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

AUDIO DATABASE

Query System Input Processor

AUDIO TEXT

Feature Extraction

User Profile Processing

Feature Extraction

Taxonomy Driven

Taxonomy Driven

Abstract Representations

Abstract Representations

Similarity Matching

QUERY RESPONSE

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

- Imitating known pieces without hearing them first — Part 1
- Imitating pieces after hearing them in their entirety first — Part 2
- Imitating a fixed fragment in four different ways — Part 3

Preparatory stage

Collecting info on the subject

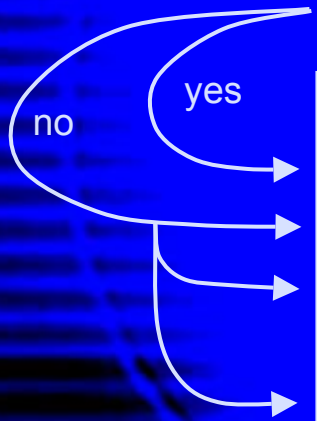
unique ID, age, gender, listening to music (how much), playing music (yes/no + how much), highest level of musical education

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":



Set1	fixed set of pieces from MAMI target database
Set3	known and imitable
Set4K	not known
Set4R	thought to be known, but not remembered
Set5	fixed fragment to be imitated in different ways
Set6	known, but not imitable

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)

Experiment part 3

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

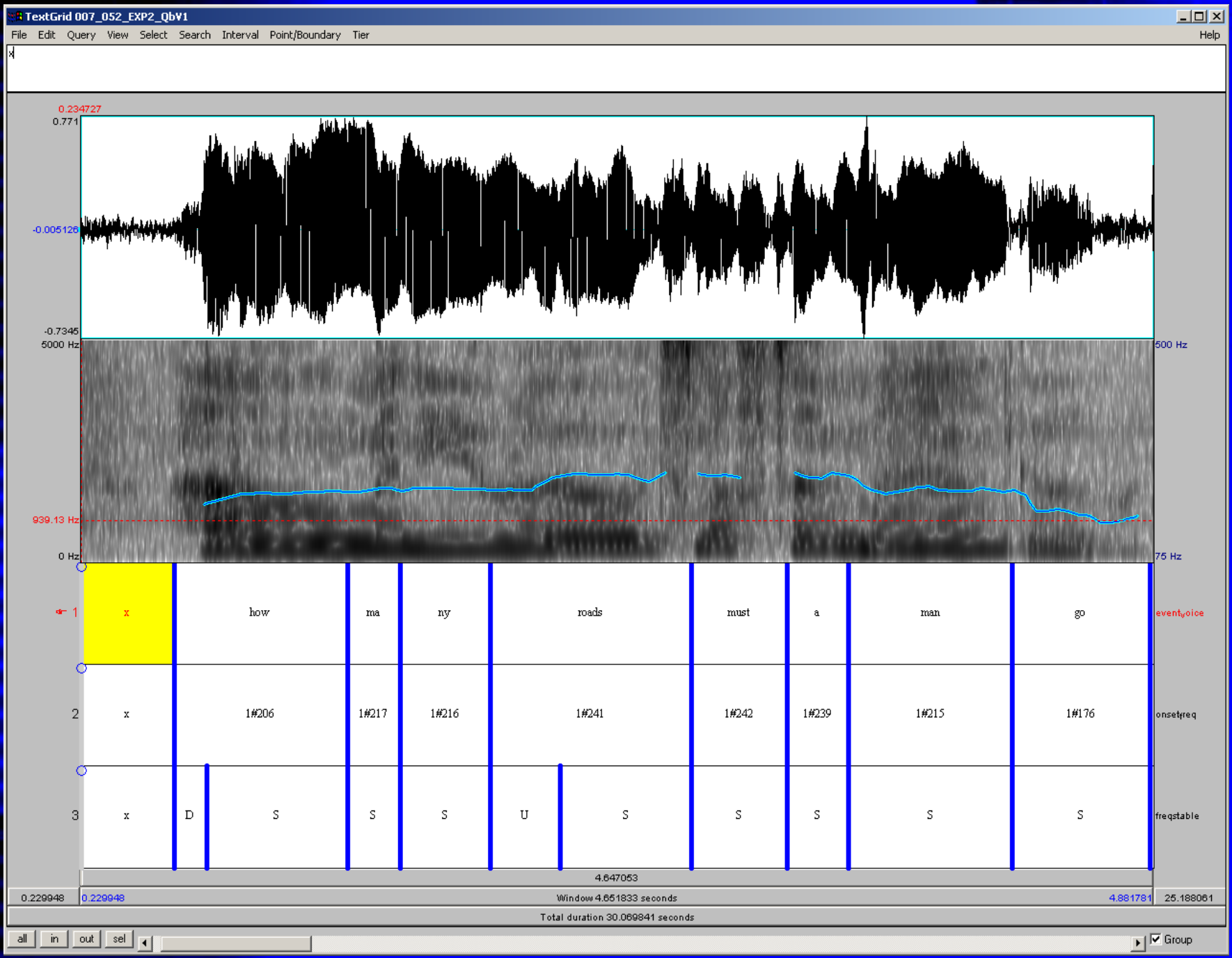
- detailed description of low en mid level acoustical features
- for testing transcription modules

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


TextGrid 007_052_EXP2_QbV1
File Edit Query View Select Search Interval Point/Boundary Tier Help
0.234727
0.771
-0.005126
-0.7345
5000 Hz
500 Hz
939.13 Hz
0 Hz
75 Hz
1 x how ma ny roads must a man go eventvoice
2 x 1#206 1#217 1#216 1#241 1#242 1#239 1#215 1#176 onsetfreq
3 x D S S S U S S S S S freqstable
4.647053
0.229948
0.229948
Window 4.651833 seconds
4.881781
25.188061
Total duration 30.069841 seconds
all in out sel
Group

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

Timing

- Average starting time

  634 msec

- Mean query length

  14.04 sec

Query methods

query method	# of segments	% of segments	total time	% of total time
text	926	45.60 %	5558959	37.40 %
syllabic	766	37.80 %	6056644	40.80 %
whistle	174	8.60 %	2544864	17.10 %
hum	101	5.00 %	541815	3.60 %
comment	42	2.10 %	65108	0.40 %
percussion	20	1.00 %	77394	0.50 %

Query methods: user categories

METHOD	N SUBJECTS	
	(total N =71)	
one	38	
		18 : text
		16 : syllable
		04 : whistle
two	17	
		15 : text +syllable
		01 : text + whistle
		01 : syllable + whistle
more	16	

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**

Institute for Psychoacoustics
and Electronic Music
Dept. of Musicology Ghent University
Blandijnberg 2, B-9000 Ghent, Belgium
info@ipem.UGent.be

MAMI: Musical Audio-Mining

"Query by Humming"

Public	Public stuff: papers, software, attended conferences, meetings and presentations, links, ...
MAMI External	Only for MAMI crew, user panel and external project responsibles
MAMI Internal	Only for MAMI crew
Calendar	Dates of meetings, deliverables, workshops, ...

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 telematica, 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 epistomology 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.

Institute for Psychoacoustics
and Electronic Music
Dept. of Musicology Ghent University
Blandijnberg 2, B-9000 Ghent, Belgium
info@ipem.UGent.be

MAMI PUBLIC

Papers

- *"An Auditory Model Based Transcriber of singing sequences"*, L.P. Clarisse, J.P. Martens, M. Lesaffre, B. De Baets, H. De Meyer and M. Leman, in: Proceedings of ISMIR, Paris, France, 2002 (PDF)
- *"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)
- *"Tendencies, Perspectives, and Opportunities of Musical Audio-Mining"*, M. Leman, L.P. Clarisse, B. De Baets, H. De Meyer, M. Lesaffre, G. Martens, J.P. Martens, and D. Van Steelant, in: A. Calvo-Manzano, A. Pérez-López, J. Salvador Santiago (Eds.), Forum Acusticum Sevilla 2002, 16-20 September, Madrid, (Special issue of Journal Revista de Acústica Vol XXXIII, no. 3-4), Madrid: Sociedad Española de Acústica -SEA, 2002 (PDF)
- *"Discovering Structure and Repetition in Musical Audio"*, D. Van Steelant, B. De Baets, H. De Meyer, M. Leman, J.-P. Martens, L. Clarisse, M. Lesaffre, in: Proceedings of Eurofuse Workshop, Varenna, Italy, 2002 (PDF)
- *"A Tonality-oriented Symbolic Representation of Musical Audio Generated by Classification Trees"*, G. Martens, H. De Meyer, B. De Baets, M. Leman, J.-P. Martens, L. Clarisse, M. Lesaffre, in: Proceedings of Eurofuse Workshop, Varenna, Italy, 2002 (PDF)
- *"User-dependent taxonomy of musical features as a conceptual framework for musical audio-mining technology"*, M. Lesaffre, M. Leman, K. Tanghe, B. De Baets, H. De Meyer and J.-P. Martens, in: Proceedings of SMAC 03 Stockholm Music Acoustics Conference 2003 (PDF)
- *"User Behavior in the Spontaneous Reproduction of Musical Pieces by Vocal Query"*, M. Lesaffre, D. Moelants, M. Leman, B. De Baets, H. De Meyer, G. Martens and J.-P. Martens, in: Proceedings of the 5th Triennial ESCOM Conference, Hanover, Germany, 2003 (PDF)
- *"The MAMI Query-By-Voice Experiment: Collecting and annotating vocal queries for music information retrieval"*, M. Lesaffre, K. Tanghe, G. Martens, D. Moelants, M. Leman, B. De Baets, H. De Meyer and J.-P. Martens, in: Proceedings of the International Conference on Music Information Retrieval, Baltimore, Maryland, USA and Library of Congress, Washington, DC, USA, October 26-30, 2003 (ISMIR 2003) (PDF)
- *"An Auditory Model Based Transcriber of Vocal Queries"*, T. De Mulder, J.P. Martens, M. Lesaffre, M. Leman, B. De Baets, H. De Meyer, in: Proceedings of the International Conference on Music Information Retrieval, Baltimore, Maryland, USA and Library of Congress, Washington, DC, USA, October 26-30, 2003 (ISMIR 2003) (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

Examples

Singing lyrics 010_030_EXP2_QbV1.wav	Mixed: percussion and singing lyrics 022_062_EXP1_QbV1.wav
Whistling 132_036_EXP2_QbV1.wav	Mixed: singing lyrics, whistling and percussion 074_073_EXP2_QbV1.wav
Humming 012_019_EXP3_hum.wav	Mixed: singing syllables and percussion 132_054_EXP2_QbV1.wav
Percussion 027_078_EXP1_QbV2.wav	Mixed: singing lyrics and comments 022_006_EXP1_QbV1.wav
“Good" query 052_058_EXP1_QbV1.wav	Mixed: singing lyrics and syllables 041_011_EXP2_QbV2.wav
“Bad" query 045_071_EXP2_QbV1.wav original	Mixed: comments and singing lyrics 052_067_EXP1_QbV1.wav