The MAMI Query-By-Voice Experiment
Collecting and annotating vocal queries for music information retrieval

IPEM, Dept. of musicology, Ghent University, Belgium
Micheline Lesaffre & Koen Tanghe @ ISMIR 2003
Outline

• About the MAMI project
• Aim of the QBV experiment
• Description of the setup of the experiment
• Methods used for annotation
• Global view on results of statistical analysis
• Some examples of output files
Taxonomy Driven Feature Extraction

Query System Input Processor

Audio Database

Feature Extraction

User Profile Processing

Feature Extraction

Taxonomy Driven

Abstract Representations

Similarity Matching

QUERY RESPONSE

Audio

Text

User

Abstract Representations
Aim of the QBV experiment

• Analysis of spontaneous user behavior
• Collecting raw data
• Setting up an annotated database for developing and testing QBV MIR systems
• Making the data available for MIR research
The rough guide to the QBV experiment

Input
• 30 pieces of music (different styles), presented using title + performer, or using audio itself
• 72 human subjects

Output
• profile files of the subjects
• log files of the experiment flow
• around 1500 query sound files (44.1 kHz, 16-bit mono)
• around 270 of these: imitations of the same fragment performed by different subjects in different ways

Physical setup
• software written in C++, running on Windows
• normal "office" environment
• standard consumer-level equipment
• duration: about 35 minutes
Experiment overview

Preparatory stage

- Collecting info on the subject
- Collecting info on the subject's knowledge of the musical pieces

Experiment parts

Part 1
- Imitating known pieces without hearing them first

Part 2
- Imitating pieces after hearing them in their entirety first

Part 3
- Imitating a fixed fragment in four different ways
Preparatory stage

Collecting info on the subject
unique ID, age, gender, listening to music (how much),
playing music (yes/no + how much), highest level of
musical education

Collecting info on subject's knowledge of
the musical pieces
presentation of title + composer/performer
classification into different sets according to:
"would you be able to imitate a fragment of this piece":

<table>
<thead>
<tr>
<th>Set</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set1</td>
<td>fixed set of pieces from MAMI target database</td>
</tr>
<tr>
<td>Set3</td>
<td>known and imitable</td>
</tr>
<tr>
<td>Set4K</td>
<td>not known</td>
</tr>
<tr>
<td>Set4R</td>
<td>thought to be known, but not remembered</td>
</tr>
<tr>
<td>Set5</td>
<td>fixed fragment to be imitated in different ways</td>
</tr>
<tr>
<td>Set6</td>
<td>known, but not imitable</td>
</tr>
</tbody>
</table>
Experiment part 1

Focus: reproduction of known pieces from long-term memory

Presentation: only title and composer/performer/…

Subject is asked to "imitate the piece vocally"
  • free choice of fragment and voice/instrument
  • suggested examples of vocal imitation:
    - humming
    - singing the text
    - singing using a syllable
    - whistling
    - mixed
  • two attempts allowed

Other ways to describe the musical piece
  • sound recording (other ways than before)
  • verbal description of the piece
  • description of another method
Experiment part 2

Focus
imitation from short-term memory
what tends to "stick" after just hearing a piece

Presentation
entire piece + title and composer/performer/…
aim: 2 "not known" and 2 "known, but not remembered"

Subject is asked
• if he/she heard the piece before
• to "imitate the piece vocally" (same as in Part 1)
Focus
differences in performances of same melody by various subjects using different query methods

Presentation
short musical fragment + title and composer/performer/…
can be listened to up to three times

Subject is asked
• if he/she heard the piece before
• to imitate the piece using the following methods:
  - humming
  - singing the text (text is shown on screen)
  - singing using "tatata"
  - whistling (if possible)
Annotation strategy

1. Model-oriented annotation
   • detailed description of low and mid-level acoustical features
   • for testing transcription modules

2. User-oriented annotation
   • knowledge about human attitudes
   • concentrate on naturally expressed vocal queries
   • user-friendly systems for content-based access
   • carried out for 1148 queries
   • focus on:
     • Impact of memory recall
     • Effects of gender, age and musicianship
     • Performance way
     • Query method
Features: model-oriented annotation

- Onset + sureness quotation
- Frequency
- Pitch stability
- Query method
Features: user-oriented annotation

General aspects

• Timing
• Segmentation

Segment specific aspects

• Timing
• Vocal query method
• Performance style
• Target similarity
• Syllabic structure
Overview user-oriented annotation

- Timing
- Query methods
- Syllable structure
- Effects of age, gender, musical experience
- Effects of memory

Micheline Lesaffre & Koen Tanghe @ ISMIR 2003
• Average starting time
  634 msec
• Mean query length
  14.04 sec
<table>
<thead>
<tr>
<th>query method</th>
<th># of segments</th>
<th>% of segments</th>
<th>total time</th>
<th>% of total time</th>
</tr>
</thead>
<tbody>
<tr>
<td>text</td>
<td>926</td>
<td>45.60 %</td>
<td>5558959</td>
<td>37.40 %</td>
</tr>
<tr>
<td>syllabic</td>
<td>766</td>
<td>37.80 %</td>
<td>6056644</td>
<td>40.80 %</td>
</tr>
<tr>
<td>whistle</td>
<td>174</td>
<td>8.60 %</td>
<td>2544864</td>
<td>17.10 %</td>
</tr>
<tr>
<td>hum</td>
<td>101</td>
<td>5.00 %</td>
<td>541815</td>
<td>3.60 %</td>
</tr>
<tr>
<td>comment</td>
<td>42</td>
<td>2.10 %</td>
<td>65108</td>
<td>0.40 %</td>
</tr>
<tr>
<td>percussion</td>
<td>20</td>
<td>1.00 %</td>
<td>77394</td>
<td>0.50 %</td>
</tr>
</tbody>
</table>
### Query methods: user categories

<table>
<thead>
<tr>
<th>METHOD</th>
<th>N SUBJECTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>(total N = 71)</td>
<td></td>
</tr>
<tr>
<td>one</td>
<td>38</td>
</tr>
<tr>
<td>18 : text</td>
<td></td>
</tr>
<tr>
<td>16 : syllable</td>
<td></td>
</tr>
<tr>
<td>04 : whistle</td>
<td></td>
</tr>
<tr>
<td>two</td>
<td>17</td>
</tr>
<tr>
<td>15 : text + syllable</td>
<td></td>
</tr>
<tr>
<td>01 : text + whistle</td>
<td></td>
</tr>
<tr>
<td>more</td>
<td>16</td>
</tr>
<tr>
<td>01 : syllable + whistle</td>
<td></td>
</tr>
</tbody>
</table>

5 user categories:
- 1/4 prefer one method text
- 1/4 prefer one method syllable
- 1/4 prefer two methods text + syllable
- 1/4 prefer more methods
- ---- one method whistlers
Effects of age

Increase of

• similarity
• use of comment
• average starting time
• use of syllable nuclei [a]
• use of onset [l]
Effects of gender

Timing
women start querying later

Syllable choice
onset: men prefer [t]
nuclei: women prefer [a]
men vary more
Effects of musicianship

Timing
Musicians produce longer queries

Methods used
Musicians less often sing the text
Effects of memory

On query method

**Textual dominance decreases**

- LTM: 48.7% / 41.7%
- LTM+STM: 39.7% / 33.3%
- STM: 34.4% / 26.6%

**Syllabic dominance increases**

- LTM: 34.9% / 36.0%
- LTM+STM: 43.1% / 47.2%
- STM: 49.1% / 58.3%

**Importance of whistling decreases**

- LTM: 8.6% / 18.0%
- LTM+STM: 9.5% / 15.3%
- STM: 4.3% / 8.0%
Effects of memory

On performance style

Melodic performances decrease

LTM: 73,9% / 79,6%
LTM+STM: 69,0% / 73,7%
STM: 47,2% / 51,7%

Intermediate performances increase

LTM: 19,1% / 18,2%
LTM+STM: 25,6% / 22,8%
STM: 45,5% / 41,9%

Rhythmic performances increase

LTM: 4,7% / 1,8%
LTM+STM: 3,7% / 3,2%
STM: 5,5% / 5,8%
Access to the files

MAMI project web site:

http://www.ipem.ugent.be/MAMI

QBV experiment files:
  • go to the Public section
  • look for: Test collections and annotation material
MAMI: Musical Audio-Mining

"Query by Humming"

Situating the project

MAMI is a data-mining project for audio recognition that investigates ways of searching an audio archive as easily as you can search a text archive. The project starts from the observation that given the current state-of-the-art in telematics, the technological orientation of the music culture and the interest of the music industry to sell musical commodities and services via the Internet, there is a high need to develop advanced tools that support new ways to deal with content concerning musical audio and associated processing. Current technology makes it possible to retrieve music from a database using new content-based methods. Performing feature extraction on a wide range of sound characteristics opens the possibility for multiple ways of querying on data not only by text queries but also by music-based query techniques such as query-by-humming or query by specification of a list of musical variables.

A main characteristic of the MAMI-project is its focus on music as audio signal. This includes all kinds of music, including electro-acoustical music as well as ethnic and world music.

Project aims

- Develop a background epistemology for audio mining that is based on auditory modelling and perception theory.
- Work out methodologies, techniques and software tools for content-based musical audio mining taking into account all kinds of music.
- Development of an integrated system for audio description using different levels of representation.
- Work towards a practical application which demonstrates its usefulness by means of the so-called "query-by-humming" paradigm.
- Allow users to retrieve a musical piece by describing sound characteristics, either by humming or playing or describing the piece on the basis of its sound characteristics.
- Set up representational structures in compliance with the MPEG-7 standard, an interface for Multimedia Content Description.

Description levels

The MAMI-research project uses different description levels to describe music, such as:

- Waveform representation.
- Frame-based representations.
- Parameter-based representations.
- Event-based representations.
Papers

- "Musical Audio Mining", M. Leman, in: J. Meij (Ed.), Dealing with the Data Flood: Mining data, text and multimedia, Rotterdam: STT Netherlands Study Centre for Technology Trends, 2002 (PDF)

Software

- A melody transcription demo program can be found here.
- A library (C/C++) for transcription of monophonic query melodies can be found here.
- A QBH demonstrator (MAMI melody transcription demo coupled to the Philips back end) (not publicly available)

Also working towards:
- prototype application incorporating existing modules

Test collections and annotation material

- The material gathered by and for the MAMI query by voice experiment (test sets, annotations and corresponding documentation) can be found here.

Participation in conferences and meetings
<table>
<thead>
<tr>
<th>Example Description</th>
<th>Wavefile 1</th>
<th>Wavefile 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singing lyrics</td>
<td>010_030_EXP2_QbV1.wav</td>
<td>022_062_EXP1_QbV1.wav</td>
</tr>
<tr>
<td>Mixed: percussion and singing lyrics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whistling</td>
<td>132_036_EXP2_QbV1.wav</td>
<td>074_073_EXP2_QbV1.wav</td>
</tr>
<tr>
<td>Mixed: singing lyrics, whistling and percussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humming</td>
<td>012_019_EXP3_hum.wav</td>
<td>132_054_EXP2_QbV1.wav</td>
</tr>
<tr>
<td>Mixed: singing syllables and percussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percussion</td>
<td>027_078_EXP1_QbV2.wav</td>
<td>022_006_EXP1_QbV1.wav</td>
</tr>
<tr>
<td>Mixed: singing lyrics and comments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Good” query</td>
<td>052_058_EXP1_QbV1.wav</td>
<td>041_011_EXP2_QbV2.wav</td>
</tr>
<tr>
<td>Mixed: singing lyrics and syllables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Bad” query</td>
<td>045_071_EXP2_QbV1.wav</td>
<td>052_067_EXP1_QbV1.wav</td>
</tr>
<tr>
<td>Mixed: comments and singing lyrics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>original</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>