COGITCH
COgnition Guided Interoperability beTween Collections of musical Heritage
NWO-CATCH project 640.005.004

1a Project Title
COgnition Guided Interoperability beTween Collections of musical Heritage

1b Project Acronym
COGITCH\(^1\)

1c Principal Investigator
Prof. dr. Remco Veltkamp

2 Summary
Sound & Vision (S&V) possesses a unique collection of popular Dutch music. The Meertens Institute (MI) possesses a unique collection of Dutch folk songs. These two collections of musical heritage belong to the same culture, but are only separated for institutional reasons. S&V wishes to make these musical archives accessible in an integrated way for the general public. MI wishes the same, to enable musicological research on song evolution. Driven by these demands, COGITCH’s general objective is to develop generic techniques to index distributed sources by developing an interoperable system. In a collaborative research, cutting across the boundaries between music cognition and computer science, we develop generic techniques for relating music from different collections. In developing retrieval methods, we will take a top-down approach, working from musical knowledge and cognitive psychology towards the identification and processing of audio features. On-line annotations provided by listeners will support establishing the relationships between ‘hooks’ (perceptually salient musical patterns) and music.

COGITCH involves three intertwined strands of research objectives:
- We design and develop a novel infrastructure to let listeners collectively provide annotations and to derive cognitively relevant features.
- Based on these cognitively relevant features we invent and implement new music thumbnail extractors and music similarity methods.
- Using content-based retrieval methods, a generic interoperable search infrastructure is designed and implemented to access the collections of S&V and MI.

The practical results of COGITCH are an interoperable search infrastructure and a workflow for music thumbnails extraction. These are beneficial to S&V and MI, and generally to institutions and industry that preserve collections of musical audio. The scientific results are models of music cognition, cognition based similarity measures, and ground truth data. These

\(^1\) A cognitive itch refers to an ‘earworm’, a fragment of music that you can’t get out of your head.
enable future music cognition research, musicological research, and music retrieval benchmarking.

3 Classification
I. Interoperability of large and distributed sources

4 Composition of the Research Team

The Multimedia group of the Institute of Information and Computing Sciences at Utrecht University will be denoted UU. Its research focuses on the development of algorithms for content management and personalized usage. We address the big challenge in multimedia research for the next ten years: the handling of multimedia information that is perceptually and semantically relevant in a way that is guaranteed to be effective, robust, and efficient. Characteristic for our approach is the emphasis on provable and rigorous properties of algorithms, and close cooperation with domain experts. UU organized the International Conference on Music Information Retrieval (ISMIR) in 2010, and hosts the prestigious VIDI project Modelling musical similarity over time through the variation principle by Anja Volk. The research quality of the group was rated as ‘excellent’ by the latest national Computer Science research assessment in 2010.

The Music Cognition Group at the Institute for Logic, Language and Computation at University of Amsterdam (www.hum.uva.nl/mmm) will be denoted UvA. The proposed research will take advantage of its long-standing expertise in the computational modelling of music cognition, supported by, e.g., the prestigious NWO PIONIER grant (see final report: [19]). Currently, the UvA group is supported by grants from the Netherlands Organisation for Scientific Research (NWO ‘Foundations of the Humanities’ programme) and the European Commission (Sixth Framework IST programme) both in the field of music cognition. The aim is to arrive at a cognitive science of music—bridging expertise from the humanities and the sciences—with a special focus on its temporal aspects, such as rhythm, tempo and timing. The research quality of the group was rated as ‘excellent’ by the latest national (VSNU) Computer Science research assessment in 2010.

The Meertens Institute is to study diversity in Dutch language and culture. The Meertens Institute will be denoted MI. The ‘DOC Lied’ is MI’s expertise centre for Dutch songs. Documentation is realized by means of the Dutch Song Database (Nederlandse Liederenbank, www.liederenbank.nl), a database that gives access to more than 125,000 songs, from the Middle Ages to the modern times, including the oral tradition. Encoded music notation is gradually being added to the Dutch Song Database. An important next step is to add recordings of popular music. MI’s research comprises several aspects of Dutch song culture in past and present, like the mechanism of oral tradition, the contrafact, street songs and tearjerkers. The activities of the DOC Lied also include teaching, realized through the special chair of Dutch Song Culture at Utrecht University.

Sound & Vision (Nederlands Instituut voor Beeld en Geluid) will be denoted S&V. It collects, preserves, and makes accessible the audiovisual cultural heritage of the Netherlands. It acts as the National Music Depot. Its holdings include 250,000 hours of music, among which are valuable collections of old gramophone recordings and unique radio broadcasts. These contain a significant number of Dutch popular songs from the first half of the 20th century. The recordings are catalogued, but there is no advanced access or enriched metadata.
Radio 5 Nostalgia is a Dutch public-service network radio station operated by NPO. It aims at broadcasting easy listening music, and listener participation (games, etc.). The target audience matches the intended audience of the MI and S&V collections. Radio 5 Nostalgia thus is an ideal portal for obtaining on-line annotations provided by listeners, which will support establishing the relationships between ‘hooks’ (perceptually salient musical patterns) and music.

In the tables below, the “Role” in the project listed is further explained in section 7.

### 4.1 Academic partners

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Expertise</th>
<th>Role</th>
<th>Financed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prof. Remco Veltkamp</td>
<td>Utrecht University (UU)</td>
<td>Computer Science (algorithmics, multimedia)</td>
<td>Supervision PhD student, programmer</td>
<td>UU</td>
</tr>
<tr>
<td>Prof. Henkjan Honing</td>
<td>University of Amsterdam (UvA)</td>
<td>Humanities (music cognition, computational humanities)</td>
<td>Supervision postdoc</td>
<td>UvA/ KNAW</td>
</tr>
<tr>
<td>Dr. Frans Wiering</td>
<td>UU</td>
<td>Computer Science (musicology, Music Information Retrieval)</td>
<td>Music information retrieval</td>
<td>UU</td>
</tr>
<tr>
<td>Dr. Anja Volk</td>
<td>UU</td>
<td>Computer Science (musicology, Music Information Retrieval)</td>
<td>Music variation modeling</td>
<td>NWO (VIDI)</td>
</tr>
</tbody>
</table>

### 4.2 Heritage institutions

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Expertise</th>
<th>Role</th>
<th>Financed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prof. Louis Grijp</td>
<td>Meertens Institute (MI)</td>
<td>Humanities/Cultural Heritage (musicology, folk songs)</td>
<td>Musicology knowledge</td>
<td>MI</td>
</tr>
<tr>
<td>Dr. Peter van Kranenburg</td>
<td>MI</td>
<td>Humanities/Cultural Heritage (Music Information Retrieval, folk songs)</td>
<td>Folk song researcher</td>
<td>MI</td>
</tr>
<tr>
<td>MSc. Martine de Bruin</td>
<td>MI</td>
<td>Cultural Heritage (music collections, ICT development)</td>
<td>Dutch Song Database manager</td>
<td>MI</td>
</tr>
<tr>
<td>MSc. Johan Oomen</td>
<td>Sound&amp;Vision (S&amp;V)</td>
<td>Cultural Heritage (music collections)</td>
<td>Crowd sourcing</td>
<td>S&amp;V</td>
</tr>
<tr>
<td>MSc. Maarten Brinkerink</td>
<td>S&amp;V</td>
<td>Cultural Heritage (music collections)</td>
<td>Legal aspects, frontend</td>
<td>S&amp;V</td>
</tr>
<tr>
<td>Esther Herder</td>
<td>Netherlands Public Broadcasting (NPO)</td>
<td>Social media</td>
<td>Site manager Radio 5 Nostalgia</td>
<td>NPO</td>
</tr>
</tbody>
</table>
### 4.3 Vacancies

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Expertise</th>
<th>Role</th>
<th>Financed</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN PhD student</td>
<td>UU, Meertens Institute, Sound&amp;Vision</td>
<td>Computer Science (pattern recognition)</td>
<td>Segmenting, similarity measures</td>
<td>NWO</td>
</tr>
<tr>
<td>NN Programmer</td>
<td>UU, Meertens Institute, Sound&amp;Vision</td>
<td>Computer Science (audio processing, distributed systems)</td>
<td>Development annotation tool and search framework</td>
<td>NWO</td>
</tr>
<tr>
<td>Dr. NN Postdoc</td>
<td>UvA, Meertens Institute, Sound&amp;Vision</td>
<td>Humanities (music cognition)</td>
<td>Cognitive hook model</td>
<td>NWO</td>
</tr>
</tbody>
</table>

### 5 Research School

We participate in the research schools ASCI (Advanced School for Computing and Imaging) and SIKS (School for Information and Knowledge Systems).
6 Description of the Proposed Research

6.a Scientific aspects

6.a.1 Objectives

COGITICH’s general objective is to develop generic techniques to index distributed music sources by developing an interoperable system. In a collaborative research, crossing the boundaries between music cognition and computer science, we develop generic techniques for relating music from different collections. The recordings at S&V are catalogued, but there is no advanced access or enriched metadata. In order to make the archives accessible, the recordings must be annotated with semantic metadata. We will develop new techniques allowing to locate and retrieve music based on musical features that are known to be relevant to the intended users. We will do this by letting listeners annotate music on-line, and then deriving cognitive features from their feedback.

The specific case that will be studied in order to reach this general objective involves two Cultural Heritage institutions, MI and S&V. MI is lacking necessary access from the Dutch Song Database to the recordings of S&V, which is needed to do research into musical cultural heritage. An example research question is: how was music replaced by popular music? In the early 20th century, folk singers increasingly used existing ‘modern’ popular melodies. For example, melodies from Willy Derby’s popular songs (in S&V’s collection) were used for songs about murders and other disasters by traditional singers (in MI’s collection). To investigate such relations, advanced content-based search methods are needed.

Online music is an important part of digital culture. S&V and MI possess two unique collections of Dutch cultural heritage that must be made interoperable and accessible. We will design and implement the interoperability between these collections. We address the above challenges by pursuing the following three intertwined strands of research objectives:

1. We will design and develop a novel infrastructure to let listeners collectively provide annotations and to derive cognitively relevant features. Research questions are:
   (i) What is an effective framework to let people perform annotations? Should they explicitly indicate relevant features, or must they be derived from the user actions?
   (ii) Which are the candidate ‘hooks’ of a melody, i.e. the fragment of the melody most people remember, or will start singing when asked to do so. As such, a ‘hook’ can be considered the ‘essence’ of a song, and might facilitate search in a large database of songs.

2. Since a ‘hook’ is composed of melodic as well as rhythmic information, both need to be captured in a model to be able to use it for annotation and search. From these derived cognitively relevant features (e.g. the ‘hook’, melodic and rhythmic similarity), we will invent content-based retrieval methods for musical audio. These will be designed such that they have necessary properties such as robustness and invariance. Implementation and rigorous evaluation will experimentally verify them. Research aspects are:
   (iii) How can we map the cognitively relevant features into computational models? How can existing measures of melodic similarity be enhanced with a measure of rhythmic similarity?
The invention of similarity measures for these features, and the design, development and evaluation of efficient and robust algorithms to compute these similarity features for large collections.

3. Using these content-based retrieval methods, an interoperable search infrastructure will be designed and implemented to access the collections of MI and S&V. Research aspects are:

(v) What is the best strategy to store the features, such that consistency is guaranteed, and search from the Dutch Song Database into the S&V archives is efficient?

(vi) How to aggregate the retrieval results from the various features?

6.a.2 Related work

Annotation and audio feature extraction are preliminary steps that precede the central modelling tasks. Two distinctions are important:

1. Monophonic vs. polyphonic music, i.e. one versus many voices.
2. Audio vs. symbolic musical data: the challenge is to be able to connect the advantages of both representations.

Computational musicology and cognitive modelling

In computational musicology, much attention has been given to traditional music, the Essen folksong collection in particular [42,44]. Traditional music is generally learned by listening and participating, not from notation, leading to considerable ‘oral’ variation between versions of a melody. This variation depends in turn on the properties of the human musical memory [26]. Therefore computational models of folksong variation must be informed by music cognition [45]. Conversely, computational modelling contributes to understanding music perception and cognition by formalizing the mental processes involved in listening to music [19].

Our idea to focus on the ‘hook’ of popular songs in order to improve the state-of-the-art music annotation and MIR methods is novel [30,37]. While the term is common among musicians and musicologists [7], what precisely constitutes a hook and what makes it stick in your mind is unknown [32,96] and how structural features of music contribute to the hook is an open question [22,35,41]. Insights from cognitive science about the way people remember a melody and recall it from memory (e.g. [4,51]) can help in creating more intuitive search engines for music.

Missing: There has been a body of music cognition models developed, but not of the kind we need in this project: cognitive and computational models for musical hooks.

Music annotation

Several experiments have been done with online collective tagging of musical audio [47,54,77]. There is hardly any research to link these tags to measurable musical properties [68]. Also, collectively created tags have not been linked to specific locations or ranges within the example. For the WITCHCRAFT project, a method for annotation of music at the phrase level was designed and used for evaluating classification features [87].
Missing: The current state-of-the-art lacks methods for collective online annotation of musical audio that allow (i) tagging of specific locations or ranges, and (ii) the annotations to be linked to measurable musical properties.

Audio feature extraction

Effective systems only exist for the transcription of monophonic audio into a symbolic representation [18,34]. Transcription of polyphonic audio has recently made significant advances, especially by reducing the problem to one of finding melody, bass, and intermediate chords. Even for the most advanced methods [55,67], the reported error rates are still considerable. Chroma features [10] offer an effective way of dealing with the pitch content of musical audio without creating an actual transcription. Chroma can be efficiently searched [9] and has proven to be an important feature for modelling audio similarity [61] and for aligning audio and notation [60].

Several important toolboxes for processing musical audio exist, for example Tzanetakis’ Marsyas [82] and Lartillot’s MIRtoolbox [46].

Missing: Research has been focussed on the notated music and music audio separately. What we need for this project is research to integrate audio-symbolic searching.

Segmentation

Complete pieces of music must be subdivided into manageable units before they can be used in a retrieval application. For COGITCH, two segmentation tasks are particularly important: separation of melody from accompaniment and division of music in units at the phrase level and below. In symbolic representation, several voice separation algorithms have been proposed, yet considerable improvement is possible [91]. In audio, the focus is rather on stream separation as a preliminary step towards polyphonic transcription [52], and on separating larger temporal units such as verse and chorus [60,65].

Missing: Voice separation algorithms are not yet sufficiently effective, and methods still need to be invented for segmenting audio in perceptually relevant units at phrase and subphrase level.

Melody

Traditionally, melody has been modelled using string-based representations [43]. More recently, geometric representations have been introduced [48,59]. A geometric similarity measure developed at UU [81] gave the best results in the MIREX music similarity evaluation [57,80] on melodic similarity (used in http://yahmuugle.cs.uu.nl). In the CATCH-funded project WITCHCRAFT [92], effective sequence alignment algorithms have been developed that incorporate musical knowledge [45] (used in http://www.liederenbank.nl/index.php?wc=true). Much ongoing research in musical memory is about the salience of melodic features [1]. Discovering salient patterns is considered an open problem in MIR [17].

Missing: Cognition research has not yet resulted in computational similarity measures that model salience and stability of melodic features, which we need in the COGITCH project.

Harmony

The abundant research on models of musical harmony [14,56] has produced only few harmonic similarity measures [66]. UU researchers developed two measures for matching
sequences in music encodings, one based on a cognitive [27,29] and one of a grammar [28] model of harmony.

**Missing:** Similarity measures for harmony retrieval need to be extended to be able to retrieve harmonic sequences from audio.

**Rhythm**

Inference of beat and rhythmic patterns detection are well-studied areas [16,31]. Inner Metric Analysis [23,84,85] derives from a hierarchy of such patterns a model of the metrical structure of a piece of music that can be used as a similarity measure [86]. Humans can quickly and accurately interpret musical rhythmic structure, and can do so very flexibly; for example, they can easily distinguish between rhythmic, tempo and timing changes [9, 19]. Rhythm perception can be viewed as interplay between top-down (memory-based, schematic) and bottom-up (data-oriented, perception-based) processes [34]. Rhythm contributes strongly to the identity of a melody [21,69].

**Missing:** What is still missing is a formalized and evaluated cognitively motivated model of musical time that makes explicit the distinction between the discrete (categorical) and the continuous (expressive) aspects of musical rhythm.

**Music information retrieval systems**

Most music information retrieval (MIR) services on the Internet are based on specialist (pandora.com) or collective annotation of audio (last.fm). Shazam and similar services use audio content to enable the precise identification of recordings, not compositions [89]. Musical similarity occurs at many levels, from the global similarity of genres to the specific similarity of sections from the same piece. Therefore it makes sense for a system to contain various similarity measures, to let these users interactively choose some of these and to aggregate the rankings from their outputs in one presentation. While there is some research in this field [95], this is an important research direction.

**Missing:** The current state-of-the-art does not offer effective systems for combining content-based retrieval of music notation and musical audio, using aggregation of multiple feature retrieval results.

**Evaluation**

MIREX (Music Information Retrieval Evaluation eXchange, http://www.music-ir.org/mirex/) has become the most important platform for the performance evaluation of MIR methods, and it is now common to benchmark MIR methods against human ground truths, whether provided by experts or novices [58,81]. Several MIREX tracks are relevant to COGITCH, for example Audio and Notation Melody Retrieval, and Structural Segmentation. There is no MIREX track yet for discovery of salient patterns, such as hooks.

**Missing:** What we miss for our project is a MIREX track on pattern discovery in music.

**Overall conclusion:** The above issues is partly addressed in ongoing projects of UU, UvA and MI, from which COGITCh will benefit. Pattern recognition in the music domain has already been done a lot, but pattern detection, e.g. hooks, and the matching on the basis of hooks is innovative and would be a scientific breakthrough.
6.a.3 Approach

Data-oriented vs cognition-based approach

The widespread availability of the Internet and the development of efficient compression techniques for audio encoding in the last ten years gave an enormous boost to the discipline of Music Information Retrieval (MIR) [10,20]. A common approach in MIR is to use information-theoretic models to extract information from the musical data using advanced machine learning techniques. Overall, this approach is based on the assumption that all relevant information is present in the data and that it can, in principle, be extracted from that data (data-oriented approach).

Several alternatives have been proposed, such as models based on perception-based signal processing [3] or mimetic and gesture-based queries [50]. However, with regard to the cognitive aspects of music information retrieval (the perspective of the listener), some information might be implicit or not present at all in the data [30,93], yet needed in the design of similarity measures 3,50]. Elaborating state-of-the-art MIR techniques with recent findings from music cognition seems therefore a natural next step in improving search engines for music and audio (cognition-based approach).

In the current project we propose to develop a cognition-based search engine based on the most salient, memorable, and easy to recall moment of a musical phrase or song, the so-called ‘hook’, in an attempt to identify which cognitively relevant musical features affect the appreciation, memorization and recall of music.

The ‘hook’ of a song

The availability of a model of a hook will help in addressing a key problem for musicologists: trying to relate versions of songs that have been transformed over the years [26]. For this one needs to understand which cognitively relevant musical features affect the appreciation and memorization of music. Furthermore, one needs large amounts of semantically annotated musical data. We propose to obtain such data by collecting information from large numbers of listeners via a web-based environment where listeners are encouraged to mark specific locations in an actual recording, locations where s/he experienced something special or that s/he considers musically striking or intriguing. Large quantities of annotated musical data will allow for evaluating the explanatory power of cognitive models of melody and rhythm perception, insights about the way people remember a melody and recall such a melody from memory [4,41].

Note that this approach is not just relevant for folk song, but for (popular) music in general. The conceptual and scientific results are applicable to a wide range of musicological domains.

Objective 1: Cognitive modelling

A web-based environment, so-called ITCH environment (Identification, Tagging and Characterisation of Hooks) will be designed and constructed to obtain large amounts of judgements from the lay audience on what makes a fragment of music easy recognisable and/or stick in one’s mind. This will allow for evaluating—in an empirical and controlled way—the explanatory power of cognitive models of melody and rhythm perception in their prediction of what structural (e.g., pitch, key, rhythm, meter) and non-structural (e.g., associative, emotional, cultural) aspects of a melody play a role in the memorization, recall and appreciation of music. Part of the research is what the exact functionality of the ITCH environment will be.
In an initial stage, volunteers (master students, PhD students in master classes) will produce an initial test corpus, with annotations and classification into melody norms (classes of variations of the same song) on which we can start developing our cognitive modelling, and the segmentation, extraction, and matching algorithms. Within the WITCHCRAFT project, this approach has proven to be successful. In the subsequent stage, the potential of social-tagging or crowd annotation will be explored using co-called “games with a purpose” [2,88]. From the annotations, a cognitive hook model will be extracted, see the figure.

The attractiveness of this approach is that no ‘ground-truth’ is necessary. (We, in fact, want to reveal the ground-truth: what makes a fragment of music a potential hook? So starting with a ground-truth has the risk of becoming circular). Crowdsourcing has proven to be very useful in other initiatives in the cultural heritage domain, tapping into what Shirky calls our ‘cognitive surplus’ [71] to tag for example art [76] and photographs [13,74]. In the audiovisual domain, S&V has created a game that is used to collaboratively annotate videos [62]. We will leverage the user community of Radio 5 Nostalgia. This guarantees reaching a large audience of potential annotators. We will ensure copyright issues are dealt with. In addition, such ‘games’ will promote the project and the scientific problem at hand to a wider audience that participates in an active, natural and engaging way.

**Objective 2: Computational modelling**
The experimental research into perceptual and cognitive musical features and their relationship to audio features leads to a computational model for musically salient fragments, or ‘hooks’ [7], in the following way. From the audio signal, features (such as chroma) are derived, in order to obtain an abstract representation of pitch information [60] yielding a symbolic transcription of the audio stream. Next, we will segment the transcription into meaningful fragments, which correspond to the listeners’ annotations. The hook model and the features per segment are combined into a computation hook feature model, see figure below.
In order to compare two songs, two such segmented transcriptions are aligned. On the basis of the best alignment, an appropriate similarity score is computed. The research questions here are how to best segment, how to define a mathematical similarity function between two segments based on the hook features, and the invention of algorithms how to compute the similarity over a large collection of songs. This will be used in the content based search infrastructure below.

**Objective 3: Infrastructure**

A cognition-based set of thumbnails is extracted from both collections, that forms a computational model of the ‘hooks’ that make people memorise music. In addition to the web-based infrastructure to collect annotation data, we will develop an interoperable retrieval framework. Using this framework we will build an interface for expert users (folksong researchers) that allows searching between the MI Dutch Song Database and the S&V archive. Using the same framework, public users are allowed to search in both the Dutch Song Database and the S&V archives.

**Summary**

The above approach sketches our line of research, and we have indicated why we believe this will be successful. Here we summarise the open research issues that need to be handled during the project.

**Objective 1 (blue blocks):**
- The design of a listener annotation environment (the ITCH environment) using controlled lab-based experiments and crowd annotation using social media.
- Identifying the elementary audio ‘hooks’ from the obtained metadata identifying the cognitive features that contribute to the essence of a recalled melodic fragment (the ‘hook’ model).

**Objective 2 (green blocks):**
- The modelling of salient patterns of symbolic features corresponding to these fragments, in terms of harmony, rhythm, and melody.
- The design of similarity measures that allow similarity search on these features.

**Objective 3 (red blocks):**
- The generation of thumbnails for both collections.
- The design of a networked framework to make the collections of MI and S&V interoperable.
6.b Multidisciplinary cooperation

The project offers a unique possibility to compare the results of these experiments with musicological analysis of large groups of similar melodies. Will the ‘hooks’ indicated in the experiments correspond with the ‘motifs’ which the musicologist finds as most stable elements within one group of similar melodies transmitted in oral tradition, i.e. from mouth to ear? If this were the case, this would prove both the cognitive relevance of the musical motifs and the musical/musicological relevance of the cognitive hooks.

COGITCH’s objectives require multidisciplinary expertise. Each of the consortium partners contributes essential data, expertise, and skills, which are used by others. The proposed research is securely founded upon previous research by the partners. In the area of Music Information Retrieval, the group at UU is the leading institute in the Netherlands. An important contribution of this group is the design and evaluation of similarity measures for notated music that have provable computational properties and moreover model the human perception of music similarity [48]. UU and MI collaborated in the WITCHCRAFT project in designing a melody search engine for the Nederlandse Liederenbank [29,52]. An annotation method for better understanding of ground truth data was established as part of the project. UU and MI also collaborate in the MultimediaN-project C-MINOR, which has delivered components for music search engines and measures for harmonic similarity that will be further developed in this project [16,18].

The Music Cognition Group at UvA has an outstanding record in music cognition. In particular, temporal features such as onset detection in singing [14] and rhythmic expectancy [44] are studied, often by Web-based listening experiments [24,25]. This work will form the basis for the modelling of ‘hooks’ and the music thumbnail extraction.

S&V participates in the projects PrestoPRIME, COMMIT and AXES, in which techniques are developed to support the annotation of audiovisual materials [38]. These techniques will contribute to the creation of COGITCH’s annotation infrastructure.

The researchers financed by CATCH will work at least 60% of the time meaning within one of the cultural heritage partners MI and S&V, so that they have access to the relevant collections, data, and systems.

The contribution from the partners, not financed by CATCH, is as follows.

- MI provides access to the Dutch Song Database data folksong encodings and recordings, and provides domain knowledge of folk songs and popular music. Researchers provide musicological research questions, and contribute to annotation tests and evaluations. The personnel contribution of MI is about 0.25 fte per year.
- S&V digitises collections of musical audio, gives access to catalogue data, and contributes domain knowledge of musical recordings. Together with MI they will identify the collections that must be linked to the Dutch Song Database. They also provide expertise on copyright issues. The personnel contribution of S&V is about 0.15 fte per year.
- UvA contributes expertise in music cognition, particularly knowledge about temporal aspects of music. Together with MI, it provides research questions about the relationship between listening experience and musical structure, and contributes to the annotation tests.
- UU provides expertise in Music Information Retrieval, and hosts the server infrastructure for annotation and searching. Together with UvA, the cognitively relevant features will be mapped into computational models.
NPO is a radio broadcasting channel that provides the social media platform to effectively perform the crowdsourcing, building on its expertise from earlier, more general initiatives.

6.c Relevance

6.c.1 Scientific merits and innovation

Cultural Heritage
For the first time, the collections of MI and S&V will be interoperable and made accessible in combined form to both the general public and music researchers. In doing so, a bridge will be built between the traditionally separated domains of folk song and popular music. The retrieval application will allow content-based access to both collections. For the song documentation at the MI the most needed functionality is to be able to identify songs with an unknown origin. Access to audio collections will enable research into the relationship between popular music and folksongs. As all methods and tools will be designed to be generic, folksong and traditional music research elsewhere, and musical audio research in general, will be able to benefit from them.

Humanities/ Musicology
The new music search engine and the combined access to the collections of MI and S&V will greatly facilitate solving research questions at the intersection of folk and popular music, such as the ones described in 6.a.1. Thus both, ethnomusicology and popular music studies will take advantage of the project results. It will help to blur the strict boundaries between folk and popular music, in analogy with recent insights in ethnology that deconstruct the traditional antithesis of folk and popular culture.

Another innovation concerns the perceptual and cognitive aspects of music, which are still little understood. With the emerging Web 2.0, some tagging games for music have been presented, similar to the ITCH environment we propose. Unique in our project is that the resulting annotations are used to derive cognitive features of musical ‘hooks’, which are linked to music similarity measures. An explicit aim of the ITCH infrastructure is to be usable in different contexts, including the Internet. Furthermore, the empirical data obtained will form a solid starting point for an anticipated research project in cognitive science that will study what could explain that some melodies behave like ‘earworms’ and others don’t, that is, what music structural aspects makes these melodies spontaneously appear in one’s mind.

Computer Science
COGITITCH will contribute scientific methods that connect the human perception of music to processing of musical audio, notably by employing chroma features in the matching of harmonic patterns. Within multimedia, the branch of Music Information Retrieval has been dominated by statistical pattern recognition based on low-level features. In contrast, we employ a top-down approach exploiting musical knowledge, which is necessary in applications where musical meaning is relevant [10,90]. The retrieval methods will provide experimental content-based access to audio, supporting the aim of improving the ‘findability’ of the music. This will open up the collection to novel forms of cultural expression, and to studies by music scholars.
6.c.2 Timing

Our research in cognitive music modelling and music retrieval in the area of folksong during the WITCHCRAFT project has made clear that the next step should be integrating insights from music cognition [38,44,81,86,87]. The development of domain-specific music retrieval that is perceptually relevant and cognition-based, is an emerging research field. UU, UvA and MI currently have a strong position in the international arena, but there is international competition. It is necessary to implement the lines of research as outlined by us [10,37,44,92] to stay at the frontline of music retrieval through large-scale research efforts. COGITCH will benefit from the synergy with two new projects at UU, the VIDI project on modelling musical similarity (Anja Volk) and the COMMIT WP ‘Sensing emotion in music’.

COGITCH will contribute to the upcoming field of Computational Humanities, for which the KNAW finances an ambitious research programme. A major aim of the programme is to automatically detect high-level, novel patterns and concepts in historical, musical, textual and artistic data that are (practically) impossible to find by hand [94]. MI has submitted a proposal in the KNAW programme for the study of oral transmission of songs and tales. COGITCH, focusing on detecting hooks as high-level patterns in music, will complement and reinforce this proposed project through music retrieval methods and infrastructure.

Both in the Netherlands and Flanders, music is increasingly recognized as an important form of cultural heritage [5]. MI and S&V participate in the Steering Group of the Nederlands Muziek Instituut to establish a national portal for integrated access to musical heritage. MI has a strong tradition in folk song research, rooted in the former Dutch Folk Song Archive and continued in the present Dutch Song Database. MI has an urgent need to connect traditional and popular music for scholarly purposes and for the general public. In the current NWO Investment project Dutch Songs On–Line, song texts are digitized and their searchability is enhanced. COGITCH will greatly contribute to the next goal of MI, to enhance the searchability of song melodies, especially audio recordings, for which no acceptable content-based solution has been found.

For S&V, the ITCH infrastructure will support the aim of extending the music catalogue into a knowledge base about the musical holdings. This is an important part of the general strategic goal to develop a digital infrastructure for durable and efficient preservation and access. Until now, most effort has been spent into making collections other than music accessible. Without the COGITCH project, the accessibility of music collections will stay behind.

6.d Research utilisation

The research output will be utilized in the following ways:

- The interoperability framework to access S&V’s and MI’s collections will be made available for the general public. It will be a powerful addition to the existing online offerings of both heritage institutions.
- The access framework will be tailored towards musicological research, enabling experts to find relations between folk songs and popular songs. This involves expert requirement elicitation and designing an expert user interface.
- The annotation framework will be utilized by Radio 5 Nostalgia.
- The hook detection methods will be implemented in a workflow for creating meaningful musical thumbnails, integrated in the current service of S&V to provide rich metadata to music retailers through Phononet. It reduces a minute long
composition to an easily recognisable fragment of music. A cognition-based thumbnail generator will have a wide impact on the music industry. We will also demonstrate the technology to commercial entities, such as Cloudspeakers and Genolabs.

- The heritage partners will host workshops (2nd and 3rd year of the project) that will show the approach to other institutions. The collaboration with the Netherlands Music Institute will be forged, as this institution is currently leading an initiative to provide integrated search over multiple collections.

We will apply for a separate grant to integrate the thumbnail generation in an operational B2B service. We will also let a web developer design a visually appealing and web-scale front end on top of the annotation infrastructure. It will include social media support that will allow taggers to share their results with the community and hence maximise the number of users.

As part of the utilisation, the following issues will be investigated:

- Motivational factors for stimulating users to collaboratively tagging the songs. The service needs to be of mutual benefit in order for it to attract users.
- The effectiveness of the resulting infrastructure for the target audiences. Methods for individual features will be evaluated in the appropriate MIREX tracks.
- Copyright issues that arise from making the collections interoperable and accessible. These will be investigated and addressed building on S&V’s general copyright policies and expertise. Copyright issues affect both the thumbnails as a ‘preview’ mechanism for accessing the collection, and the use of complete pieces in the annotation infrastructure. NPO manages IPR as part of their daily operations.
- Establishing an organisational framework that will maintain the technical outcomes after the formal duration of the project.

### 7 Description of the proposed plan of work

The global division of tasks is as follows:

- The PhD student will design retrieval methods for musical audio, and its components such as segmentation and similarity measures.
- The postdoc will model the relationship between the listeners’ perception and structure of music.
- The programmer will design and implement the music search engine and its components.
- Remco Veltkamp is the general project leader, and will supervise the PhD student and programmer.
- Henkjan Honing is the site leader at UvA, and provides the expertise of cognitive modelling of music, and will supervise the postdoc.
- Frans Wiering provides the music information retrieval expertise and will supervise the PhD student and programmer.
- Anja Volk contributes with expertise on music variation modelling from her VIDI project.
- Louis Grijp is the site leader at the cultural heritage institute MI, and will direct the musicological research in folksong and popular music.
- Peter van Kranenburg is contributing expertise in modelling stable motifs in oral transmission of folk songs.
Martine de Bruin is responsible for the access to the MI Dutch Song Database and music collection. Johan Oomen is the site leader at the cultural heritage institute S&V. He will contribute expertise on crowdsourcing and access to distributed collections. Maarten Brinkerink (S&V) will provide legal expertise and will be the liaison between UU, Radio 5 and the external software agency in creating the annotation front end. Esther Herder is site manager of NPO Radio 5 Nostalgia, and will supervise the hosting of the annotation tool on the portal http://www.radio5nostalgia.nl/.

At the beginning of the project we investigate the copyright issues, as they might affect the architecture of the search engine.

**Expected results**

The envisioned research results are the following.

- A web-based annotation infrastructure for listeners (ITCH infrastructure).
- A ground truth collection of annotated musical audio.
- Cognitive models of music, in particular of rhythmic and harmonic similarity, appraisal and memorization.
- An architecture for storage and retrieval of distributed collections of musical data (audio, features, metadata).
- Methods for perceptually relevant audio segmentation.
- Audio retrieval methods based on patterns of melody, harmony and rhythm.
- Musicological research results on the connections between folk song and popular music.

These results are embodied in peer-reviewed journal papers, a PhD thesis, software modules, and a prototype system for musicological research at MI and searching in the collection of S&V.

**8 Literature**

**8.a References**


### 8.b Most important publications of the research group


