Content-Based Audio Retrieval

Eugene Weinstein, NYU
June 16th, 2008
Content-based Audio Retrieval

• Say I play 5-10 seconds of audio of a song for you
• You have to figure out what the song is
  • Not so hard, right?
• But what if I ask you to do this for 1,000,000 songs and get it right 99% of the time?
  • A little harder
• And what if I add some noise to the recording?
  • Even harder!
Building a Song ID System

• Take something like this

• And match it to something like this
Feature Selection

- Looking at the spectrogram, pick out identifying information
- E.g., spectral peaks
Temporal Sweep

- Go over recording of entire song
- Record spectral peak positions for each 100ms chunk
- e.g.,

\[(2,5) \quad (22,1) \quad (3,11)
(3,8) \quad (3,4) \quad (29,6)
(5,23) \quad (5,16) \quad (17,8)...
(15,7) \quad (28,6) \quad (15,6)
(8,17) \quad (15,2) \quad (8,17)\]
Index: Hash Table

- Hash table is a key-value mapping, e.g.
  - Key = student, Value = seat number
  - Key = date, Value = high temperature
  - Key = spectral peak, Value = corresponding song

- We are interested in two hash table operations:
  - Insert(key, value)
  - Lookup(key)
Hash Table Example

- Key = student name, Value = ID number
- Idea: split into “buckets” by first letter of name

Adam \[\rightarrow A \rightarrow \text{Adam:123-45-6789}\]
Brian \[\rightarrow B \rightarrow \text{Brian:987-65-4321}\]
Mike \[\rightarrow M \rightarrow \text{Mike:321-54-9876}\]
Mary \[\rightarrow N \rightarrow \text{Mary:111-22-3333}\]

...
Hash Table Lookup

- What is Brian’s ID number?

```
Adam → A → Adam: 123-45-6789
Brian → B → Brian: 987-65-4321
...                        ...
Mike → M → Mike: 321-54-9876
Mary → N → Mary: 111-22-3333
...                        ...
```
System Training

- Compute spectral peaks over audio for each song
- Insert into hash table

<table>
<thead>
<tr>
<th>Keys</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2,5)</td>
<td>(22,1)</td>
</tr>
<tr>
<td>(3,8)</td>
<td>(3,4)</td>
</tr>
<tr>
<td>(5,23)</td>
<td>(5,16)</td>
</tr>
<tr>
<td>(15,7)</td>
<td>(28,6)</td>
</tr>
<tr>
<td>(8,17)</td>
<td>(15,2)</td>
</tr>
</tbody>
</table>

Hash Table

Ben Folds Five, Brick
Using the System

• Compute Spectral Peaks: Key
• Lookup in table: Value

(22,1)
(3,4)
(5,16)
(28,6)
(15,2)

Which Song?

Hash Table

Ben Folds Five, Brick