CrossMark

Documenting a song culture: the Dutch Song Database as a resource for musicological research

Peter van Kranenburg[1] · Martine de Bruin[1] · Anja Volk[2]

Abstract The Dutch Song Database is a digital repository documenting Dutch song culture in past and present. It contains more than 173 thousand references to song occurrences in the Dutch and Flemish language, from the Middle Ages up to the present, as well as over 18 thousand descriptions of song sources, such as song books, manuscripts and field recordings, all adhering to high quality standards. In this paper, we present the history and functionality of the database, and we demonstrate how the Dutch Song Database facilitates and enables musicological research by presenting its contents and search functionalities in a number of exemplary cases. We discuss difficulties and impediments that were encountered during the development of the database, and we sketch a future prospect for further development in the European context.

Keywords Dutch song culture · Database · Musicology · Literature · Metadata · History · Information retrieval

✉ Peter van Kranenburg
peter.van.kranenburg@meertens.knaw.nl

Martine de Bruin
martine.de.bruin@meertens.knaw.nl

Anja Volk
a.volk@uu.nl

[1] Meertens Institute, P.O.Box 10855, 1001EW Amsterdam, The Netherlands

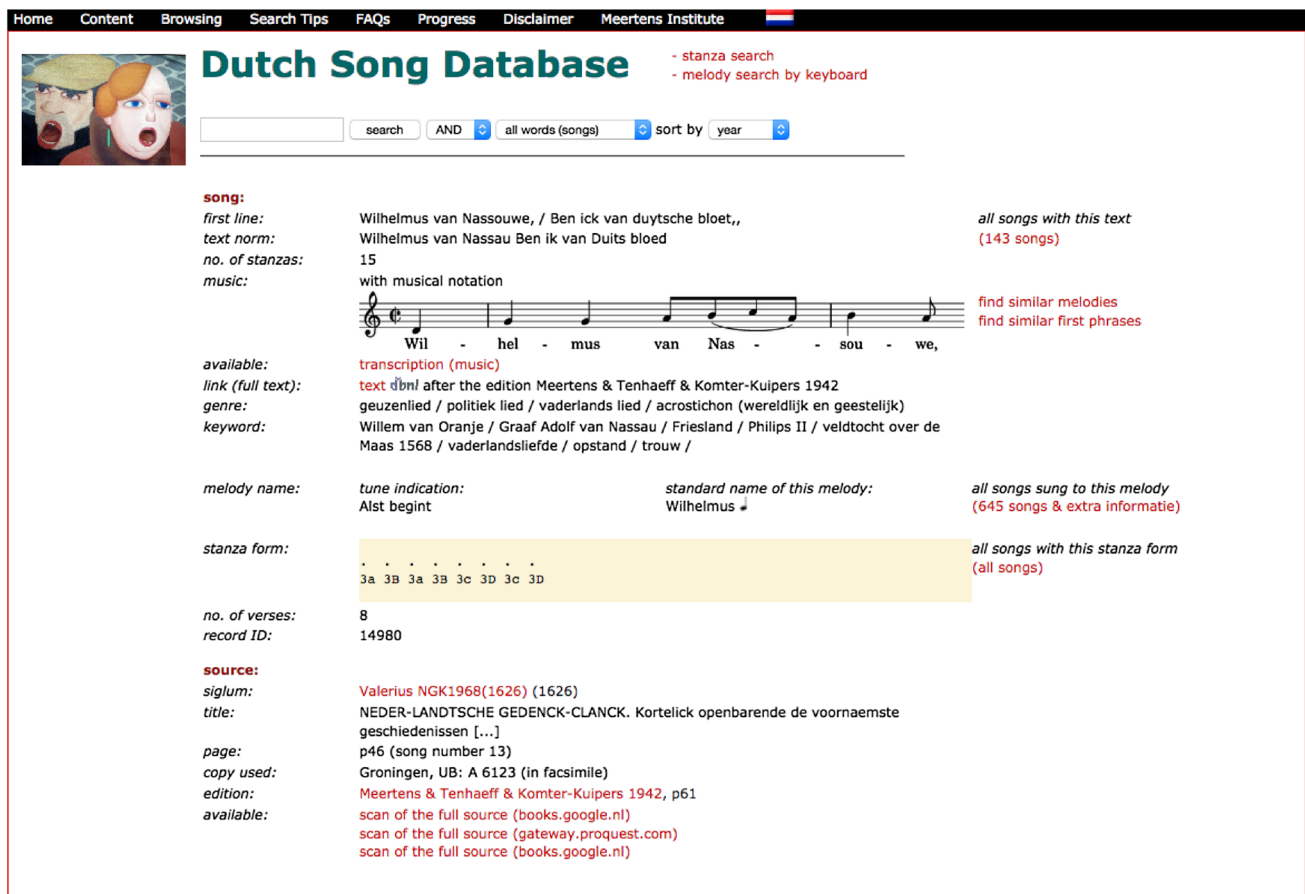[2] Utrecht University, Utrecht, The Netherlands

1 Introduction

The Dutch Song Database is a digital repository documenting Dutch song culture throughout the ages. The database was initiated in the early 1990s by the Dutch musicologist Louis P. Grijp (1954–2016), who continued to lead the development of the database till 2015. During these years, many research and documentation projects have been carried out, and gradually an enormous amount of high-quality data has been collected.

In this paper, we demonstrate how the Dutch Song Database facilitates and enables musicological research by presenting its main features (Sect. 2), the way in which melodies and song texts are linked in the database (Sect. 3), an overview of projects that have been carried out (Sect. 4), the Meertens Tune Collections (Sect. 6), search functionalities (Sect. 7) the academic context (Sect. 8), a number of exemplary research cases (Sect. 9), the lessons learned during development of the database (Sect. 10), and future perspectives (Sect. 11).

2 Introducing the Dutch Song Database

The main feature of the Dutch Song Database is the availability of collected references to occurrences of Dutch songs. The database can be accessed by the general public through an online interface at http://www.liederenbank.nl. Figure 1 shows the contents of a typical song record as presented in the online interface. In general, the most basic fields of a song description are the first line, a tune indication as it appears in the source, and a reference to the source.

At the moment, the database contains metadata on c. 173 thousand occurrences of Dutch songs in a variety of sources dating from the twelfth century up to the present day. A large

Fig. 1 Full description of a song as presented in the online interface of the Dutch Song Database

number of song texts and melodies have been digitized and are currently accessible within the collection. The database documents many kinds of (folk) songs, including love songs, satirical songs, beggar songs, psalms, other religious songs, children's songs, St. Nicholas songs and Christmas songs. The main sources in which these songs were found are songbooks, songsheets (broadsides), song manuscripts, and fieldwork recordings. The catalogued record for each song contains information about the source in which the text and/or the melody occurs. In most cases, direct links are provided to the complete song text, to a scan of the source, to the notated sheet music, or to a recording of an individual song.

The Dutch Song Database is hosted by the Meertens Institute[1] of the Royal Netherlands Academy of Arts and Sciences (KNAW) in Amsterdam and is maintained and developed further by the Oral Culture Group of the institute, in cooperation with several partners.

Since the initiation of the database, quality of the metadata has been of utmost importance. Therefore, most of digitizing has been done by hand and all entered songs have been double checked by database co-workers, resulting in an invaluable source of reliable song data. Nevertheless, it is important to realize that much of the data in the database is the interpretative work of several researchers and co-workers that worked on documenting the database materials over a period of years. In summary, working on such a digital archive as the Dutch Song Database remains a complex endeavour, which requires constant attention and a critical approach.

The online database is continuously being consulted by music and literary scholars as well as by the general public. Recently, we sent out a questionnaire to a relatively large number of Dutch active musicologists with the question whether they are interested in the functionality of the Dutch Song Database. The responses were very positive, and a majority of the respondents indicated an interest in collaboration. This exemplifies the infrastructural value of the database.

3 Linking melodies and song texts: standard names and stanza form

One of the initial aims in establishing the collection was for the database to map connections between song texts
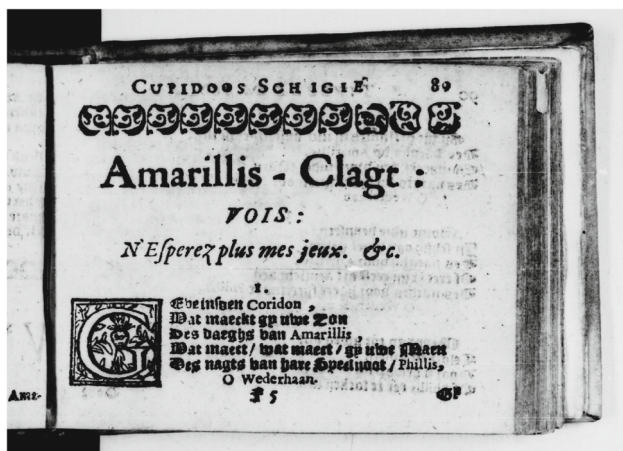
---

[1] http://www.meertens.knaw.nl.

Fig. 2 Facsimile of the song "Amarillis-Clagt" [Amarillis' complaint] as published in the songbook *Tweede Delfs Cupidoos schighje […]* (1656) [Second Delft Cupid's arrow]. Collection Royal Library of the Netherlands

and melodies from seventeenth-century sources. The typical appearance of a song in a historical songbook is without music notation, but with a *tune indication*, which is a written indication of the tune for the given song text. An example of this can be found in Figure 2, where "Vois: N'Esperez plus mes jeux. &c." indicates that the following text is supposed to be sung to the tune known as "N'Esperez plus mes jeux". Historic users of the book were expected to know the indicated tune by heart, so they could sing the song simply with the aid of the tune indication. Unfortunately, this knowledge is not available for the researcher of today and one needs to undertake complex investigations with the help of the database for all this lost knowledge to be revealed. One way to reconnect a tune indication with an actual tune is to examine contemporary songbooks that employ music notation and to compare these with songbooks that incorporate the same name of the tune indication. While this seems straightforward, there are several issues that complicate this task. First, tune indications were far from standardized. The same melody can be combined with various tune indications in various sources. Secondly, since, historically, popular song culture was, in large part, an oral culture, a great deal of variation occurred both in texts and melodies

To cope with these difficulties, two key concepts were introduced in creating and organizing song data in the database, namely the *melodienorm* (standard name of the melody) and the *incipitnorm* (standard name of the text). Whenever the person documenting "recognizes" a text or melody while entering a song description into the database, he or she enters the standard name of it in the respective field of the song record. Thus, a *melodienorm*—the standard name of the melody—unites all songs in the database that were presumably sung on the same melody. It is important to realize that assigning a standard name to a text or a melody

is actually an act of interpretation by the person who is documenting occurrences of a given song text and song melody.

Through this method of identifying songs melodies, it is possible for users of the database to retrieve song texts that were sung to a given melody with one search in the database. This even is the case if the actual melody is not traceable anymore because there is not a single known source with the actual music notation of the melody. In these cases, we know that the melody existed, and we are able to identify the tune indications with which the melody was combined, but we do not know any variant of the melody itself.

The database provides a second mechanism to link texts and melodies, namely the *stanza form* [6]. The *stanza form* is a symbolic representation of the structure of a stanza with regard to the number of stressed syllables per line and the rhyme, indicating whether the form uses masculine or feminine rhyme structures. A very simple example is: .3A .3b .3A .3b, which denotes a stanza of four lines with three stresses in each line, with a masculine rhyme structure (indicated with capital letters) between the first and third lines, and feminine rhyme structure (indicated with lowercase letters) between the second and fourth lines. The dots represent a line starting with an unstressed syllable ("pick up"), like in an iamb or anapest.

The crucial insight here is that all stanzas with the same stanza form can be sung to the same melody. Therefore, a strategy to find the melody on which a given song text could have been sung would be through the examination of contemporary songbooks, to find songs with the same stanza form that include a notated melody or tune indication. For that reason, the stanza form of a song is important information for identifying songs and is included as part of the song data to be entered in the database. A search facility to retrieve songs with a certain stanza form is included in the user interface.

4 Data entry projects

Since entering song descriptions is a laborious task, which requires an advanced set of skills, adding new information to the database is a slow process. Within the last two decades, a large number of projects have been carried out at the Meertens Institute, each of which provided specific contents to the database. All data entry was supervised by musicologists and literary scholars, and many of the project workers were graduated musicologists, or musicology students.

As a consequence of the project-based progress, some historical epochs are more prevalent and some are underrepresented in the database. A full overview of the projects is provided online at the site of the database.[2] We now will present a few major projects.

---

[2] http://www.liederenbank.nl/NLBprojecten.php.

The Nederlands Volkslied Archief [Dutch Folk Song Archive] was the predecessor of the Dutch Song Database. It consisted of an index card system that was established in the 1950s and maintained till the 1980s [32, pp. 351ff]. In a major project (1999–2004), the information contained on the c. 80 thousand song cards was entered into the Dutch Song Database.

During the project *Repertory of Dutch songs until 1600* (1993–2001), which was performed in collaboration with the University of Antwerp (UFSIA), virtually all known Dutch monophonic songs contained in sources dating until 1600 were entered. The sources, both in print (625) and in manuscript (314), were found in libraries all over the world. They originate from both the Northern and the Southern Netherlands (today, The Netherlands and Belgium). The project resulted in a book and an accompanying CD-ROM (2001), presenting the repertory both in printed form and as digital database [4]. In total, the repertory comprises 11,077 variants of 7,621 song texts, and 1,158 melodies.

With a grant from Metamorfoze, the Netherlands' national programme for the preservation of paper heritage, the project *Songsheets 1750–1950* was carried out (1998–2001). As many as 12,359 songs from a collection of songsheets and broadside ballads, owned by the Meertens Institute and the Royal Library of the Netherlands, were restored, stored in acid-free boxes and documented in the database. This was directly followed by the project *Streetsongs* (2002–2004), in which this collection and some additional sources were scanned and published through the national heritage website *The Memory of the Netherlands*.[3]

The large collection of field recordings *Under the Green Linden* (see Sect. 9.3) was digitized in several stages. As part of the Meertens Institute's aim to preserve all the audio recordings, the original tapes were digitized in the Meertens audio studios in 2001. The metadata were entered during 2002–2007. A total of 1,300 singers biographies were added in 2004–2009, and 3,754 available handwritten and typed transcripts were scanned by the Meertens staff in 2005 as a preparation for the research project WITCHCRAFT (2006–2010) [34].

Although the database started as a collection of metadata, in recent years, complete song texts and song melodies have been digitized in large quantities. *Dutch Songs OnLine* (2009–2014) was a collaborative project involving the Dutch Song Database and the Digital Library for Dutch Literature (DBNL). The DBNL is a website holding 2 million XML-pages containing full text editions in the fields of Dutch literature, language and cultural history. The project allowed for the corresponding parts of the data sets (more than 53 thousand full song texts) of Dutch Song Database and DBNL to be linked.

In the projects *Dutch Folksongs as Musical Content* (2007–2009),[4] *Speelmuziek* (2010–2011),[5] and *Tunes & Tales* (2012–2016),[6] almost 17 thousand full melodies have been digitized and stored in a parallel database that is directly connected with the Dutch Song Database.

Although—apart for the period until 1600—the coverage of the database is far from complete, the present contents of the database offer a very rich cross section of Dutch song culture throughout history. The availability of this material enables analysis of Dutch songs on a large scale.

5 The Dutch Song Database in numbers

As of September 2016, the Dutch Song Database contains 173,964 descriptions of songs and melodies. The songs are contained in 18,014 different sources located in c. 300 different libraries and archives. Over 48,573 songs incorporate music notation in their sources and 16,684 of these melodies have been digitized and are searchable through the melody search engine, while many more are accessible as scanned items. There are 9,393 audio recording accessible online, and 53,390 thousand links to full song texts. The database includes 51,339 searchable stanza forms. In total, there are 70,772 different standard names of texts and 12,293 different standard names of melodies.

6 Meertens Tune Collections

The digital availability of melodies, song texts and metadata allows for statistical analysis of the corpus as a whole, or of substantial parts of it. The online interface, however, provides access to the data on the basis of individual songs. To facilitate large-scale investigations, parts of the contents of the Dutch Song Database have been released as a collection of downloadable data sets under the name *Meertens Tune Collections* (MTC) [14,16]. These data sets provide song or music researchers with large collections of songs and metadata in various formats. The initial release of the MTC consisted of a collection of early twentieth-century vocal songs (MTC-FS), a collection of seventeenth- and eighteenth-century monophonic instrumental pieces (MTC-INST), a small quantity of vocal songs that carefully has been selected for representativeness (MTC-ANN), a large corpus of mixed contents that has been used in various

---

studies (MTC-LC), scans of handwritten transcriptions of songs from the collection *Onder de groene linde* (MTC-OGLSCANS), and the full set of audio files comprising the collection *Onder de groene linde* (MTC-OGLAUDIO).

7 Searching the Dutch Song Database

Songs can be found using a simple online search form, which is included in the opening page of the online interface of the database. A search action will lead to a result list, which shows short representations of the individual songs in chronological order. By clicking on one of the results, a full description of the song appears, as depicted in Figure 1. This full description offers a range of extra search possibilities. One may find all other songs with the same or similar versions of the melody by following the link "all songs sung to this melody". In a similar way, the link "all songs with this text" leads to a list of songs with identical or similar song texts. Clicking the link "all songs with this stanza form" results in a list of songs with exactly the same stanza form. For those songs for which a full melody has been encoded in the database, there are two further search possibilities: "find similar melodies" and "find similar first phrases", which, respectively, lead to a list of songs with similar melodies and to a list of songs with a melody that starts similarly.

It is also possible to click the siglum (or library code) of the source of a song. This leads to a full description of the source, including (if applicable) title, author(s), year and place of publication, publisher, literature, copies (library sigla), number of songs and kind of source (manuscript or print, songbook or songsheet, etc.), other editions, and a link to a list of all the songs in a source (provided that these have been entered into the database). The database also contains titles of sources which have not (yet) been excerpted. One can also search sources using the simple online search form on the opening page of the database.

There is also the possibility for the user to enter a stanza form or a melody to query the database with, by choosing the "stanza search" option, or "melody search by keyboard" from the opening page. The melody search engine was developed in the WITCHCRAFT-project [13].

For most of Dutch history, spelling was not standardized. This impedes a textual search. Therefore, all first lines, titles, choruses and tunes indications have also been entered into the database in modern spelling by the documentalists. For instance, in sixteenth- and seventeenth-century spelling, the Dutch word for song "lied") could have been written as "lied", "liedt", "liet", "leidt", etc. Because the modern spellings are included in the database, both the original and modernized words can be searched for.

8 The Dutch Song Database in context

Originally stimulated by Romantic and Nationalistic ideologies, collecting folk material has been an important part of ethnological research [26]. From the late eighteenth century onwards, ethnological fieldwork activities resulted in many collections of folk songs, either in transcription or in recordings. Many of these collections have been preserved until the present day in institutions such as libraries, universities or museums.

To our knowledge, there are no digital databases in existence that are equivalent to the Dutch Song Database in that they intend to document an entire nation's song culture, encompass an extended period of many centuries, include a rich variety of sources, provide a diversity of data representations, and are online searchable in a variety of ways, including both musical and textual content and metadata.

One comparable digital database is the online searchable Roud Folk Song Index, compiled by Steve Roud,[7] which is hosted by the Vaughan Williams Memorial Library (VWML) in London, England's national folk music and dance archive. This index provides information on English-language traditional songs. Each identified song text obtained a Roud number, roughly corresponding to the *incipitnorm* of the Dutch Song Database, and each entry in the index corresponds with an occurrence of a song. The index contains over 200 thousand occurrences of around 27 thousand songs. Compared to the Dutch Song Database, more variants of less songs are documented. Furthermore, the Roud index does not provide the mechanisms to connect melodies and texts that are central to the Dutch Song Database. The VWML further provides a rich variety of source material on English folk arts.

Another example is the German Historisch-kritisches Liederlexikon, which offers editions of different versions of songs and summaries of song commentaries in earlier editions.[8] This database is hosted by the Zentrum für Populäre Kultur und Musik in Freiburg (formerly known as the German Folksong Archive) and contains 285 different songs in 1,247 editions.[9] Next to full texts, this database contains scans of music notation and links to commercial recordings. Although the number of songs is small, the information provided for each song is extensive.

In the course of the twentieth century, the interest of musicological scholars shifted away from musical folk song material as object of research, as noted by Bruno Nettl [22, p.130]. In recent years, however, the rise of digital approaches in humanities and the advances within the field of Music Information Retrieval, caused a renewed academic interest in

---

7 http://www.vwml.org/search/search-roud-indexes.

8 http://www.liederlexikon.de/.

9 http://www.liederlexikon.de/register/lieder.

Waer **werd** op**rech**ter **trouw**	3	a	m	Where was faith more sincere
Dan **tus**schen **man** en **vrouw**	3	a	m	Than between man and woman
Ter **wee**reld **oit** ge**von**den?	3	b	f	Ever found in the world?
Twee **zie**len **gloe**nde aan **een** ge**smeed**,	4	c	m	Two souls forged glowingly together
Of **vast** ge**scha**keld **en** ver**bon**den	4	b	f	Or steadily linked and united
In **lief** en **leedt**.	2	c	m	In love and sorrow.

Fig. 3 Text and stanza form of "Waer werd oprechter trouw". The columns show the Dutch text with accented syllables in bold, the number of accents, the rhyme scheme, whether the rhyme is masculine or feminine, and an English translation

digitized collections of folk song material [15]. For example, the EsAC Folksong Databases [25], which currently include over 20 thousand digitized folk song melodies, were used extensively within the field of Music Information Retrieval in order to test algorithms that operate on monophonic songs. Among both digital and non-digital resources, the Dutch Song Database is unique in the world, as its highly organized digital corpus of metadata exists as a resource serving to document a nation's song culture.

9 Examplary research cases

The database, as it has been developed until the present day, enables empirical research in various disciplines such as Historical Musicology, Music Cognition, Music Information Retrieval, History and Philology.

The following presents a number of diverse case studies showing how the Dutch Song Database assists music and song research. These case studies demonstrate that content-based search facilities remain crucial to addressing specific kinds of research inquiries. While some studies have concentrated on individual songs, other investigations have involved the collection as a whole. The first two cases are specific for the history of Dutch song culture. The following case addresses a more general issue in folk song research: variation caused by oral transmission. To show that the data are also very well suitable for studying general concepts in Musicology, MIR, and Music Cognition, we finally present cases on music similarity, pattern discovery and absolute pitch.

9.1 Rediscovering melodies

The Dutch Song Database provides search facilities that allow researchers to find connections between texts and melodies in historical sources. The organization of the database allows for given song texts to be reunited with the melody on which they once had been sung. An example of this is found in the chorus "Waar werd oprechter trouw" from the play *Gysbrecht van Aemstel* (1637) by Joost van den Vondel (1587–1679). This is a very famous text in the literary history of the Netherlands, and there has been debate as to whether this text was traditionally sung or spoken during performances of the play. The first stanza, including the stanza form, is depicted in Figure 3.

In this study, which was performed by Louis P. Grijp in 2005, the first step was to establish the stanza form, which consists of the number of accents in each line, the rhyme scheme, and whether the rhyme is masculine (m) or feminine (f). The stanza form is denoted next to the text in Figure 3. The accented syllables are indicated in underscore. Each line starts with an unaccented syllable. Combining all these observations, we establish the stanza form as: `.3A .3A .3b .4C .4b .2C`.

The strategy Grijp used in this investigation was to search in the Dutch Song Database with this stanza form and to see whether there exists a song with the same stanza form in another source that has either a tune indication or even music notation. The result list presented a song called "Amarillis-Clagt" [Amarillis' complaint], which was published in the songbook *Tweede Delfs Cupidoos schighje […]* (1656) [Second Delft Cupido's arrow]. Figure 2 shows a facsimile. This song has exactly the same stanza form as the poem by Van den Vondel. It has as tune indication: "N'Esperez plus mes yeux. &c.". Although there is no music notation in this source, this finding provides a name of a melody on which "Waer werd oprechter trouw" can be sung. There appears to be an *air de cour* by Antoine de Boësset (1586–1643) with the same title. The song texts of both "Amarillis-Clagt" and "Waer werd oprechter trouw" fit perfectly on this music. Boësset's composition was published in *Harmonie Universelle* (1636) by Marin Mersenne. This book had had a wide dissemination and had been read throughout Europe and in the Netherlands as well. Therefore, it is quite certain that this melody would have been known in the Netherlands by 1637. The particular stanza form is rare; therefore, it is unlikely that another song with exactly the same stanza form would have been sung to an independent melody. Furthermore, the dates of Boësset's composition and Mersennes publication closely precede the year in which Vondel wrote his play. Therefore, it is highly plausible that this indeed was the melody on which "Waer werd oprechter trouw" was sung. The discovery of the melody by Louis P. Grijp generated much media attention and ended the controversy whether the chorus has been sung or not [7].

9.2 Lully in Holland

The ongoing digitization of melodies from Dutch instrumental sources from the seventeenth and eighteenth centuries enables large-scale investigations of the origin of these melodies. An initial study of these melodies showed that many of them had their origin in France, more particularly in operas and in so-called *airs de cour*. In 2012, a project was started at the Meertens Institute to investigate the French influence, in particular of Jean-Baptiste Lully (1632–1687), on Dutch popular song culture, using the facilities of the Dutch Song Database.

A main challenge of this research is that the origin of a given melody is almost never mentioned in Dutch manuscripts and printed sources. In this way, a given melody with a Dutch title could originate from a French opera. Therefore, in almost all cases, a content-based search is required wherein the melody is employed as the material for the query and a database search is made using a large collection of melodies that have a known origin. The collection of the Dutch Song Database is one such collection. If a given melody also occurs in another source that already has been included and identified in the Dutch Song Database, a single melody-based search could allow for the identification of a given song. For instance, the tune "Mars" [March] in a very small manuscript found in the seventeenth-century dollhouse of Petronella de la Court could be identified as the "Marche des 4 Nations" in *Ballet de Flore* by Lully (LWV 40, 1669) and "Rondeau" from the same manuscript could be identified as "Que n'aimez-vous coeurs insensibles" from Lully's opera *Perse* (LWV 60, 1682).

Using this approach, we systematically examined a number of instrumental sources with monophonic melodies. Each given melody was digitized and used as material for query both in the search engine of the Dutch Song Database and in the online search engine of the RISM (Répertoire International des Sources Musicales). Thus, we were able to identify a multitude of melodies that were part of Dutch popular culture in the seventeenth and eighteenth centuries.

The project resulted in two musicological bachelor theses, and three master theses [2,9,33], while a fourth is in preparation.

9.3 Oral transmission of songs

Traditionally, folk songs were sung by ordinary people during work or social activities. These folk songs were part of a larger oral culture in the Netherlands and the songs—especially the melodies—were learned primarily through listening and participation. Within the process of oral transmission, the songs underwent continuous change.

As a consequence of industrialization and the introduction of the radio, this form of traditional singing has completely disappeared from the Netherlands. The fact that we still have access to this tradition is because of a long-term project of transcribing the songs either into musical notation or recording them on tape. The first to collect written transcriptions of folk songs from the Dutch language area was the Flemish writer Jan Frans Willems (1848). Among the first Dutch sound recordings were those made by Will Scheepers in the early 1950s. The Dutch ethnologist Ate Doornbosch continued this fieldwork project. In total, more than seven thousand field recordings were made. For c. five thousand recordings the aid of Doornbosch's radio programme *Onder de groene linde* (Under the Green Linden, 1957–1994) was of great importance. This programme was an early example of interactive radio, as listeners were encouraged to contact Doornbosch if they knew more about the songs that were broadcast. Doornbosch would then record the listener's version and broadcast it. In this manner, a collection was created that documents not only an aspect of Dutch cultural heritage, but also the textual and melodic variation that results from oral transmission. Doornbosch was specifically interested in Dutch ballads: old strophic songs of considerable length in which an often tragic story is told. The collection is currently known as *Onder de groene linde* (OGL) and is preserved by the Meertens Institute.

The concept of *melodienorm*, which had been introduced to identify different tune indications that refer to the same tune, was re-used for the OGL collection in a related but somewhat different manner. In Folk Song Research, the concept of *tune family* had been introduced to identify a number of folk song melodies that presumably have one common "ancestor" melody in the process of oral transmission [1]. For the songs in OGL, the field of *melodienorm* is used to indicate the tune family the song belongs to according to the person documenting the song. The various individuals documenting songs for the Dutch Song Database have been able to group many melodies from oral transmission into tune families. This is a process that is based on human recognition of the songs, either through human memory or through results of the melody search engine of the Dutch Song Database.

The availability of this large collection of material derived from oral transmission enables large-scale investigations of processes of oral transmission within this song tradition. In the project Tunes & Tales, hypotheses regarding the stability of melodies in oral transmission are being tested. In a previous project, the WITCHCRAFT-project [34], a search engine for melodies was developed that takes melodic variation into account [13]. Furthermore, the data have been used by many researchers for investigations into monophonic songs from oral cultures, e.g. in [3,10,17,24,28].

9.4 Music similarity

Music similarity is a key concept for the computational processing of music information in Music Information Retrieval (MIR), for investigating musical structure in Musicology, and for cognitive processes involved in human engagement with music as investigated within the field of Music Cognition [29]. Music similarity constitutes a very challenging concept for various scientific and scholarly disciplines [30]. For instance, in MIR the concept of music similarity is researched in order to enable the retrieval of musical pieces from a large digitized collection that are similar to a given query, and in Musicology, analyses of musical pieces, such as paradigmatic analysis [21], are based on similarity of musical patterns.

The lack of reference data regarding which musical pieces are considered similar is often a stumbling block for developing and evaluating computational models of similarity in MIR. Musicologists, on the other hand, work implicitly with the notion of similarity on a daily basis, without explicating the underlying processes that lead to assessing similarity, these existing as an integral part of their musicological expertise. The Dutch Song Database provides, with its rich musical material, a valuable source for researchers working on music similarity in different disciplines, and has made it possible for investigators to form important insights regarding melodic similarity, especially through empirical investigations of the concept of tune family. These insights enrich the usefulness of the tune family as reference data for melodies that are considered similar in MIR, and provide seminal examples for understanding human categorization processes in music, as investigated in cognitive studies.

In Cognitive Science, the relationship between human categorization processes and similarity is a major topic [8], assuming that similar items are placed in the same category. These models of similarity-based categorizations have been opposed by models of theory-based categorization [20,27] with the argument that the concept of similarity is in some cases not sufficient in order to explain how humans form categories. While many experiments regarding categorization processes in the field of Cognitive Science have been carried out in other domains than music, several studies in music cognition have shown that similarity is crucial for categorization processes in music. However, it remains an open question as to what and how musical features contribute to similarity in categorization processes.

The concept of the tune family in the Dutch Song Database is specifically interesting within the context of categorization processes in music. Since the historical processes of oral transmission have been lost, experts rely on similarity for categorizing the folk song melodies into tune families. This makes the tune families an ideal subject for studying the question as to which musical features contribute to the assessment of similarity of folk songs that have been categorized within the same tune family. Experts make intuitive and holistic decisions without an explicated system about rules of categorization that determines how musical features may contribute to a given similarity assessment. We have therefore developed an annotation study to explicate the features that play a role in the categorization of similar melodies in tune families [31]. The motivation for the annotation study came from the goal of the WITCHCRAFT-project—to develop a search engine for the folk songs that will allow researchers and laymen to automatically find similar melodies. The annotation study was an important step in gaining insight into the nature of the similarity between the songs in order to develop computational approaches.

When developing this annotation method in conjunction with experts, it turned out that certain melodies were allocated the status of a prototypical melody within their tune family. The prototypical melody hence is considered the most typical representative within the family. As reported by the experts, this prototypical melody most distinctively stays in their mind as the most characteristic version of all songs belonging to one family. All other possible melodies are compared to the prototypical melody in order for researchers to decide whether the melodies belong to this tune family. Together with co-workers of the Dutch Song Database, we discussed which musical features play a seminal role in similarity assessment. Melodic contour, rhythm, lyrics and motifs had been identified as relevant features contributing to similarity assessments. We then defined criteria for numerical ratings for each feature that best reflect a given experts' working method. As a result, the experts annotated 360 melodies from 26 representative tune families from *Onder de groene linde* following this procedure. All melodies of a tune family were compared to the reference melody, and the level of similarity for all musical features was rated. This formal procedure proved similar to the daily working practice of the experts consulted.

The corresponding data set MTC-ANN that resulted from this annotation study delivers valuable information on the contribution of musical features to the process of categorization. For most melodies, the similarity assessment is a multidimensional process, since several musical features have been determined to be important for the categorization process. While global characteristics of melodies, such as discussed by Wiora [35], are relevant for the corpus of *Onder de groene linde*, it turned out that in general the recurrence of characteristic motifs seems most important for the perceived similarity between melodies. The annotation study confirmed the crucial role of similarity for categorization processes in music and provided valuable insights into the musical features that contribute to perceptions of similarity. This study has provided important data for MIR researchers

who seek to develop models to determine music similarity (e.g. [3,12,17,19,28]).

9.5 Pattern discovery

The Dutch Song Database has developed into an important reference tool for research in the area of automatic pattern discovery. Melodies of *Onder de groene linde* are classified into tune families, each family consisting of a set of similar melodies. According to an annotation study that was carried out to reveal the role of several musical features in creating perceptions of similarity between melodies [31], the most important ingredient of perceived melodic similarity appeared to be the occurrence of similar motifs between songs belonging to the same tune family. Therefore, the collection of the songs, together with the annotated motifs as distributed in the data set MTC-ANN, provides an invaluable resource within the field of automatic pattern discovery in Computational Music Analysis, and for the investigations into the role of repetition in music. According to musicologists, repetition is an essential feature of music [5], which occurs on all levels of musical organization [18], existing both within a given musical work and between musical works. Yet, there are many open questions with regard to the specifics of repetition in musical works and styles [31], such as the question as to whether some composers employ repetition more than others [18].

Modelling repetition in music computationally constitutes a prominent area in Computational Music Analysis in the task of detecting repeated patterns through automatic pattern discovery. Due to music's extremely repetitive nature, most pattern detection algorithms discover large amounts of patterns [11], which often require extensive filtering techniques for determining patterns with relevant musical structures. Hence, as of today, within the field of Computational Music Analysis automatic pattern discovery still presents challenges. One of the challenges relates to the question of what constitutes a salient or musically meaningful pattern, and how similarity in music is based on the occurrence of similar patterns [31]. The annotated motifs of MTC-ANN therefore provide an invaluable data source of recurring patterns considered to constitute similarity between songs belonging to one tune family.

In [3], the annotated motifs have been used as a way to compress the songs. All notes that do not belong to an annotated motif were removed. While thus 60% of the note material was removed, the accuracy of the employed classification algorithm, which classifies the melodies into tune families, decreases only by 3 percentage points. Hence, the annotated motifs of MTC-ANN carry indeed salient information for the similarity between the songs. Therefore, the annotated motifs of MTC-ANN provide an outstanding source of musical materials for studying the role of repetition

within melodic similarity and to study techniques for music compression. Another approach for using the tune families as a source for investigating the use of compression in computational methods is reported in [19]. Therein, pattern discovery is used to calculate the compression distance between songs. Different point-set compression algorithms, i.e. pattern discovery algorithms that treat a musical piece as a set of points in a multidimensional space, are compared for the purpose of classification.

9.6 Absolute pitch

A recent study performed by Olthof et al. [23] makes use of the availability of the recordings of the digital collection *Onder de groene linde* in another manner. They researched an aspect of music cognition, as the collection served as material to test a specific problem of absolute pitch. These researchers investigated the similarity of the absolute pitch height within songs from the same tune family. Receiving positive results for a number of tune families indicated to the researchers that the absolute pitch height of a given song is, to a certain extent, preserved in the process of oral transmission. From the perspective of music cognition, this is an important finding, as it contributes to the understanding of how the parameter of pitch functions within processes of oral transmission and (collective) musical memory.

10 Lessons learned

One of the major lessons learned in the 25 years of developing the Dutch Song Database is that it can be hard to "capture" a song in a database. Over time, songs could change radically in terms of their functionality and content. For instance, a given song can transform over time to be used in various genres. What was once a theatre tune can become very popular as a street song, and what was once a song with strong political connotations can be used later for a wedding song. In this process, people can change both the text and melody either consciously or unconsciously. These kinds of transformations are fascinating to study, but it can be quite difficult to describe the processes of transformation. For this reason, songs in the database need to be described in their historical context with proper information about where, when and by whom they were sung or published. What makes the process of defining a given song even more complicated is the fact that each song consists of both textual and musical information. For this reason, structural relationships between songs exist as a web of musical and textual citations, imitations, variations, and adaptations, all being very challenging for the archivist and researcher to describe. In this approach, it is crucial to determine structural connections between songs and to see how relationships between them exist. In order to
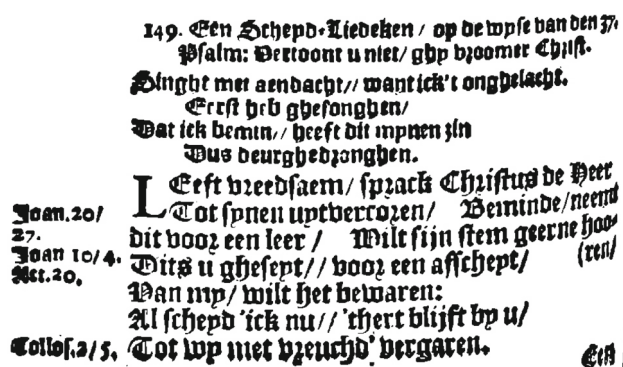
Fig. 4 The beginning of a song from Leendert Clock, *Het groote liede-boeck van L.C.* (1625) [Leendert Clock's big songbook], fol. T8v, exposing the index and title "149, a farewell song", a tune indication "to the tune of Psalm 37: Don't show yourself, thee pious Christ.", a four-line poem, and a song text starting with "Leeft vreedsaem" [Live peacefully]. Special Collections, University of Amsterdam (OK 65-66)

do this, the metadata scheme of the Dutch Song Database has been extended over and over again: the *incipitnorm*, for example, was introduced only after about eight years.

Dutch song culture as recorded in the database spans several centuries. Ideally the database would provide the information that is needed to trace individual melodies and song texts in their variants during their transmission over long periods of time. One problem in defining transmission is the lack of solid data within historical sources: songs are often published without a proper title and often the name of the author or composer is missing in sources that provide not even a hint of where and when they were made. Yet this is crucial information in our diachronic approach and it is a very time-consuming operation to make the metadata as complete as possible.

In the process of linking the Dutch Song Database with the database of the Digital Library for Dutch Literature during the project *Dutch Songs Online* (as introduced in Sect. 4), many difficulties were encountered. Just one example is an incompatibility of ways in which the "title" of a song is determined. To illustrate this problem, Figure 4 depicts a scanned image of a song as occurs in *Het groote liede-boeck van L.C.* (1625) [Leendert Clock's big songbook]. The seemingly straightforward question what is the title of this song has several potential answers. For example, whether or not to include the tune indication, or the four-line poem. Many of those questions of definition were encountered, which is easy to underestimate at the beginning of such a project.

Another issue is the international context. The database contains songs in foreign languages like Latin, French and German that were part of Dutch song culture. Many of these songs were included in Dutch songbooks and it is clear that even more "foreign" tunes were known in various times in the history of the Netherlands. For example, the database

contains more than 10,000 French tune indications. While during some periods in history, some Dutch social groups spoke a foreign language among themselves, it is also certain that many songs and melodies had been imported from other countries as well. Many folk song collections—like the Dutch Song Database—are national collections, but melodies and songs do not stop at international borders. Therefore, international collaboration is required and, ideally, all existing European data sets should be linked.

11 Future directions

There are many ways in which the Dutch Song Database can be extended and improved. In this section, we present some ideas and wishes.

Due to the fact that much work on the database was project based, concentrating on specific historical epochs, song repertoires from other epochs are still underrepresented. This remains a problem within a collection that aims to represent Dutch song culture in its historical development. For example, song repertoire from the nineteenth century is underrepresented. We expect many twentieth-century field recordings in *Onder de groene linde* to have origins in the nineteenth century, which are presently lacking within the database. Furthermore, several Flemish song sources are missing, including choruses sung in theatre in the seventeenth and eighteenth centuries, children's songs from the twentieth and twenty-first centuries, and so on. Yet, it is an illusion to aim for a "complete" data set. A great deal of historical song material has simply been lost, and the number of sources showed a substantial increase starting in the nineteenth century. Collecting these materials remains a gigantic undertaking. Nonetheless, efforts to add materials from underrepresented epochs would render the database more representative of Dutch song culture, as it existed throughout history. Once additional texts and melodies become included in the database, this will enable new content-based comparisons and investigations on an even larger scale than is possible at the present time.

The study *Lully in Holland* shows that Dutch popular culture was not isolated from the rest of Europe. Many melodies are of French origin. We know that there are also relations with, for example, German and English song cultures. Therefore, a sensible next step would be to make connections with researchers and institutions conducting folksong research in other European countries and to study popular song culture in a larger European context. This would require similar documentation projects for national song histories of other countries as well. The approach of the Dutch Song Database in documenting relationships between song texts and song melodies, as these are found within historical sources, can be seen as exemplary. We aim to actively continue pursu-

ing national and international collaboration to continuously improve the Dutch Song Database as an invaluable source of knowledge about Dutch national musical heritage in the context of the larger cultural domain of Europe.

References

1. Bayard, S.: Prolegomena to a study of the principal melodic families of British–American folk song. J. Am. Folk. **63**(247), 1–44 (1950)
2. Van den Berg, M.: Boertige en ernstige minnezangen 1696–1709: De plaats van Sweerts' liedboek in de geschiedenis van het Nederlandse lied. Master's thesis, Utrecht University (2014)
3. Boot, P., Volk, A., de Haas, W.B.: Evaluating the role of repeated patterns in folk song classification and compression. J. New Music Res. **45**(3), 223–238 (2016)
4. De Bruin, M., Oosterman, J., Strijbosch, C. (eds.): Repertorium van het Nederlandse Lied tot 1600/Repertory of Dutch songs until 1600. Meertens Instituut and Koninklijke Academie voor Nederlandse Taal- en Letterkunde, Ghent and Amsterdam (2001)
5. Fitch, W.T.: The biology and evolution of music: a comparative perspective. Cognition **100**(1), 173–215 (2006)
6. Grijp, L.P.: Het Nederlandse Lied in de Gouden Eeuw. P.J. Meertens Institute, Amsterdam (1991)
7. Grijp, L.P.: Waer werd oprechter trouw: over de muziekhistorische uitvoeringspraktijk op de amsterdamse schouwburg. Tijdschrift Oude Muziek **22**(4), 16–20 (2007)
8. Hahn, U., Ramscar, M.: Similarity and Categorization. Oxford University Press, New York (2001)
9. Heldens, W.: Lully als speelmuziek: De plaats van handschrift Berkeley in de vroegmoderne populariteit van Lully in nederland. Master's thesis, Utrecht University (2015)
10. Hillewaere, R., Manderick, B., Conklin, D.: Alignment methods for folk tune classification. In: Spiliopoulou, M., Schmidt-Thieme, L., Janning, R. (eds.) Data Analysis, Machine Learning and Knowledge Discovery, pp. 369–378. Springer, Berlin (2014)
11. Janssen, B., de Haas, W.B., Volk, A., Van Kranenburg, P.: Finding repeated patterns in music: state of knowledge, challenges, perspectives. In: Aramaki M., Derrien O., Kronland-Martinet R., Ystad S. (eds.) Sound, Music, and Motion-10th International Symposium, CMMR 2013, Lecture Notes in Computer Science, vol. 8905, pp. 277–297. Springer (2014)
12. Janssen, B., Van Kranenburg, P., Volk, A.: A comparison of symbolic similarity measures for finding occurrences of melodic segments. In: Proceedings of the 16th International Society for Music Information Retrieval Conference, pp. 659–665. Málaga (2015)
13. Van Kranenburg, P.: A computational approach to content-based retrieval of folk song melodies. Ph.D. thesis, Utrecht University, Utrecht (2010)
14. Van Kranenburg, P., De Bruin, M., Grijp, L.P., Wiering, F.: Meertens Tune Collections. Meertens Online Reports 2014-1, Meertens Institute, Amsterdam (2014)
15. Van Kranenburg, P., Janssen, B.: What to do with a digitized collection of western folk song melodies? In: A. Holzapfel (ed.) Proceedings of the Fourth International Workshop on Folk Music Analysis, pp. 117–119 (2014)
16. Van Kranenburg, P., Janssen, B., Volk, A.: Meertens Tune Collections: the annotated corpus (mtc-ann) versions 1.1 and 2.0.1. Meertens Online Reports 2016-1. Meertens Institute, Amsterdam (2016)
17. Van Kranenburg, P., Volk, A., Wiering, F.: A comparison between global and local features for computational classification of folk song melodies. J. New Music Res. **42**(1), 1–18 (2013)
18. Margulis, E.H.: On Repeat. How Music Plays the Mind. Oxford University Press, New York (2013)
19. Meredith, D.: Music analysis and point-set compression. J. New Music Res. **44**(3), 245–270 (2015)
20. Murphy, G.L., Medin, D.L.: The role of theories in conceptual coherence. Psychol. Rev. **92**(3), 289–316 (1985)
21. Nattiez, J.J.: Music and Discourse: Toward a Semiology of Music. Princeton University Press, Princeton (1987)
22. Nettl, B.: The Study of Ethnomusicology: Thirty-one Issues and Concepts, 2nd edn. University of Illinois Press, Urbana (2005)
23. Olthof, M., Janssen, B., Honing, H.: The role of absolute pitch memory in the oral transmission of folk songs. Empir. Music. Rev. **10**(3), 161–174 (2015)
24. Rodríguez-Lopez, M.E., Volk, A.: Selective acquisition techniques for enculturation-based melodic phrase segmentation. In: Proceedings of the 16th International Society for Music Information Retrieval Conference, pp. 218–224. Málaga (2015)
25. Schaffrath, H.: The essen associative code: a code for folksong analysis. In: Selfridge-Field, E. (ed.) Beyond Midi-The Handbook of Musical Codes, pp. 343–361. The MIT Press, Cambridge (1997)
26. Schepping, W.: Lied-und Musikforschung. In: Brednich, R.W. (ed.) Grundriss der Volkskunde, 2nd edn, pp. 467–492. Dietrich Reimer, Berlin (1994)
27. Smith, E.E., Sloman, S.A.: Similarity-versus rule-based categorization. Memory Cognit. **22**(4), 377–386 (1994)
28. Velarde, G., Weyde, T., Meredith, D.: An approach to melodic segmentation and classification based on filtering with the haar-wavelet. J. New Music Res. **42**(4), 325–345 (2013)
29. Volk, A., Chew, E., Margulis, E.H., Anagnostopoulou, C.: Music similarity: concepts, cognition and computation. J. New Music Res. **45**(3), 207–209 (2016)
30. Volk, A., de Haas, W.B., Van Kranenburg, P.: Towards modelling variation in music as a foundation of similarity. In: Proceedings of the 12th International Conference on Music Perception and Cognition, pp. 1085–1094. Thessaloniki, Greece (2012)
31. Volk, A., Van Kranenburg, P.: Melodic similarity among folk songs: an annotation study on similarity-based categorization in music. Music. Sci. **16**(3), 317–339 (2012)
32. Vos, J.L.M.: De spiegel der volksziel. Ph.D. thesis, Katholieke Universiteit Nijmegen, Nijmegen (1993)
33. Warris, M.: Andries Pels' Mengelzangen (1717). Master's thesis, Utrecht University (2013)
34. Wiering, F., Veltkamp, R.C., Garbers, J., Volk, A., Van Kranenburg, P.: Modelling folksong melodies. Interdiscip. Sci. Rev. **34**(2–3), 154–171 (2009)
35. Wiora, W.: Systematik der musikalischen Erscheinungen des Umsingens. Jahrbuch für Volksliedforschung **7**, 128–195 (1941)