Sonderdruck aus

Ruth-E. Mohrmann (Hg.)

Audioarchive

Tondokumente digitalisieren,
erschließen und auswerten

Louis Peter Grijp

From Audio in the Dutch Song Database to a Singing Map of Europe

1.   Introduction

The Meertens Institute for research and documentation of Dutch language and culture, part of the Royal Dutch Academy of Arts and Sciences in Amsterdam, has concentrated its activities in the field of Dutch song culture in the Center for Documentation and Research of Dutch Songs, presently a unit of the Department of Dutch Ethnology. The center is rooted in the old Dutch Folksong Archive (*Nederlands volksliedarchief*), founded in the mid-1950's on the model of the *Deutsches Volksliedarchiv* in Freiburg. Three of the collaborators of the center belong to the permanent staff of the institute (Martine de Bruin, Ellen van der Grijn, and myself); the others work on a temporary basis, such as a post-doc researcher, Ph.D. students, project documentalists, programmers, and trainees. They all contribute in one way or another to our central system, the Dutch Song Database (*Nederlandse liederenbank*). Since 2007 it has been available online at <www.liederenbank.nl>.

The Dutch Song Database, worldwide one of the biggest of its kind, unlocks about 150,000 songs from several repertories, including songs from the Middle Ages, songs from religious and profane printed songbooks from the sixteenth to the twentieth century, broadsides from the eighteenth until the early twentieth century, and field recordings from the twentieth century. The database includes 63,000 songs from the old card files of the *Volksliedarchief* that were added during a digitalization project in 1999–2002. Most of the songs are represented as metadata, i.e. first lines, tune indications, stanza forms, refrains, key words etc. For some of the songs full texts, melodies, scans, and/or sound files are supplied, and we are working hard to add more content. At the moment we are adding the texts of 100,000 songs in a project called Dutch Songs On Line, together with Utrecht University and the Digital Library of Dutch Literature DBNL in Leiden.

For the present volume the most relevant part of the database is the audio recordings. Apart from a huge number of dialect recordings, the Meertens Institute houses four musical audio collections. The most important is a collection of field recordings, called *Onder de groene linde* after a famous program on the Dutch national radio (7,000 songs). The audio of this collection is available on line, as a subset of the Dutch Song Database. Three additional collections have been digitized as well but are not yet on line: the field recordings by Harrie Franken, from the Kempen region on the Dutch-Belgian border (4,000 songs); field recordings from the Flemish VRT radio-archives made in Flanders by radio reporter Pol Heyns (from a pilot project of 250 songs); and about 1,500 tunes from eighteenth-

43

century bell clocks from the Netherlands, about which Marieke Lefeber is preparing a dissertation (see elsewhere in this volume).

2. The *Onder de groene linde* collection of field recordings

Ate Doornbosch (1926–2010) recorded elderly people, most of them living in the countryside, for his national radio program *Onder de groene linde* (Under the green linden). When he broadcast an old song, listeners wrote him letters with their own version of the song, or with other songs. Doornbosch visited them at their homes, recorded the songs, broadcast the recordings, received new letters, and so on. This snowball rolled on for thirty-six years (1957–1993) and yielded some 5,000 songs on magnetic tapes. Doornbosch collected the songs from all over the country, with a slight accent on the three northern provinces Groningen, Friesland, and Drenthe. During the project Doornbosch concentrated more and more on ballads in the sense of tragic story-telling songs, at the cost of other oral genres such as children's songs, broadsides including murder songs, labor songs, songs of sailors and fishermen, etc. To Doornbosch's collection we added the roughly 2,000 field recordings made by his predecessor Will Scheepers, who was inspired by ethnomusicologist Jaap Kunst and conducted her fieldwork in the period 1950–1964. The united collection is known under the name of the radio program, *Onder de groene linde.*

All 7,000 recordings have been digitized in the Meertens sound studio[1] and are available in mp3-format through the Dutch Song Database. They can be found using metadata such as first line and refrain, but also by the singer's name and the place where s/he was recorded. Because the field recordings have been integrated into the other song items in the database, one can also find Doornbosch's and Scheepers's field recordings through written sources. For example, if one has arrived at the edition of the famous ballad of the women's killer *Heer Halewijn* by J. F. Willems (printed in 1848), one can easily click on a list with all other versions of this song, whether in printed form, handwritten, or recorded as audio. Another option is to draw a geographic map of all the places where this song has been recorded. Fig. 1 shows the result. This "singing map" also shows differences in the melodies to which the ballad has been sung. For example, in the northeast of the country, in the province of Groningen, the song of *Heer Halewijn* was sung to a tune indicated on the map by a square. In the east, the same song text was sung to a different tune, indicated by a triangle.

Another option, which has not yet been fully developed, is the possibility of searching for similar melodies. This option uses advanced Music Information Retrieval technology, i.e. an algorithm that calculates the "distance" of the query

---

1   In 2001–02 the original recordings on magnetic tape were digitized following the PCM (Puls Code Modulation) principle. The audio has been stored in AIFF-format. The quality of the CD-Roms has been checked according to the Red Book standard.
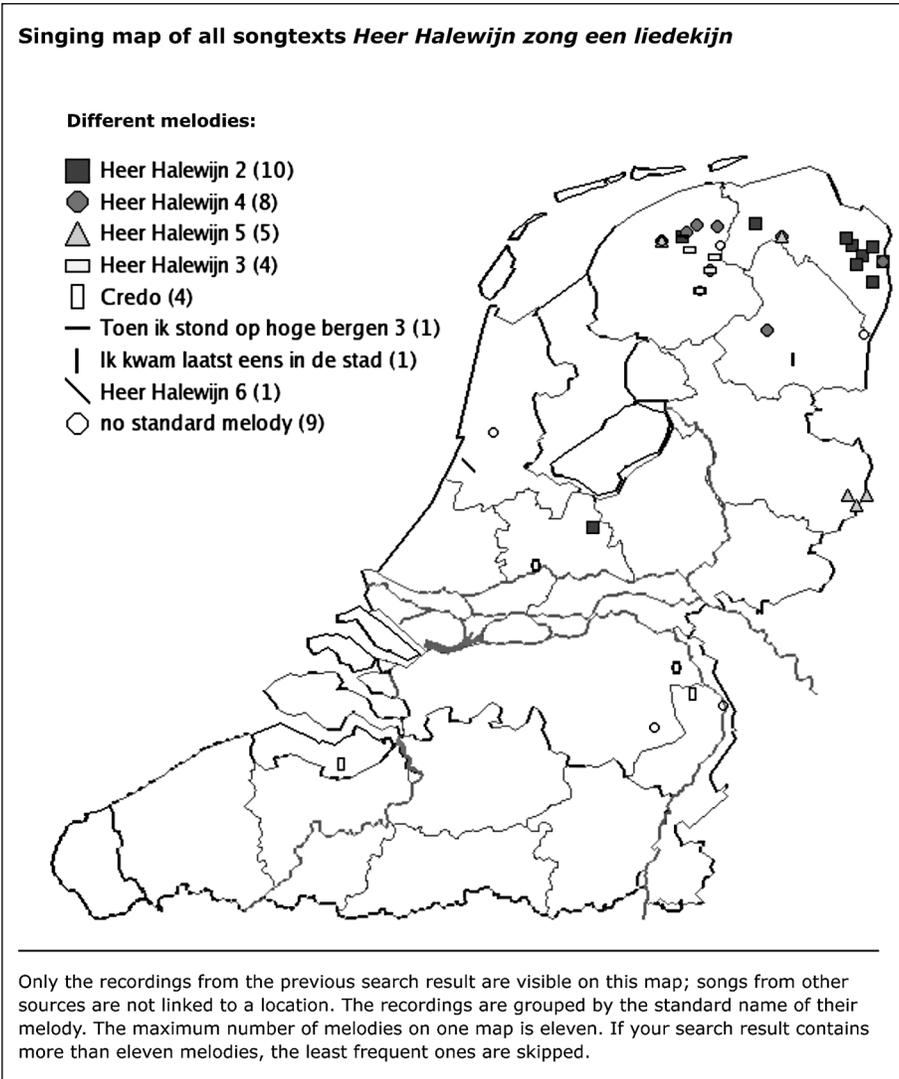
Singing map of all songtexts *Heer Halewijn zong een liedekijn*

Different melodies:

■ Heer Halewijn 2 (10)
● Heer Halewijn 4 (8)
▲ Heer Halewijn 5 (5)
▭ Heer Halewijn 3 (4)
☐ Credo (4)
— Toen ik stond op hoge bergen 3 (1)
| Ik kwam laatst eens in de stad (1)
＼ Heer Halewijn 6 (1)
◇ no standard melody (9)

Only the recordings from the previous search result are visible on this map; songs from other sources are not linked to a location. The recordings are grouped by the standard name of their melody. The maximum number of melodies on one map is eleven. If your search result contains more than eleven melodies, the least frequent ones are skipped.

Fig. 1:   Singing map of all recorded text versions of the song "Heer Halewijn".

to all melodies available in the database. This technique was developed during a project called WITCHCRAFT (see below).

In short, we did our best to make the audio recordings of *Onder de groene linde* as accessible as possible through the Dutch Song Database. Thus we tried to fulfill our task of making this important part of the Netherlands' immaterial heritage accessible both for scholars and for the general public. But are they really interested in these recordings? This is another task, convincing people, namely, of the

45

value and importance of such collections of field recordings.[2] To make the general public aware of this aspect of cultural heritage we produced a ten-CD-box with the most important ballad themes from *Onder de groene linde*, in total 163 recordings (2008).[3]

The box generated a lot of media exposure and eighty-four-year old Ate Doornbosch gave many interviews, especially for the radio – his own medium, after all. It received two prizes, the Jahrespreis der deutschen Schallplattenkritik (2009) and the Dutch prize Edison 2009 World & Jazz Special Releases. During all this excitement it became clear that the general public no longer knew the old songs, but people *did* remember the radio program, which had been very popular. The medium had become more important than the message.

3.　Research projects on *Onder de groene linde* material

We also wanted to confirm the value of the collection by encouraging scholarly research into it. After all, the Meertens Institute is primarily a research institute. It was thus useful to think of interesting research questions that might be addressed with this specifically oral material. One possibility I liked was the diversity of the voices. Imagine hundreds of male and female voices, high and low, beautiful and ugly, bright and hoarse, pure and out of tune ... Can such a diversity of voices be classified? Do people have individual singing styles and do they have their own way of remembering and changing songs?[4]

Another question is the problem of oral variation as a transmission system. *Onder de groene linde* contains dozens of different versions of the same texts and melodies. When two people sing the same song, there are always differences, in both text and music. But why are they different, and to what extent may they vary? What is the mechanism behind oral transmission? Could this be deduced from the recordings?

The WITCHCRAFT project mentioned above represents a first attempt to address such questions. We made them concrete by building a melody search engine. The central question was: Is it possible for a computer to recognize melodic variants caused by oral transmission, i.e. melodies that contain different notes, although we humans recognize the melodies as "the same"?

2　On this problem see L. P. Grijp, "Selling Songs: Dutch Results and Experiments." in T. Widmaier, E. John (eds.), *From Wunderhorn to the Internet: Perspectives on Conceptions of Folk Song and the Editing of Traditional songs* (B.A.S.I.S., 6) (Trier 2010), 234–45.

3　L.P. Grijp, I. van Beersum (eds.), *Onder de groene linde. 163 verhalende liederen uit de mondelinge overlevering, opgenomen door Ate Doornbosch e.a. / Under the green linden. 163 Dutch Ballads from the oral tradition recorded by Ate Doornbosch a.o.* (1 vol. + 9 CDs + 1 DVD) (Amsterdam/Hilversum 2008).

4　Ineke van Beersum, "Onder de altijdgroene linde II. Persoonlijke inbreng van zangers in de mondelinge liedcultuur." In L.P. Grijp, F. Willaert (eds.), *De fiere nachtegaal. Het Nederlandse lied in de middeleeuwen* (Amsterdam 2008), 311–29.

WITCHCRAFT stands for What Is Topical in Cultural Heritage – Content-based Retrieval Amongst Folk Tunes. The term "Cultural Heritage" is included in the acronym as a reference to the funder: CATCH, or Continuous Access to Cultural Heritage, from *Nederlandse Organisatie voor Wetenschappelijk Onderzoek* (NWO), the Dutch *Forschungsgemeinschaft* or research foundation. The idea is that university-based computer scientists should cooperate with cultural heritage institutions such as museums in order to make their collections more accessible to the public. For this purpose we regarded the Meertens Institute as a cultural heritage institute, because of its collections, and cooperated with the Department of Information and Computing Sciences of Utrecht University. Three computer scientists worked for four years (2006–10) to develop a search engine for melodies that had been transmitted orally.[5] The result is a prototype of such a search engine, based on the principle of string alignment.[6] The implementation of the prototype into the Dutch Song Database implies that for each transcribed melody from the *Onder de groene linde* collection it is possible to search for melodic variants. Pressing a button called "find similar melodies" yields a ranking list of musical transcriptions: the first melody resembles the original melody (the "query") most, the second one somewhat less, and so on.

Fig. 2 shows such a ranking list for the melody "Het was laatst op een zomerdag." The first, most similar melody is the query itself, recorded in the village of Gees in 1967. The second melody (recording Rockanje 1959) is identical, at least as far as we can see in this presentation; only the text is different. The third melody (recording Groningen 1968) begins with three different notes; the continuation is the same as the preceding melodies, but notated an octave lower. The following items of the list are also notated in the low register. This search engine works rather well and will become accessible for the public on the Internet in 2012.[7] The only shortcoming is that we don't know exactly why it does what it does. And when sometimes the algorithm fails, we cannot explain why. This is an interesting problem. As a next step we want to develop algorithms whose workings we understand, so that we might improve them.

This idea has resulted in two new projects, which both started in 2012. One is another CATCH project, called COGITCH (COgnition Guided Interoperability beTween Collections of musical Heritage). Its aim is to integrate cognitive knowledge into the melody search engine. Contrary to WITCHCRAFT it should work on audio, not on music notation. The idea is to identify so-called "hooks," the moments in a melody that catch our attention and make us recognize it. These hooks should be found by human annotation, preferably by crowd sourcing. This new

---

5   F. WIERING, R.C. VELTKAMP, J. GARBERS, A. VOLK, P. VAN KRANENBURG, L.P. GRIJP, "Modelling Folksong Melodies." *Interdisciplinary Science Reviews* 34/2-3 (2009), 154–71.

6   P. VAN KRANENBURG, *A Computational Approach to Content-Based Retrieval of Folk Song Melodies* (Diss. Utrecht University 2010).

7   A robust version of the prototype and the public user interface are produced in a separate project, WITCHCRAFT Plus (2009-2012, Meertens Institute, project manager Martine de Bruin).

Fig. 2:   Result list of a search for melodies similar to a song recorded in Gees in 1967. The first line is "Daar was laatst een meisje loos, meisje loos". The song turns out to be sung to a tune which is usually called "Het was laatst op een zomerdag".

audio music search engine should be able to handle folk music as well as popular music. Therefore, the collections of the Meertens Institute will be made interoperable with the sound archives of the Netherlands Institute for Sound & Vision in Hilversum, i.e. the Dutch media archives.[8]

The other new project is called Tunes & Tales. This was granted by the new Computational Humanities program of the Royal Dutch Academy of Arts and Sciences. The general idea is to stimulate the humanities to cooperate with computer science. Tunes & Tales has a more theoretical focus than COGITCH. The plan is to build a model that can simulate oral transmission including its variation, both in music and storytelling, so in two different fields of oral culture.[9] In the WITCHCRAFT project we found that motifs might be a clue for understanding stability in oral transmission. Therefore, an important part of the Tunes & Tales project will be work on the automatic recognition of motifs in melodies and folk tales.

We have combined these two projects, Tunes & Tales and COGITCH, together with a third project on automatic recognition of folktale types (called FACT),[10] in an e-Laboratory for Oral Culture.[11] We hope in this way to create synergy by having so many investigators – at least ten – and their promoters cooperating on similar problems.

4. Singing Map of Europe

I would like to conclude this overview of projects on oral collections with something that is not yet a project but rather still a vague idea, or a dream: a Singing Map of Europe. Exactly what this map should look like, let alone how to realize it are still unknown. One option would be that the map should make visible the stylistic differences in folk musics. What is the difference between Irish and Polish music, or between Spanish and French music? The Hungarian mathematician and folk musician Zoltán Juhász has proposed an interesting method for visualizing such differences.[12]

Another problem that could be expressed in a Singing Map of Europe is that of the international diffusion of music. For instance, the origins of most of the melodies of *Onder de groene linde* are unknown; did they come from Germany, from France, or from elsewhere? The same question may be asked for many European regions and countries. Such a map showing international musical influences

---

8   <www.cs.uu.nl/research/projects/music/Cogitch-online.pdf> (25 May 2012).
9   <ehumanities.nl/tunes-tales> (25.5.2012).
10   **F**olktales **A**s **C**lassifiable **T**exts, see <www.nwo.nl/nwohome.nsf/pages/NWOP_6CCC3L_
     Eng> (25.5.2012).
11   <www.elab-oralculture.nl> (25.5.2012).
12   Z. Juhász, "A Systematic Comparison of Different European Folk Music Traditions Unsing
     Self Organizing Maps." *Journal of New Music Research* 35/2 (2006), 95–112.

should probably be drawn for different historical periods. This implies that historical sources such as song books and fiddle books should also be incorporated.

Such a dream can only be realized by the international cooperation of musicologists, ethnologists, and computer scientists.

(The conference took place on 15–16 September 2011.)