What Is Topical in Cultural Heritage: Content-based Retrieval Among Folksong Tunes (WITCHCRAFT)

1a Project Title

What Is Topical in Cultural Heritage: Content-based Retrieval Among Folksong Tunes

1b Project Acronym

WITCHCRAFT

1c Principal Investigator

dr. F. Wiering, Center for Content and Knowledge Engineering, Institute of Information and Computing Sciences, Utrecht University

2 Summary of research proposal

Music Information Retrieval (MIR) research has produced numerous methods and tools, but the design of complete, usable working systems is still in its infancy. The urgent need for such systems is strongly felt, in particular by cultural heritage institutions that possess large musical holdings, such as the Meertens Instituut. Therefore, the WITCHCRAFT project sets as its objective to develop a fully-functional content-based retrieval system for large amounts of folksong melodies stored as audio and notation, building on the best practices of MIR research. The system will be integrated in the *Nederlandse Liederenbank* ('Database of Dutch Songs') of the Meertens Instituut. Other participants are the Institute of Information and Computing Sciences (Utrecht University) and the Theater Instituut Nederland.

The principal issues to be addressed in the project are:

1. framework of the search engine;
2. preparation of a realistic test corpus;
3. transcription of audio data;
4. selection of relevant features for matching;
5. designing similarity measures;
6. designing efficient matching algorithms;
7. modelling user tasks and visualisation;
8. performance and usability evaluation;
9. integration in the *Nederlandse Liederenbank*.

We expect WITCHCRAFT to have an impact on MIR research by providing insight into the modelling of high-level musical features, by setting a standard to compare future systems against, and by creating tools, data and evaluation methods that can be used in other MIR research projects, and in the song databases of other cultural heritage institutions in the Netherlands and abroad.

3 Classification

3–Personalisation through presentation

4 Composition of the Research Team

	Name/function	Role/expertise
1. Vacancies	PhD student	
	postdoc	
	programmer	
2. Utrecht University		
Institute of Information and Computing Sciences	prof. dr. J van den Berg (CCKE)	promotor, information systems
	drs. M. Bosma (PhD student, GIVE)	music retrieval
	ir. R. van Gulik (PhD student, GIVE)	visualisation
	dr. H.J.M. (Tabachneck-) Schijf (CCKE)	cognition and retrieval
	dipl.-inf. R. Typke (PhD student, GIVE)	music retrieval
	dr. R.C. Veltkamp (GIVE)	multimedia algorithmics
	dr. F. Wiering (CCKE)	principal investigator, music retrieval
	dr. ir. R. van Zwol (CCKE)	retrieval evaluation
3. Cultural Heritage Institutions		
Meertens Instituut, KNAW	prof. dr. L.P. Grijp	promotor, musicological research
	drs. E. Brinkhuis	database, indexing
	drs. M. de Bruin	database management, *Nederlandse Liederenbank*
	M. van Dijk	data entry (2 years)
Theater Instituut Nederland	drs. P.J. Post	entertainment/popular song

Abbreviations:

CCKE Center for Content and Knowledge Engineering
GIVE Center for Geometry, Imaging and Virtual Environments

The researchers listed under **(1.)** will be employed by the ICS and positioned at the Meertens Instituut.

5 Research School

SIKS (Dutch Research School for Information and Knowledge Systems)

6 Description of the Proposed Research

6a Scientific aspects

Motivation

Music is known to every culture, and in each it plays a number of roles, for example in religious ceremonies, in entertainment, or as a sign of social identity. Music is therefore an important aspect of what UNESCO describes as the 'intangible cultural heritage' consisting of practices and expressions rather than objects. Access to the musical heritage has its own set of problems that have not been addressed as yet in the CATCH core projects.

Music's most important manifestation, sound generated during performance, is volatile. It can be captured, though imperfectly, in two ways, as sound recording and as music notation. Worldwide, huge amounts of music have been recorded and notated in attempts to document and preserve traditional culture.

In the Netherlands, Ate Doornbosch created a collection of national importance for his famous radio programme *Onder de Groene Linde* ('Under the Green Limetree') [Doornbosch 1987; Grijp and Roodenburg 2005]. Between 1957 and 1994, he has made field recordings of over 7000 folksongs, mainly narrative ballads. At the Meertens Instituut, an information system has been developed around this and other song collections, the *Nederlandse Liederenbank* ('Database of Dutch Songs'). The *Liederenbank* contains song metadata such as first line, title, refrain, author, and keywords, but one cannot query the musical content itself. This is a serious lacuna because melody identification and comparison forms the basis of much musicological and ethnological research.

Figure 1: Seven versions of the beginning of the song *Toen ik op Neerlands bergjes stond*

At this moment, the only way to identify the source of a melody—or to find variants of it—would be learning it by heart and then listening through thousands of recordings and singing through hundreds of notated folk music collections—including collections of theatre songs and other popular tunes. An important characteristic of folksong is that it is orally transmitted, without using music notation. Therefore, versions of a melody may vary considerably (see Figure 1), creating yet another challenge to the researcher.

This demonstrates the urgent need for a search tool to investigate melody collections. A fundamental requirement for it is to be designed for handling the variation due to oral transmission: it must focus on the essential features of a melody while attaching little importance to accidental changes. In the WITCHCRAFT project, we propose to make such search tasks possible by using automation techniques developed for *content-based* music information retrieval (MIR).

State of the art

Music information retrieval has only recently become a major research area, even though the concept originates in the 1960s [Byrd and Crawford 2000]. The principal research topics of MIR are shown in Figure 2.

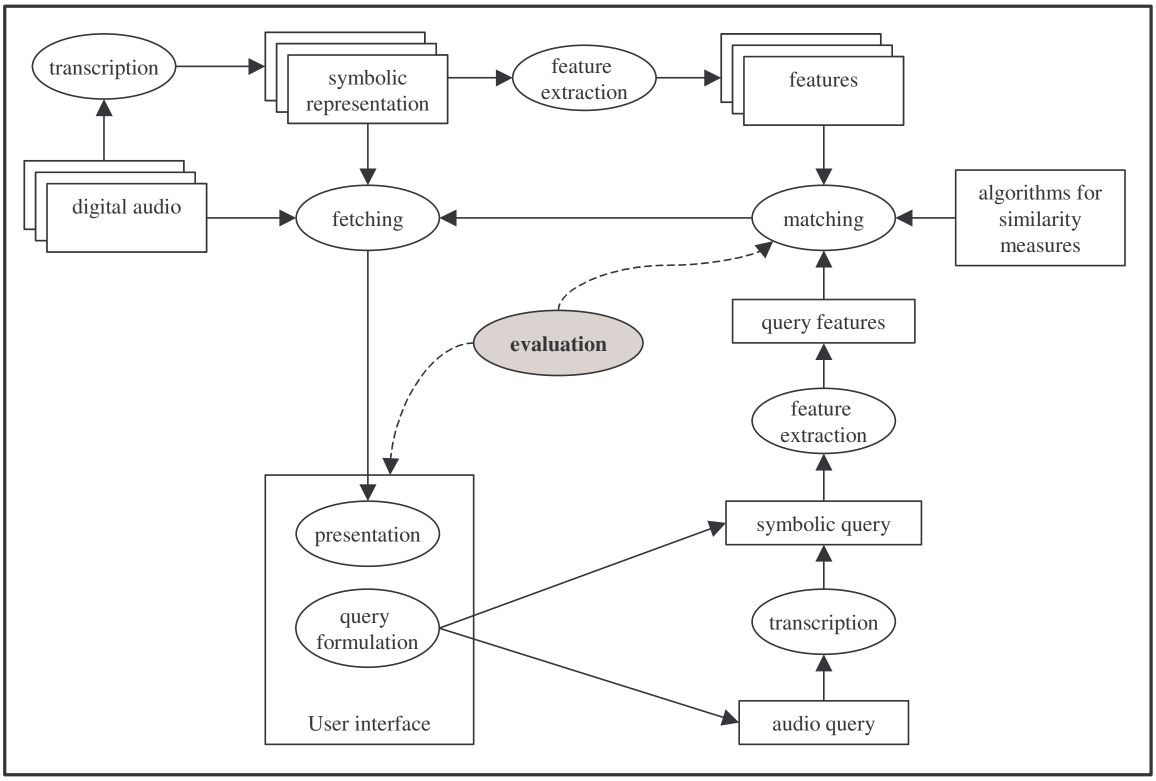

Figure 2: MIR framework, showing the principal research topics of MIR and their interrelation.

Early applications worked with encoded music notation, but currently, the focus is more on music's most universal format, audio (MP3 in particular). However, although simple audio features are suitable for audio fingerprinting (i.e. identifying *recordings*) and generic classification of style, they are not suited for modelling musical similarity at a detailed level, for example for identifying or tracing variants of melodies.

So-called 'query by humming' systems track pitch events in a hummed (sung, whistled) query and match these against a **corpus**[1] of melodies in a symbolic representation[2] [Pauws

[1] Terms in **bold** refer to the activities listed in the Approach section.

2002]. No large databases of symbolic representations derived from audio exist: all such databases have been created from notation or MIDI information.

Because of the above-mentioned limitations of ‘pure’ audio retrieval, the **transcription** of musical audio to a symbolic representation has become a focal area of MIR research. Polyphony (using simultaneous note events) remains very hard to transcribe, but transcription of monophony (unaccompanied melodies) is reasonably successful [Klapuri 2004]. One of the best transcribers of monophony, especially for song—where text is a disturbing factor—is the MAMI transcriber [De Mulder et al. 2004].

For *matching* the query to the items in a database of symbolic representations, the dominant approach used to be to compare monophonic pitch strings, using for example string matching techniques (as in Themefinder[3] [Kornstädt 1998]) or edit distances [Eliëns and Kersten 1999]. However, rhythm carries almost as much information as pitch [Byrd and Crawford 2000]. A number of techniques have been developed that take pitch *and* rhythm into account and can also handle multiple voices [Clausen 2000; Ukkonen et al. 2003].

Transportation distances—Earth Mover’s Distance (EMD) and Proportional Transportation Distance (PTD)—deal very well with local variations (ornamentation, interval and duration changes, see Figure 3), and are tolerant against pitch and rhythm errors [Typke et al. 2004a]. In addition to pitch and time, these distances employ the *weight* of the pitch event, which may depend on its duration, place in contour and/or other properties. The weight expresses the contribution of a pitch event to the ‘shape’ or ‘identity’ of a segment of music.

It is widely felt that musical **similarity measures**, which lie at the heart of MIR, should be **perceptually and cognitively** motivated [Byrd and Crawford 2000, Selfridge-Field 2004]. Properties like contours and phrases are particularly important in musical memory [Snyder 2000], and therefore in similarity judgments. The challenge is to develop methods for extracting these features from audio and notation and developing efficient **algorithms** for matching them.

MIR research has produced numerous methods and tools, but few complete, usable working systems exist.[4] Generally, these only match low-level features[5] or use a small data set. **User** needs [Lee and Downie 2004], the query formulation process [Lesaffre et al. 2003] and output visualisation [Van Gulik et al. 2004] have only recently begun to be studied. Most importantly, suitable **evaluation** methods are lacking, a matter which is now being addressed in a concerted, international effort [Downie et al. 2004] and a yearly competition (MIREX).[6]

---

[2] The information content of the music expressed in some encoding, for example MIDI or MusicXML.
[3] http://www.themefinder.org
[4] For a survey of MIR systems see http://www.mirsystems.info/
[5] Fingerprinting services such as MusicDNA (http://www.musicdna.nl) and SHAZAM (http://www.shazam.com) have large data sets but use only spectral features for identification of recordings.
[6] http://www.music-ir.org/mirexwiki/index.php/MIREX_2005

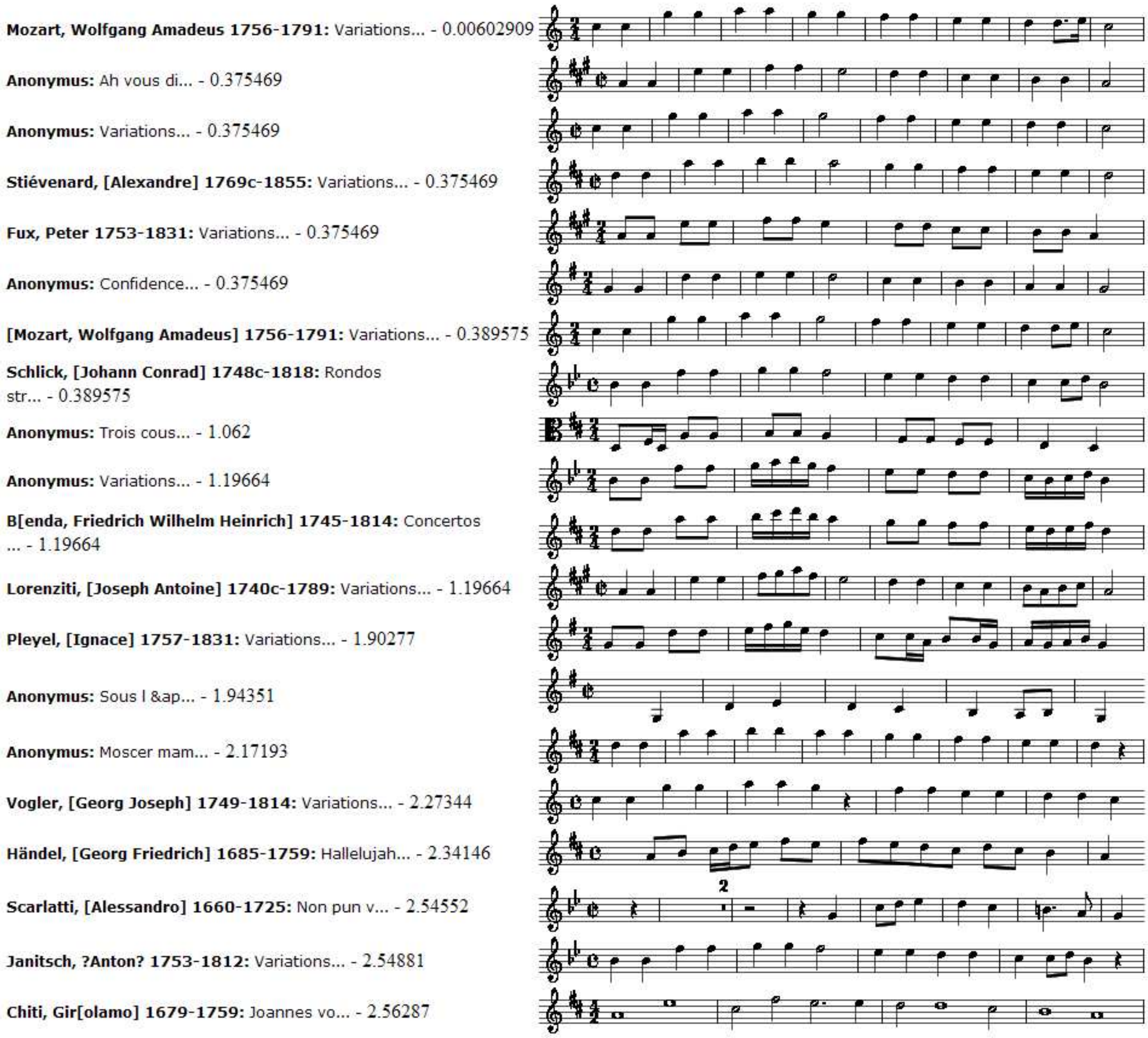

Figure 3: First 20 hits of an experimental PTD search on the RISM A/II database, using W.A. Mozart, *Ah, vous dirai-je maman* as a query. The query is the first hit. Notational duplicates have been removed.

Objectives

The WITCHCRAFT project sets as its objective to develop a fully-functional content-based retrieval system for folksong melodies stored as audio and/or notation, building on the best practices of MIR research. More precisely, the project aim is the design, implementation and evaluation of a melody search engine that:

- is capable of handling large amounts of audio and notation data;
- matches relevant low-level *and* high-level musical features using similarity measure(s) that are based on music cognition and perception and reflect the musical characteristics of the folksong repertoire, including oral transmission;
- orders search output (melodies or melody fragments) by musical similarity;
- is usable for both specialists and the general public.

This system's potential will be demonstrated by **integrating** it in the *Nederlandse Liederenbank* of the Meertens Instituut, and adding a selection of early 20th century entertainment/ popular songs from the collection of the Theater Instituut Nederland to test the strengths of the melody search engine in cross-repertoire research.

The computer science part of the WITCHCRAFT project will be executed by the ICS. In the field of MIR, the ICS is the leading institute in the Netherlands. In its Orpheus project,[7] methods are being researched for searching large quantities of music notation using similarity measures such as the EMD and PTD [Typke et al. 2003, Wiering 2004]. The performance of these measures compares favourably to other similarity measures [Typke et al. 2004b]. ICS's recently started C-minor project (part of the BSIK-3 MultimediaN consortium) focuses on musical similarity measures that have cognitive relevance. Close collaboration with partners abroad such as the Institute for Psychoacoustics and Electronic Music (IPEM) of Ghent University[8] will also enable us to build on tools and techniques developed elsewhere, in the WITCHCRAFT project.

Approach

The principal issues to be addressed in the project are:

1. **Framework.** Design and first implementation of a modular architecture for the search engine.
2. **Realistic test corpus.** A substantial test corpus that adequately mirrors the complexity of the material is an essential precondition for designing and evaluating an adequate retrieval system. It will consist of both audio and encoded notation, plus annotations to the audio needed for the transcription step [Lesaffre et al. 2004].
3. **Transcription.** The transcription module will be based on the MAMI-transcriber [De Mulder et al. 2004], which will be redesigned to accommodate the unschooled singing and low technical quality of old recordings. It will be tested against the manually prepared test corpus, and used both for transcribing the melodies in the database and the vocal user queries.
4. **Cognitive model.** On the basis of the substantial literature on music cognition, for example [Snyder 2000], high-level features need to be selected that are potentially relevant to the folksong repertory, and reflect user queries both from specialists and the general public. New methods must be developed for extracting these features from transcribed audio and notation.
5. **Similarity measures.** We will design similarity measures that take the above features into account, for example by integrating them in the weights for EMD and PTD. New similarity measures must also be able to handle errors introduced in the transcription process [Dannenberg and Hu 2004] and oral variation, and be suitable for indexing.
6. **Algorithms.** The design of computationally efficient solutions for the similarity measures. The algorithms must have provable properties that explain their behaviour, as is the case for those described in [Typke et al. 2004b].
7. **User model.** User tasks will be modelled for both specialists and the general public, resulting in a user-friendly query interface and a choice of presentation modes of the search results. Presentation modes to be researched include aligned lists of notated melodies (for specialists) and a navigable virtual space (for the general public).

---

7 http://www.cs.uu.nl/centers/give/multimedia/music
8 http://ipem.ugent.be/

8. **Evaluation.** The performance and usability of the system will be evaluated by end users and domain experts by means of quantitative experiments, using the TERS evaluation testbed [Van Zwol and Van Oostendorp 2004] and by participating in MIREX. [Typke et al. 2005] present a method for determining the ground truth for continuous measures of musical similarity and measuring system performance against it. Additional methods for rigorous evaluation will be created.
9. **Integration** of the search engine into the *Nederlandse Liederenbank*.

Deliverables

The WITCHCRAFT project will deliver:

- a PhD thesis;
- conference papers and journal articles (Digital Libraries, Music Retrieval, Multimedia, Musicology);
- software modules (transcription, search) that can be built into other systems;
- a prototype melody search engine that is integrated in the *Nederlandse Liederenbank*, with technical documentation;
- a test collection of folksongs, with guidelines for further digitisation of folksong collections;
- a workshop aimed at transferring and disseminating the acquired knowledge to relevant cultural heritage organisations.

We will attempt to host ISMIR (International Conference on Music Information Retrieval) in Utrecht in the autumn of 2009 as a conclusion to the WITCHCRAFT project.

6b Innovation

The proposed research is innovative for the following reasons:

1. It aims at a complete, working system for melody retrieval using advanced search methods on a large database. Current systems are experimental, use primitive search methods and/or a small data set.
2. Musical audio and notation will be integrated by creating a database of reliable symbolic transcriptions from audio. In current systems, only the query is transcribed from audio.
3. Knowledge from music perception and cognition will be used for designing similarity measures, that are based on high-level musical features rather than the low-level audio or notational features that have dominated MIR so far. These can be fine-tuned for the repertoire (and, in principle, for any repertoire).
4. New methods will be designed for visualisation of musical output, offering usable alternatives to the Google-style lists that are currently generally employed.
5. Performance and usability will be evaluated by general public end users and domain experts. Adequate methods will be created for this, enabling the rigorous evaluation of systems that has been lacking so far in MIR.

We expect the WITCHCRAFT to have an impact on MIR research by providing insight into the modelling of high-level musical features, by setting a standard to compare future systems against, and by creating tools, data and evaluation methods that can be used in other projects.

6c Relevance for cultural heritage

The Meertens Instituut feels that incorporating a melody search engine into the *Nederlandse Liederenbank* is an urgent matter. Access to the melodic content of recordings and notated music allows large-scale research into transmission, variation and popularity of folksongs that was previously unthinkable. The melody search engine will be used for the systematic identification of melodies from *Onder de Groene Linde* as well as for other repertories preserved in the Meertens Instituut. To realise this potential more fully, and in addition to hosting the requested 3 researchers, it will make a considerable investment in personnel (an average of 0,75 FTE, whereas CATCH requires a minimum of 0,25 FTE) for systems integration and for data entry for a test corpus of sufficient size and complexity (0,5 FTE, 2 years, if WITCHCRAFT is granted). Furthermore, it will initiate projects to digitise the entire content of complete song collections, in particular the newly-acquired collection of Harrie Franken (c. 15.000 songs). Research into the origin of folk melodies will be enhanced by linking them to collections of popular music from the early 20th century such as the one of the Theater Instituut Nederland (TIN).

The project results are relevant in the first place to other institutions that possess large collections of song data, which cooperate in the *Stichting Erfgoed van het Nederlandse Lied* ('Foundation for the Heritage of Dutch Songs'). Members include the Meertens Instituut, TIN, the Royal Library, the Library of the University of Amsterdam, and the Nederlands Muziek Instituut. An impact is also expected in Flanders, where the need for content-based access to the scattered folksong collections is similarly strongly felt, and in the numerous folksong digitisation projects elsewhere in Europe (e.g. in Finland,[9] Poland[10] and Slovenia[11]). In general, we expect that this project will generate a stimulus for new digitisation projects of musical content, as these are only viable insofar as an effective mechanism exists for providing access to this content.

Aspects of the project also have a potential for impact on other types of cultural heritage, in the following areas:

- transcription of speech corpora;
- algorithms and indexing methods for image matching;
- visualisation methods for non-textual materials;
- evaluation methodology for cultural heritage.

Melody is the most natural property for people to remember music by. Methods developed in the WITCHCRAFT project can therefore be generalised so that more tangible access to many repertoires within the world's musical heritage can be achieved.

[9] Digital Archive of Finnish Folk Tunes; http://www.jyu.fi/musica/sks/index_en.html
[10] ESAC, Polish Academy of Sciences; http://www.esac-data.org/
[11] Digital archive OSNP; http://lgm.fri.uni-lj.si/work.htm

7. Description of the proposed plan of work

The WITCHCRAFT engine will be developed in a series of iterations:

1. a simple framework;
2. a working search engine with sufficient test data;
3. a complete prototype;
4. the finished engine, integrated into the *Nederlandse Liederenbank*.

The numbered activities listed below are described in section 6a.

Phase	Activity	Team			
		PRG	OIO	PD	DE
Year 1, months 1-6					
	literature study		X	X	
	design and implement the framework (1)	X		X	
	selection of materials for the test corpus (2)		X		
	digitisation and annotation guidelines (2)		X		
Year 1, months 7-12					
	digitisation and annotation of test corpus (2)				X
	transcription tool (3)	X			
	cognitive model (4)		X	X	
Year 2, months 1-6					
	digitisation and annotation (2, continued)				X
	transcription tool (3, continued)	X			
	documentation transcription tool (3)	X			
	similarity measures (5)		X	X	
	algorithms (6)		X	X	
Year 2, months 7-12					
	digitisation and annotation (2, continued)				X
	algorithms (6, continued)		X		
	implementation of search module	X	X		
	design evaluation method (7)			X	
	test phase 1 (7)		X	X	
Year 3					
	digitisation and annotation (2, continued)				X
	iteration based on test phase 1: improve cognitive model, similarity measures, algorithms, and/or implementation	X	X	X	
	user model, visualisation (6)		X		
	test phase 2		X	X	
Year 4					
	iteration based on test phase 2	X	X		
	integration in *Nederlandse Liederenbank* (8)	X			
	technical documentation	X			
	user documentation	X			
	PhD dissertation		X		

Abbreviations:

PRG	Programmer
OIO	PhD Student
PD	Postdoc
DE	Data entry (Meertens Instituut employee)

The PhD student and the Postdoc often participate in the same tasks. The generic division of labour is as follows:

PhD student: methods of MIR;
Postdoc: framework and system integration, evaluation.

The tasks of the other project staff from the ICS and Meertens Instituut are not specified; staff members are involved in all activities specified above. In addition, their special responsibilities in year 4 are **dissemination** (organisation of workshop or ISMIR conference) and **user documentation**. Other publications will appear after the initial phase of the project: writing these is not listed as a special activity.

8. Literature

D. Byrd, T. Crawford (2000). Problems of Music Information Retrieval in the Real World. *Information Processing and Management* 38, 249-272

M. Clausen, R. Engelbrecht, D. Meyer, J. Schmitz (2000). A Web-based Tool for Searching in Polyphonic Music. *Proceedings ISMIR 2000*

R.B. Dannenberg, N. Hu (2004). Understanding Search Performance in Query-by-humming Systems. *Proceedings ISMIR 2004*, 232-237

T. De Mulder, J.P. Martens, M. Lesaffre, M. Leman, B. De Baets, H. De Meyer (2004). Recent Improvements of an Auditory Model Based Front-end for the Transcription of Vocal Queries. *Proceedings ICASSP 2004*

A. Doornbosch (1987). *Onder de Groene Linde: verhalende liederen uit de mondelinge overlevering*. Uniepers

J.S. Downie, J. Futrelle, D. Tcheng (2004). The International Retrieval Systems Evaluation Laboratory: Governance, Access and Security. *Proceedings ISMIR 2004*, 9-14

A. Eliëns, M. Kersten (1999). Musical Feature Detection in ACOI. Unpublished; http://www.cs.vu.nl/~eliens/research/@online/papers/midi/doc.ps

L.P. Grijp, H. Roodenburg (2005). *Alan Lomax en Ate Doornbosch. Twee muzikale veldwerkers*. Amsterdam University Press.

R. van Gulik, F. Vignoli, H. van de Wetering (2004). Mapping Music in the Palm of Your Hand, Explore and Discover Your Collection. *Proceedings ISMIR 2004*, 409-414

A.P. Klapuri (2004). Automatic Music Transcription As We Know It Today. *Journal of New Music Research* 33, 269-282

A. Kornstädt (1998). Themefinder: A Web-based Melodic Search Tool. *Computing in Musicology* 11, 231-236

J.H. Lee, J.S. Downie (2004). Survey of Music Information Needs, Uses, and Seeking Behaviour: Preliminary Findings. *Proceedings ISMIR 2004*, 441-446

M. Lesaffre, K. Tanghe, G. Martens, D. Moelants, M. Leman, B. De Baets, H. De Meyer, J.P. Martens (2003). The MAMI Query-by-voice Experiment: Collecting and Annotating Vocal Queries for Music Information Retrieval. *Proceedings ISMIR 2003*, 65-71

M. Lesaffre, M. Leman, B. De Baets, J.P. Martens (2004). Methodological Considerations Concerning Manual Annotation of Musical Audio in Function of Algorithm Development. *Proceedings ISMIR 2004*, 64-71

S. Pauws (2002). CubyHum: A Fully Operational Query by Humming System. *Proceedings ISMIR 2002*, 187-196

E. Selfridge-Field (2004). Towards a Measure of Cognitive Distance in Melodic Similarity. *Computing in Musicology* 13, 93-111

B. Snyder (2000). *Music and Memory*. MIT Press

R. Typke, C. Giannopoulos, R.C. Veltkamp, F. Wiering, R. van Oostrum (2003). Using Transportation Distances for Measuring Melodic Similarity. *Proceedings ISMIR 2003*, 107-114

R. Typke, F. Wiering, R.C. Veltkamp (2004a). A Search Method for Notated Polyphonic Music with Pitch and Tempo Fluctuations. *Proceedings ISMIR 2004*, 281-288

E. Ukkonen, K. Lemström, V. Mäkinen (2003). Geometric Algorithms for Transposition Invariant Content-Based Music Retrieval. *Proceedings ISMIR 2003*, 193-199

R. van Zwol, H. van Oostendorp (2004). Google's 'I'm feeling lucky', Truly a Gamble? *Proceedings WISE 2004*, 378-390

Key publications of the research team

Het Antwerps Liedboek (2004). Teksteditie door D.E. van der Poel, D. Geirnaert, H. Joldersma en J. Oosterman. Reconstructie van de melodieën door L.P. Grijp. Lannoo

J. Vleugels, R.C. Veltkamp (2002). Efficient Image Retrieval through Vantage Objects. *Pattern Recognition*, 35, 69-80. Received the Pattern Recognition Society Award for outstanding contribution to the journal *Pattern Recognition*

R. Typke, R.C. Veltkamp, F. Wiering (2004b). Searching Polyphonic Music Using Transportation Distances. *Proceedings ACM Multimedia 2004*, 128-135

R. Typke, M. den Hoed, J. de Nooijer, F. Wiering, R.C. Veltkamp (2005). A Ground Truth For Half A Million Musical Incipits, in *Proceedings of the Fifth Dutch-Belgian Information Retrieval Workshop*, 63-70

F. Wiering, R. Typke, R.C. Veltkamp (2004). Transportation Distances in Music Notation Retrieval. *Computing in Musicology* 13, 113-128