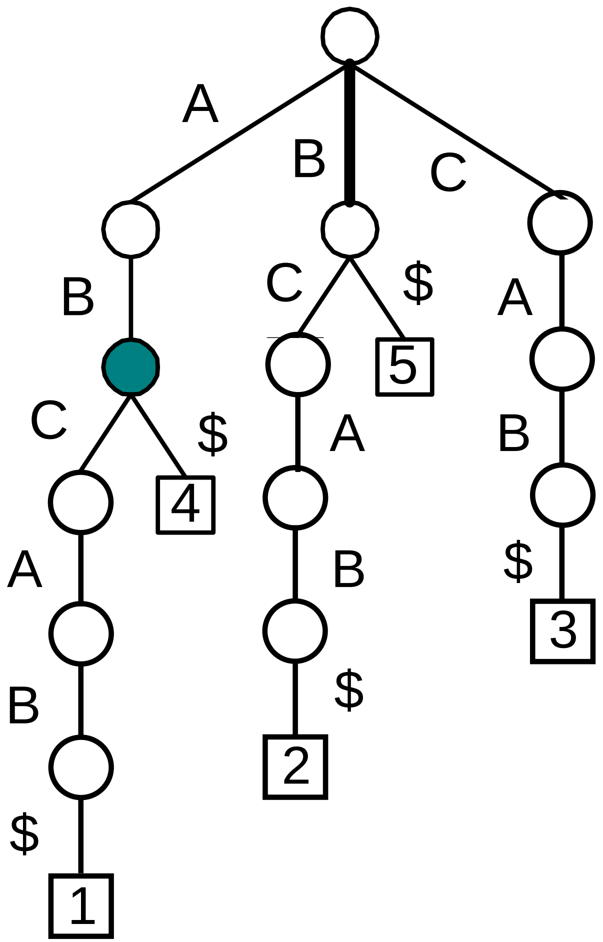


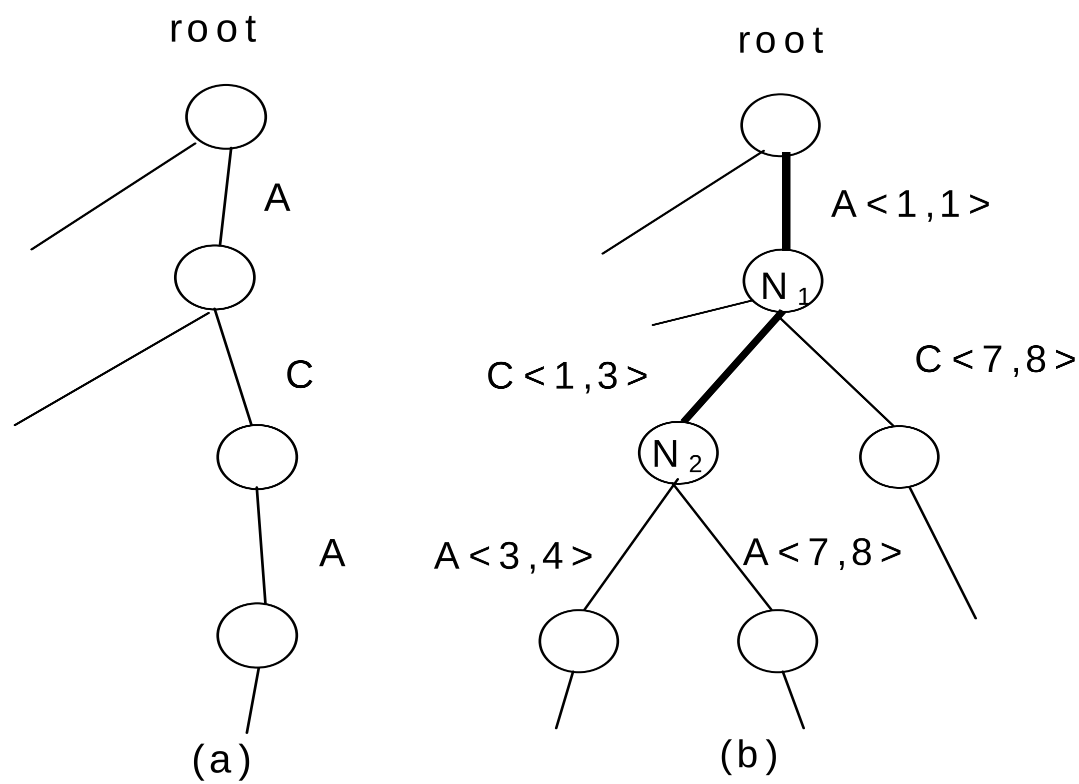
type D





Tree-Based Index (Cont.)





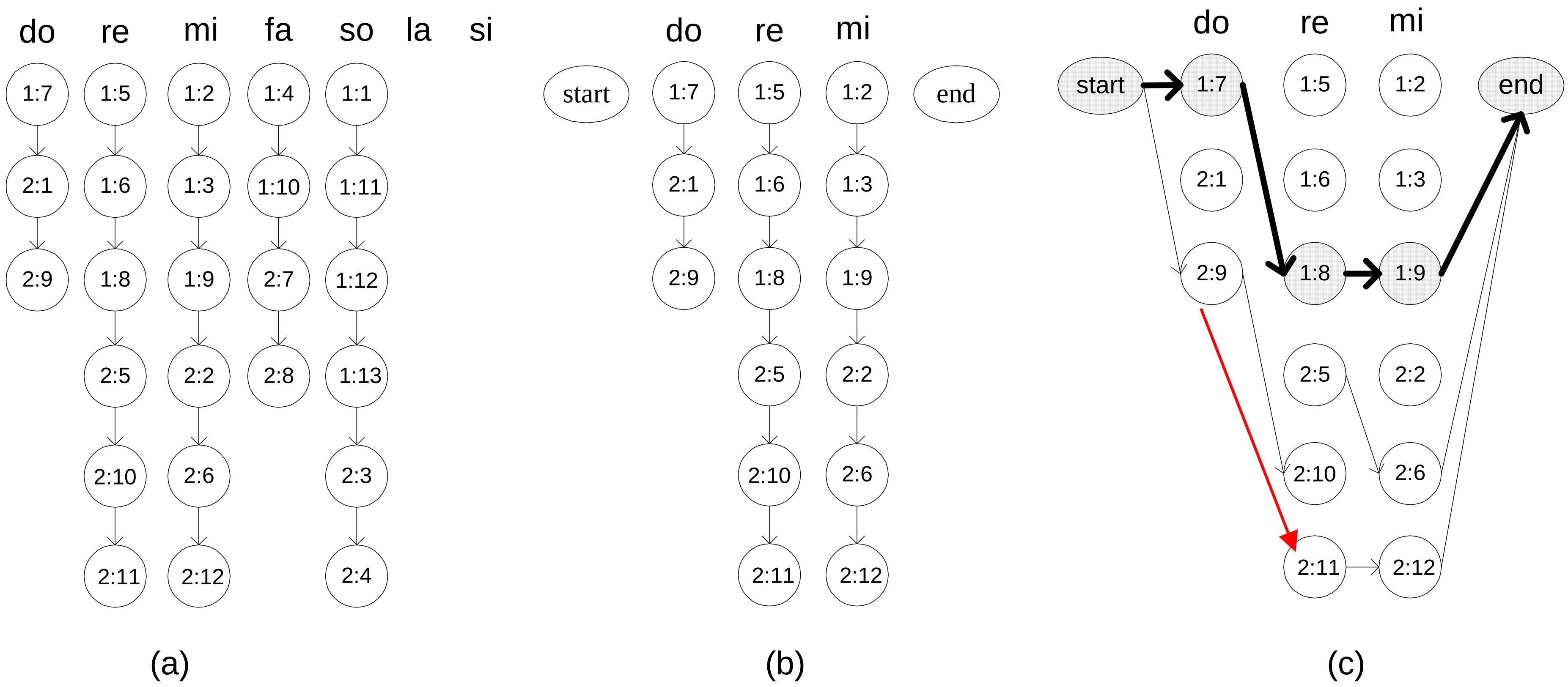
(a) An example suffix tree
(b) A 1-D augmented suffix tree

The suffix tree of the string S="ABCAB"

List-Based Index

- [Liu, Hsu and Chen, ICMCS'99]
- Music objects are coded as melody strings
  - "so-mi-mi-fa-re-re-do-re-mi-fa-so-so-so"
- Melody strings are organized as linked lists
- Both incipit/partial and exact/approximate matching can be handled
  - Exact link, insertion link, dropout link, transposition link

List-Based Index (Cont.)



N-Gram Index

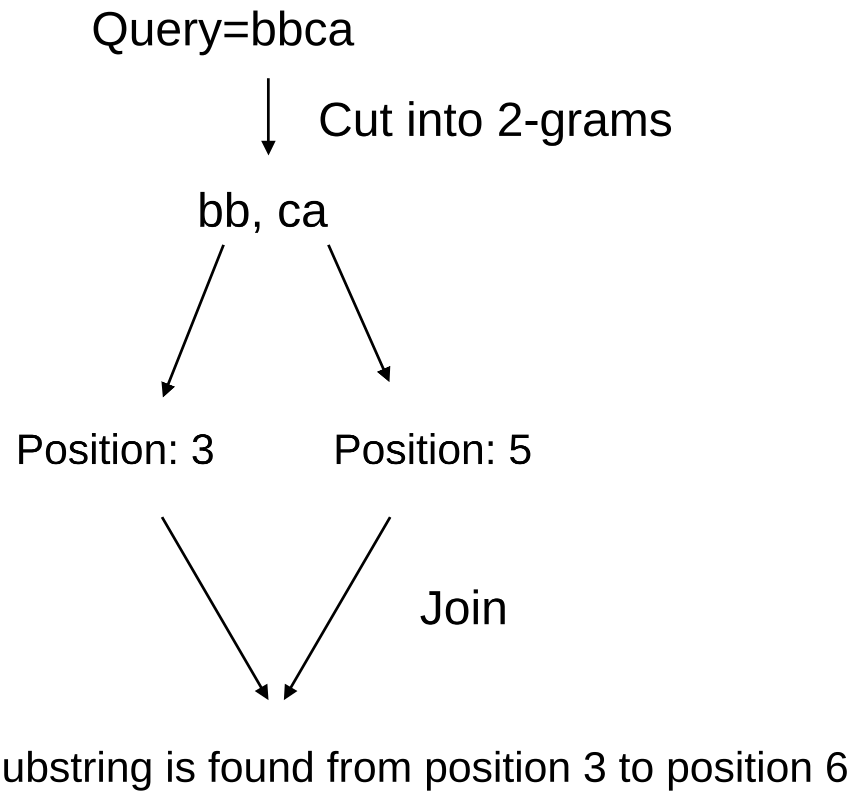
- [Doraisamy and Ruger, ISMIR'02]
- A widely used technique in music databases
- Target strings are cut into index terms by a sliding window with length N
- Index can be implemented by various methods, e.g., inverted file
- Queries are also cut into index terms with length N
- Searching and joining are then performed

N-Gram Index (Cont.)

S=aabbcaab

2-Gram	Position
aa	1,6
ab	2,7
bb	3
bc	4
ca	5

Inverted file



System Evaluation

- Traditional measures of effectiveness are precision and recall

$$precision = \frac{number\ of\ retrieved\ references\ that\ are\ relevant}{number\ of\ references\ that\ are\ retrieved}$$

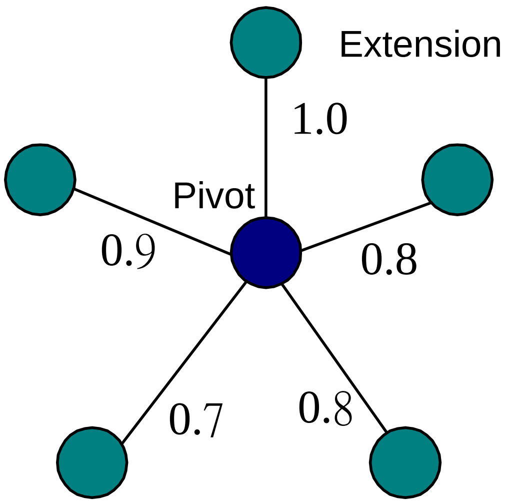
$$recall = \frac{number\ of\ retrieved\ references\ that\ are\ relevant}{number\ of\ relevant\ references}$$

A Platform for Evaluating MIR Systems

- Evaluation of various music retrieval approaches
  - Efficiency
  - Effectiveness
- The Ultima project builds such a platform [Hsu, Chen and Chen, CIKM'01]
  - Same data set and query set for various approaches
  - Compare recall-precision curves

Repeating Pattern Discovery

- A repeating pattern in music data is defined as a sequence of notes which appears more than once in a music object
- The themes or motives are typical kinds of repeating patterns
- Exact repeating patterns [Hsu, Liu and Chen, TMM'01]
  - By the string-join operator
- Approximate repeating patterns [Liu, Wu and Chen, Dasfaa'05]



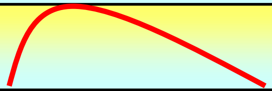
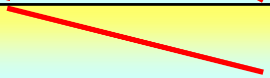
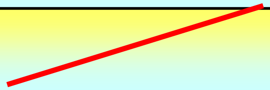
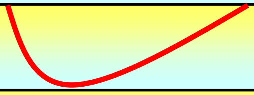
- Each vertex represents a music segment and the number on each link indicates the similarity degree between the two segments
- The center is called a *pivot;* if the similarity degree to the pivot is larger than the predefined threshold, the associated segment is then called an *extension* of the pivot
- A pivot is an ARP if the number of extensions reaches the support threshold

Phrase Extraction

- Two features used for phrase extraction
  - Duration and rest
- Melodic Shapes [Huron, Computing in Musicology'95]
  - Statistics Information in Western Folksongs
    - The most common length of a phrase is 8 notes
    - Half of all phrases are between 7 and 9 notes in length
    - Three-quarters of all phrases are between 6 and 10 notes in length

Phrase Extraction (Cont.)

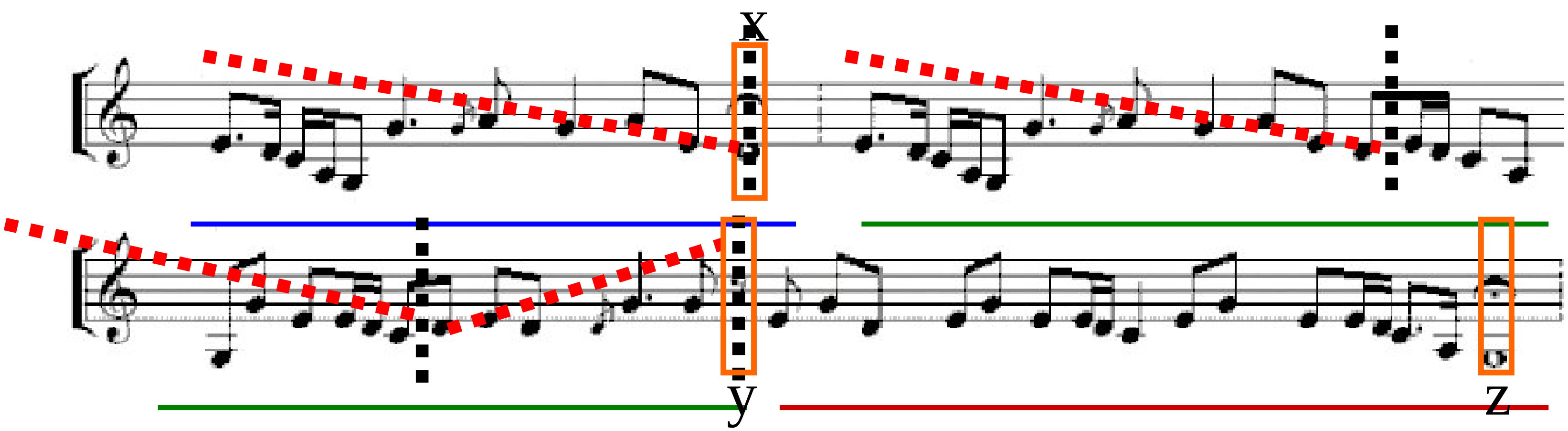
A: the pitch value of the first note in the target phrase
B: the pitch value of the last note in the target phrase
C: the average pitch value of the remaining notes in the target phrase

Contour Type	Number of Phrases	Percentage	Arch Shape	Definition
Convex	13926	38.6%		$A<C \wedge B<C$
Descending	10376	28.8%		$A>C>B$
Ascending	6983	19.4%		$A<C<B$
Concave	3496	9.7%		$A>C \wedge B>C$
Others	1294	3.5%		

Phrase Extraction (Cont.)

- Identify the positions of all the terminative notes
- Extract the music pieces according to the terminative notes
- Select the candidate music pieces for decomposition based on the length information
  - If the length ≤ 12, the music piece is marked as a phrase
  - If the length > 12, decompose the music piece into phrases
    - convex > descending > ascending > concave

Phrase Extraction (Cont.)



64 62 60 57 55 67 67 69 67 69 64 62

64 62 60 57 55 | 67 67 69 67 69 64 62 | 64 62 60 57 55 67 64 64 62 60 62 64 62 62 67 67

Convex?
No

Descending?
Length = 12, A = 64, B = 62, C = 63.7

Order	The Length of the Prefix Fragment	The Pitch of the First Note	The Pitches of the Remaining Notes	The Pitch of the Last Note
		4	62, 60, 57, 55	67
		4	62, 60, 57, 55, 67	67
		4	62, 60, 57, 55, 67, 67	69
		4	62, 60, 57, 55, 67, 67, 69	67
		4	62, 60, 57, 55, 67, 67, 69, 67	69
		4	62, 60, 57, 55, 67, 67, 69, 67, 69	64
7	12	64	62, 60, 57, 55, 67, 67, 69, 67, 69, 64	62

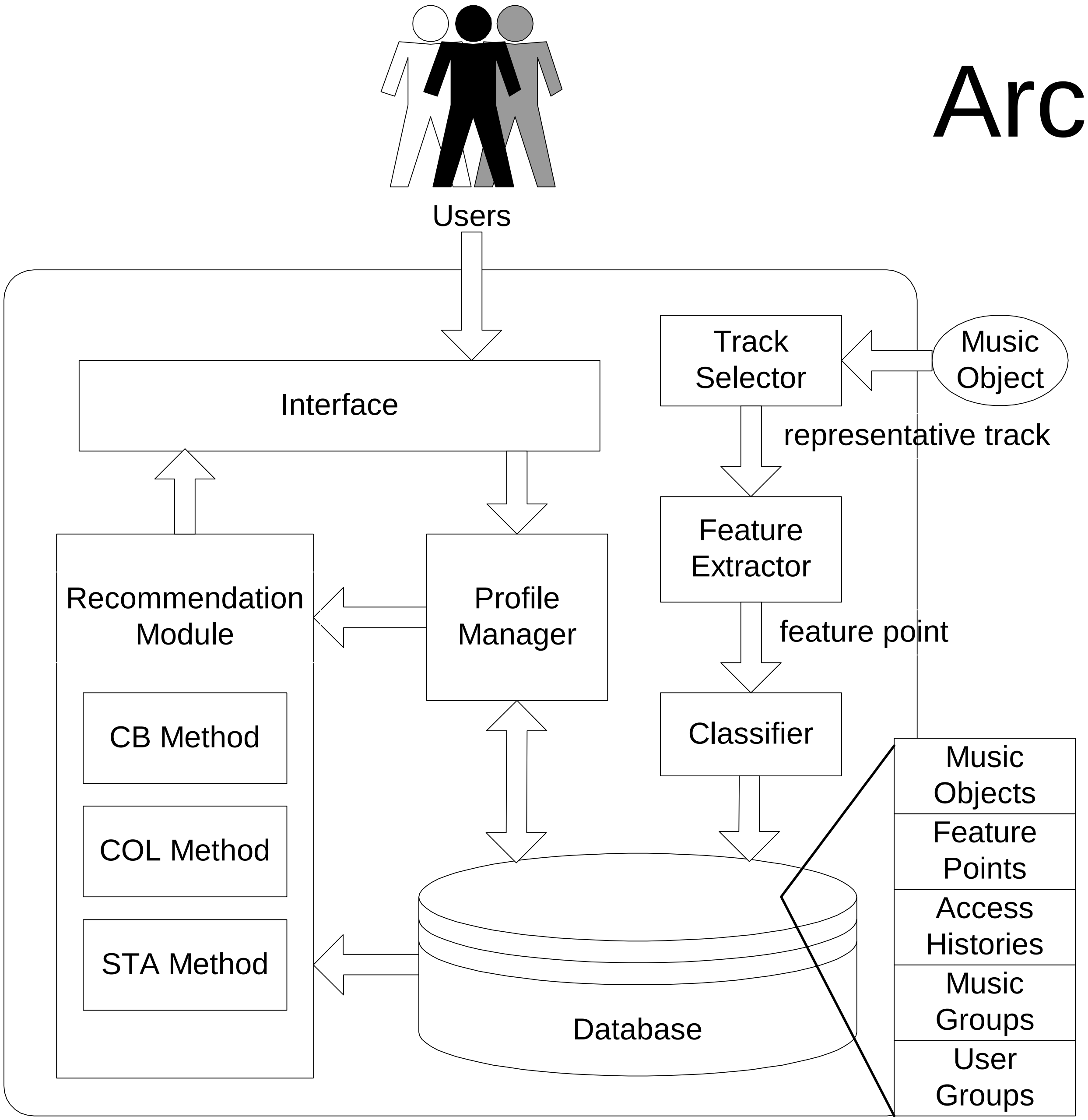
Music Classification

- Different kinds of music units can be extracted from music objects, such as repeating patterns and phrases
- Different kinds of music units may have different semantics in musicology
- These extracted music units can be used in music classification, retrieval, and analysis

Music Recommendation Systems

- [Chen and Chen, CIKM'01]
- The results of music classification can be used for music-related services
- By analyzing the user access histories, we can discover which music classes the users may be interested in and which users belong to the same group
- By using different kinds of recommendation mechanisms, we can recommend the users suitable music objects

Architecture



A polyphonic music object

- one melody track
- other accompaniment tracks

Recommendation Mechanisms

- Content-based filtering approach
  - Similarity between music objects and user profiles
  - Recommend the music objects that belong to the music groups the user is recently interested in
- Collaborative filtering approach
  - Similarity between user profiles
  - Provide collaborative recommendations to the users in the same user group
- Statistical approach
  - Recommend “hot” music objects

Future Research Directions

- Similarity measures, index structures and segmentation for polyphonic music
- Cross-media retrieval
- Music retrieval based on user's emotion/situation
  - Music mood analysis
  - music therapy
  - Creating a High Quality Learning, Relaxing and Sleeping Environment through Affective and Attentive Computing