

Meertens

Online Reports

Meertens Instituut, Amsterdam, Netherlands
www.meertens.knaw.nl

ISSN 2352-2135

The Meertens Tune Collections

*Peter van Kranenburg, Martine de Bruin,
Louis P. Grijp, Frans Wiering*

2014-1

The Meertens Tune Collections

Peter van Kranenburg,¹ Martine de Bruin,¹ Louis P. Grijp,¹ and Frans Wiering²
¹Meertens Institute, Amsterdam, ²Utrecht University

December 2014

15 February 2015: An additional remark has been added to Section 3.8.

Contents

1	Introduction	2
2	Overview of Contents	2
3	General Remarks	3
3.1	Distribution and Availability	3
3.2	Versioning	3
3.3	File Types	3
3.4	WCE Files	4
3.5	wce2krm	6
3.6	Metadata	6
3.7	Songs, Strophes and Voices	7
3.8	Hard Breaks and Soft Breaks	7
3.9	File Naming	7
3.10	The Concepts of Tune Family and Text Family	7
4	MTC-OGLAUDIO and MTC-OGLSCANS	8
4.1	MTC-OGLAUDIO	8
4.2	Metadata Table MTC-OGLAUDIO-recordings	8
4.3	Metadata Table MTC-OGLAUDIO-singers	9
4.4	MTC-OGLSCANS	9
5	MTC-FS	9
5.1	Content Files	9
5.2	Metadata Table MTC-FS	11
5.3	Metadata Table MTC-FS-sources	12
5.4	Metadata Table MTC-FS-singers	12
6	MTC-INST	13
6.1	Content Files	13
6.2	Metadata Table MTC-INST	15
6.3	Metadata Table MTC-INST-sources	16
7	MTC-ANN	16
8	MTC-LC	16
9	License and Attribution	17

Name	Description	File types	Size
MTC-OGLAUDIO	Collection <i>Onder de groene linde</i> : 7,178 audio recordings collected by Dutch field workers during the 1950s-1980s.	mp3	20 GB
MTC-OGLSCANS	Scans of 3,754 transcriptions of recordings from <i>Onder de groene linde</i> as made during the 1950s-1980s. The music is hand-written, the lyrics are mostly typed.	jpg	1.6 GB
MTC-FS	4,120 digitally encoded strophes from vocal folk songs both from <i>Onder de groene linde</i> (2,503) and from various related written sources (1,617).	**kern, midi, LilyPond, png, pdf, txt	250 MB
MTC-INST	2,368 digitally encoded instrumental tunes from 18th-century Dutch manuscripts and printed scores.	**kern, midi, LilyPond, png, pdf	130 MB
MTC-ANN	Annotated Corpus: 360 melodies used in various publications.	**kern, midi, png, pdf	1.5 MB
MTC-LC	Large Corpus: 4,830 melodies used in various publications.	**kern, midi	20 MB

Table 1: Contents of the Meertens Tune Collections.

1 Introduction

The Meertens Institute of the Royal Netherlands Academy of Arts and Sciences (Amsterdam, Netherlands) hosts a large collection of Dutch songs. These are accessible via the online interface of the *Nederlandse Liederbank*,¹ which contains meta-data of c. 170,000 songs, full texts of c. 50,000 songs, and music notation for c. 9,000 songs.² The online interface offers search functionality and access to songs on an individual basis. Each song record contains a rich collection of meta-data fields.

With the *Meertens Tune Collections* (MTC), the Meertens Institute releases several collections of digitized songs. This release is aimed to serve those who wish to use entire data-sets instead of individual songs. The release includes collections that already have been used in various research papers.

The availability of digitized musical material is important for several disciplines. From a musicological perspective, the study of these melodies is relevant for understanding transmission and dissemination of folk tunes. For research in the field of Music Information Retrieval (Downie, 2003), the availability of large annotated data sets is of crucial importance to test algorithms and tools, such as segmentation algorithms, melodic similarity measures, and pattern discovery algorithms. In the area of music cognition, analysis of the songs may yield insight in melodic segmentation, melodic similarity and the stability and variability of motifs, intervals and rhythms.

This report describes contents, background and properties of these sets.

2 Overview of Contents

The main substance of the initial release of the MTC consists of two collections of melodies: 4,120 strophes from vocal folk songs and 2,368 early-modern instrumental tunes. The folk songs are transcriptions from both twentieth-century field recordings and printed books that cover roughly the same repertoire. The instrumental tunes are taken from eighteenth-century, mostly monophonic Dutch sources, both prints and manuscripts. Table 1 provides an overview of the collections that are part of MTC along with a short description of the contents. Each collection is completed by a rich set of meta-data. The individual collections will be described in the following sections.

¹Dutch Song Database: <http://www.liederbank.nl>

²These are the numbers at the end of 2014. Data and metadata are continually added.

3 General Remarks

3.1 Distribution and Availability

MTC is provided as set of gzipped tar files. Each file contains one collection, with exception of MTC-OGLAUDIO, which has been split into files with size of c. 4GB each and a separate file for the metadata. Each tar file expands into a directory with name and version number of the collection, e.g., “MTC-LC-1.0”. This top-level directory contains subdirectories for each of the representations (png, pdf, krn, etc.), and for the metadata. The files can be downloaded from <http://www.liederenbank.nl/mtc>.

Long-term access is guaranteed by the Meertens Institute. As trusted data repository, the Meertens Institute obtained the Data Seal of Approval.³ The Meertens Tune Collections are available as: The archives of the Meertens Institute, digital collection nr. 51.

3.2 Versioning

Each of the collections has its own version number. In the initial release all have version 1.0. We anticipate future addition of songs and metadata, especially for the folk song (FS) and instrumental (INST) collections. These enlarged collections will get their own version number. By referring to the name and version number of a certain collection in research publications, the exact data that was used for the research can be retrieved by others. Permanent access to, and sustainable storage of all current and previous versions are guaranteed by the Meertens Institute.

3.3 File Types

The following file types are used in MTC:

jpg jpeg image compression format
<http://www.jpeg.org/>

png portable network graphics image format
<http://www.libpng.org/>

pdf portable document format
<https://www.adobe.com/devnet/pdf/>

krn Humdrum **kern representation
<http://www.humdrum.org/>

ly LilyPond source file
<http://www.lilypond.org/>

wce witchcraft editor file
See section 3.4.

txt UTF-8 plain text file
<http://www.unicode.org/>

mp3 MPEG1 layer III audio compression format
<http://mpeg.chiariglione.org/>

mid midi sequence
<http://www.midi.org/>

csv UTF-8 text file containing comma-separated values
<http://tools.ietf.org/html/rfc4180>

All text files (krn, ly, wce, txt, csv) have UNIX line endings and character encoding UTF-8.

³<http://www.datasealofapproval.org>

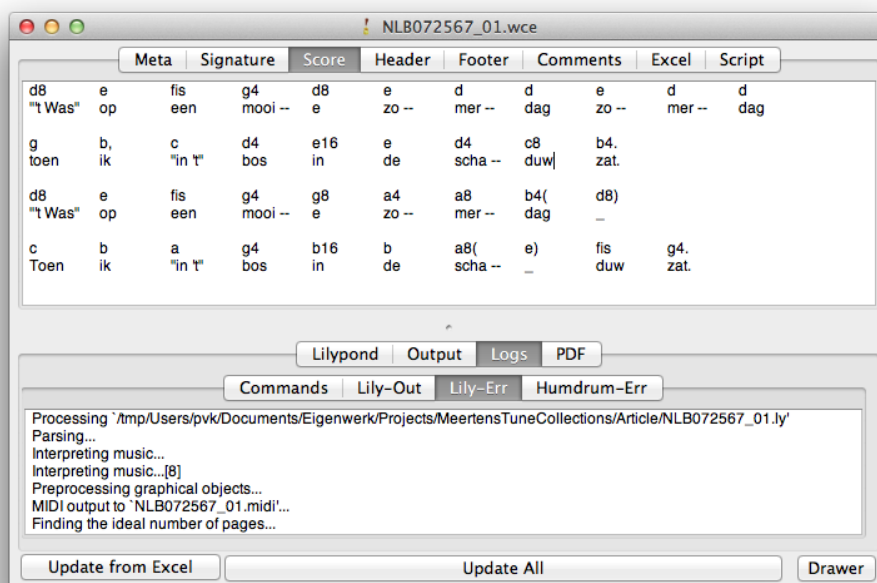


Figure 3.1: User interface of the witchcraft editor.

3.4 WCE Files

The witchcraft editor was developed within the WITCHCRAFT-project (Wiering et al., 2009) as a Mac OS X application to enter melodies (see Figure 3.1). The input is stored in witchcraft editor files (wce) which are included in the distribution. Other file types such as `**kern`, LilyPond, lyrics, score (png and pdf), and midi files are derived from the wce files. The user interface of the editor combines a number of text input fields in a tabbed interface. There are input fields for key, meter, upbeat, notes, title, lyrics, etc. The contents of all musical fields are encoded according to a subset of the input encoding for LilyPond. The contents of the fields are stored in the wce files, according to the Apple property-list XML format.⁴

The following fields are included in the wce-files:

NLBController-recordIDTextField Record number of the song in the Database of Dutch Songs.

SignatureController-stropheNumberTextField Number of the strophe/voice (first is 1).

SignatureController-titleTextField Title of the song.

SignatureController-footerTextView If a LilyPond source file is assembled from the contents of the wce-fields, the contents of this field should be appended at the end of the resulting lilypond source file. Currently, the only use is to add footnotes to a score.

SignatureController-isMeterInvisibleSwitch True if the melody has a free meter.

SignatureController-keyTextField Key in LilyPond encoding (e.g., `c \major`, `d \dorian`, `a \minor`, etc).

SignatureController-partialTextField Length of upbeat (if any) in LilyPond duration encoding (e.g., `8*3`).

⁴<http://www.apple.com/DTDs/PropertyList-1.0.dtd>

SignatureController-tempoTextField Tempo in LilyPond encoding (e.g., 4=120).

SignatureController-relativeTextField The pitch the first note is related to according to the LilyPond `\relative` command (e.g., g').

SignatureController-timeTextField The (initial) meter of the song in LilyPond encoding (e.g., $9/8$). A meter is always provided in this field, also in the case of a melody in free meter. In those cases the meter can be ignored in processing.

TextProcessingDefaults-inputTextView A text field containing the melody and lyrics, as shown in Figure 3.1. Melody and lyrics are interwoven. The melody of each phrase is directly followed by the lyrics of the phrase (if any). Phrases are separated by a blank line. For the encoding of the melody a subset of LilyPond's music encoding is used. The encoding of the lyrics is also according to the LilyPond format. In the Figure, syllables and notes are separated by tabs for readability. This is not required.

The exact subset of the LilyPond encoding that is used is defined in the input files for the lexical analyzer as included in the source distribution of `wce2krn: lilylexer.ll` and `textlexer.ll` (see Section 3.5). The encoding of the bar lines is according to LilyPond versions ≤ 2.16 . When using more recent versions of LilyPond, these should be converted.

The following LilyPond definitions are used in the melody encodings:

```
sb = {\breathe}
x = {\once\override NoteHead #'style = #'cross }
gl=\glissando
ficta = {\once\set suggestAccidentals = ##t}
fine = {\once\override Score.RehearsalMark #'self-alignment-X = #1
  \mark \markup {\italic{Fine}}}
dc = {\once\override Score.RehearsalMark #'self-alignment-X = #1
  \mark \markup {\italic{D.C.}}}
dcf = {\once\override Score.RehearsalMark #'self-alignment-X = #1
  \mark \markup {\italic{D.C. al Fine}}}
dcc = {\once\override Score.RehearsalMark #'self-alignment-X = #1
  \mark \markup {\italic{D.C. al Coda}}}
ds = {\once\override Score.RehearsalMark #'self-alignment-X = #1
  \mark \markup {\italic{D.S.}}}
dsf = {\once\override Score.RehearsalMark #'self-alignment-X = #1
  \mark \markup {\italic{D.S. al Fine}}}
dsc = {\once\override Score.RehearsalMark #'self-alignment-X = #1
  \mark \markup {\italic{D.S. al Coda}}}
pv = {\set Score.repeatCommands = #'((volta "1"))}
sv = {\set Score.repeatCommands = #'((volta "2"))}
tv = {\set Score.repeatCommands = #'((volta "3"))}
qv = {\set Score.repeatCommands = #'((volta "4"))}
xv = {\set Score.repeatCommands = #'((volta #f))}
```

The following settings are needed in the LilyPond source:

```
\set melismaBusyProperties = #'()
\override Score.MetronomeMark #'transparent = ##t
\override Score.RehearsalMark #'break-visibility = #(vector #t #t #f)
```

Normally in LilyPond if notes are slurred, one lyrics syllable is aligned with the entire group of slurred notes. The first setting (`melismaBusyProperties`) disables this behavior. This means that for each subsequent note in a melisma, an underscore is needed in the lyrics input.

3.5 wce2krn

The ****kern** files, LilyPond source files, and lyrics files have been generated from the wce source files using `wce2krn`, a converter for the wce format that has been developed by the Meertens Institute. The C++ code is available from the MTC homepage.⁵ For MTC version 1.0, `wce2krn` 1.64 was used. Usage is as follows:

```
Usage: wce2krn [-k] [-s] [-l] [-w] [-v] [-h] [-p] [-e] [wcefile]
Reads wce-file and generates various other formats.
-k: generate kernfile.
-r: generate LilyPond.
-w: generate LilyPond for pngs in the Meertens Tune Collections.
-l: generate file(s) with only the lyrics.
-s: generate a separate file for each phrase.
-p: print the contents of the music to stdout with annotations.
-e: suppress log messages.
-v: print version number and exit.
-h: print this help message and exit.
If no filename is given, or '-', standard input and output will be used.
In this case, '-s' does not have any effect.
```

It is important to note that `wce2krn` only converts a subset of the LilyPond input to the ****kern** representation. Ornaments and glissandi are skipped. All bar lines are converted to the single bar line. Structural indications (such as Da Capo, Dal Segno, etc) are skipped. Additional text (LilyPond `\markups`) are skipped as well. This has mainly consequences for MTC-INST, since repeat bar lines, structural indications and ornaments are absent in the vocal folk songs.

Lyrics are converted to a separate spine in the Humdrum files. Information about meter, key, and phrasing is converted into the ****kern** files. All phrase endings are marked by a fermata. Only the hard breaks (see Section 3.8) result in ****kern** phrase markings using braces: { }. The keys and the key signatures are indicated using the standard ****kern** tandem interpretations. Church modes are designated according to the following overview:

- *C:** C major
- *c:** C minor
- *C:mix** C mixolydian
- *C:lyd** C lydian
- *C:ion** C ionian
- *c:dor** C dorian
- *c:phr** C phrygian
- *c:aeo** C aeolian
- *c:loc** C locrian

3.6 Metadata

Metadata are provided in text-files that can be found in the directory `metadata` of each collection. The character encoding is UTF-8, and the line endings are in UNIX style for all text files. For most collections, the metadata are contained in tables that are formatted as comma-separated-values (csv). Most values are enclosed in double quotes. Field-headings are provided in separate files. Sometimes a field contains a list, for example if there is more than one singer in a recording, or if there is more than one author of a

⁵<http://www.liederenbank.nl/mtc>

song book. In that case, the pipe symbol, |, is used as separation marker. Names are formatted as: Last Name, First name. As an example:

```
Wolsey, Boy|Waldorp, Jan
```

means Boy Wolsey and Jan Waldorp.

3.7 Songs, Strophes and Voices

Most files with folk song items contain a transcription of the first strophe of the song; but sometimes, if the singer had initial problems recalling the melody correctly, the transcriber chose to transcribe another strophe next to the first as well, or even to omit the first strophe at all. Also, in case of considerable variation between strophes of one song, more than one strophe could have been transcribed.

In the instrumental sources, sometimes a second voice — often a bass part — is notated as well. In many of these cases both the first and second voice have been digitized.

Since each file only contains one strophe or one voice, there are more files than songs in most of the collections.

3.8 Hard Breaks and Soft Breaks

All melodies have been segmented into phrases by inserting blank lines at the phrase boundaries (*hard breaks*) in the manual input in the witchcraft editor (see Section 3.1). The exact positions of phrase boundaries are in most cases determined by the transcriber or encoder, which implies a certain level of subjectivity.

For vocal songs, sometimes the source provides a partitioning of the song into phrases. But in many cases decisions had to be made by the transcribers and encoders. For these decisions textual cues such as rhyme and closed syntactical units could be used.

For the instrumental melodies, there are no textual cues. Therefore, encoders were instructed only to insert hard breaks at clear structural divisions such as repeat bar lines, which, however, results in relatively long segments. To have a finer segmentation as well, an additional level of phrasing was added by introducing *soft breaks*, encoded as `\sb`. These were intended to indicate points of perceived melodic closure according to the musical intuition of the encoder. As cue the encoders were instructed to take those places in the melody that serve as a ‘natural’ breathing point during singing or playing. However, in practice, the various encoders that contributed to the MTC did not apply these instructions consistently, or not at all. Therefore, a clear distinction in meaning between encoded soft breaks and hard breaks should not be expected, and hard breaks may occur at positions other than clear structural divisions. Furthermore, because of the inherent ambiguity of melodic segmentation, the breaks should be regarded as one possible, musically valid, way to divide the melody into smaller units. In the provided scores (png and pdf), the soft breaks have been visualized as breathing marks (see e.g., Figure 6.1).

3.9 File Naming

The basenames of content files have the following format: `NLBxxxxxx_yy`, where `xxxxxx` is the record number of the song in the Dutch Song Database, preceded with leading zeros if necessary, and `yy` is a serial number which has different meanings for the different collections. For vocal folk songs `yy` indicates the number of the strophe that is in the file. For instrumental melodies `yy` indicates the voice: 01 for first voice and 02 for second voice, often a bass part. For the scans, `yy` indicates the page number of the transcription. `yy` is `all` if all strophes are present in the file, which is only the case for the audio recordings in MTC-OGLAUDIO. For example, `NLB074344_01` designates the first strophe of the song with record number 74344. `NLB074344_all.mp3` is the audio recording that contains all strophes of this song. The same format is used as identifier in the metadata tables.

3.10 The Concepts of Tune Family and Text Family

Two concepts are used to group songs: *tune family*, and *text family*. Because of the variability due to the process of (oral) transmission, almost no two instances of the same song are exactly identical in a literal

sense. In our collections, we rather have groups of variants, both for texts and melodies. For variants of the same melody, we use the concept of *tune family*.⁶ By analogy, to indicate a group of songs with the same textual contents, we use the concept of *text family*. These songs are designated in the metadata to belong to the same text family. The categorization in both tune and text families has been done over the course of many years by the collection specialists at the Meertens Institute, and is subject to constant revision. Therefore, tune family labels are not guaranteed to be consistent across different versions of the MTC. Three levels of confidence are used to assign a tune family label to a melody: *strong*, *neutral*, and *weak*. Sometimes a melody is clearly related to a tune family, but only partially, or with specific differences. In that case it is marked as *variant* of the tune family, and the confidence is designated as *weak*.

4 MTC-OGLAUDIO and MTC-OGLSCANS

The two collections MTC-OGLAUDIO and MTC-OGLSCANS contain the songs from *Onder de groene linde* (see Grijp, 2008; Doornbosch, 1987). These folk songs were sung by ordinary people during work or social activities. They were part of oral culture: the songs — especially the melodies — were primarily learned by listening and participation. In this process of oral transmission, the songs underwent continuous change.

As a consequence of labour mechanization and the introduction of the radio, this form of traditional singing has completely disappeared from the Netherlands. That we still have access to this tradition is the result of a long-term effort to transcribe the songs into musical notation or to record them on tape. The first to collect written transcriptions of folk songs from the Dutch language area was the Fleming Jan Frans Willems (1848); among the first Dutch sound recordings were those made by Will Scheepers in the early 1950s. Her work was continued by Ate Doornbosch. In total more than 7,000 field recordings were made. For c. 5,000 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. Listeners were encouraged to contact Doornbosch if they knew more about the songs that were played. Doornbosch would then record their version and broadcast it. In this manner, a collection was created that documents not only an appealing aspect of Dutch cultural heritage, but also the textual and melodic variation that results from oral transmission. Doornbosch was specifically interested in ballads: old strophic songs of considerable length in which a tragic story is told.

4.1 MTC-OGLAUDIO

The collection MTC-OGLAUDIO contains all of the recordings of *Onder de groene linde* in MP3-encoded audio representation. The collection consists of the mp3 files and two tables with metadata: recordings and singers.

4.2 Metadata Table MTC-OGLAUDIO-recordings

filename The name of the mp3 file (string).

songid The record number of the song in the Database of Dutch Songs (integer).

date_of_recording Date of recording (DD-MM-YYYY).

place_of_recording Place of recording, in most cases the name of the municipality (string).

latitude latitude of place of recording (float).

longitude longitude of place of recording (float).

⁶The concept of *tune family* has been introduced and defined by Samuel Bayard (1950) and has been used in American folklore studies. In Bayard's definition it is the (presumed) 'genetic' relation between the melodies that establish a tune family. In Europe, the concept of *type* (Typus) has been used more often, be it with a multitude of definitions. An important example is the edition of melody types of German folk song by Suppan and Stief (1976).

fieldworker the fieldworker who did the recording, if known ({Ate Doornbosch, Will D. Scheepers, \emptyset }).

singer_id_s identifiers of one or more singers (list of integers).

tunefamily name of the tune family, the song belongs to (string).

tunefamily_id short identifier of the tune family (ASCII string).

variation whether the tune should be considered a more distant variant within the tune family (variant) or a full member (\emptyset).

confidence the confidence with which the tune was assigned to the tune family ({weak, neutral, strong, \emptyset }).

4.3 Metadata Table MTC-OGLAUDIO-singers

singer_id identifier of the singer (integer).

year_of_birth year of birth of the singer (integer).

place_of_birth place of birth of the singer, in most cases the name of the municipality (string).

latitude latitude of place of birth.

longitude longitude of place of birth.

4.4 MTC-OGLSCANS

Around 3,750 of the recordings in *Onder de groene linde* were transcribed into musical notation in the period 1950–1990. In most cases, the lyrics were transcribed using a typewriter, and the music notation of the melodies was added in handwriting. In 2006 all handwritten transcriptions were digitized as images (jpeg) and subsequently made publicly available in the Dutch Song Database. The collection MTC-OGLSCANS contains scanned images (jpg) of all transcriptions.

5 MTC-FS

MTC-FS consists of digitized melodies both from the *Onder de groene linde* sound recordings, and from related printed songbooks. We selected 2,503 melodies from *Onder de groene linde* for digitization. The selection was partially based on an earlier classification of the songs into ballad types.⁷ Furthermore, a number of recordings of children’s songs were added to the set.

In addition to the transcriptions of sound recordings, 1,617 melodies are included that are taken from printed songbooks such as Jan Frans Willems, *Oude Vlaemsche Lieder* (1848), Florimond Van Duyse, *Het oude Nederlandsche lied* (1908), and Jop Pollmann and Piet Tiggers, *Nederland’s volkslied* (1941). We selected those songs from the song books that also appear among the recordings of *Onder de groene linde*. In the metadata of the collection, all variants of a melody have got the same tune family label, and all variants of a text have got the same text family label.

The collection consists of **kern encodings, midi sequences, LilyPond source files, images of the scores (png and pdf), text files with syllabilized lyrics, and three metadata tables: songs, singers and sources.

5.1 Content Files

The manual input of the melodies is contained in the witchcraft editor (wce) files (see Section 3.4). The Humdrum **kern files, the LilyPond source files, and the lyrics files have been generated from the wce

⁷In our current terminology: text families.

't Was op een mooi - e zo - mer - dag zo - mer - dag
toen ik in 't bos in de scha - duw zat.
't Was op een mooi - e zo - mer - dag
Toen ik in 't bos in de scha - duw zat.

NLB072567_01 - <http://www.liederenbank.nl/liedpresentatie.php?zoek=72567>

Figure 5.1: Example of a folk song, 't was op een mooie zomerdag, recorded on 07-05-1963 in Coevorden.

files. The midi files are generated from the `**kern` files using `humdrum extra` command `hum2mid`,⁸ and the pdf and png representations of the scores have been generated from the LilyPond source files.

As an example, the following listing shows the `**kern` file for the melody that is shown in Figure 6.1. The phrase endings are indicated with a fermata sign (`;)` and by a phrase indication in `**kern` using curly brackets, and by a global comment (`!! verse x`) indicating the number of the phrase.

```

1  !! Published by the Meertens Institute (http://www.meertens.knaw.nl)
2  !! Part of the Database of Dutch Songs (NLB) (http://www.liederenbank.nl)
3  !! NLB identifier: NLB072567_01
4  !! NLB record number: 72567
5  !! NLB strophe/voice number: 1
6  **kern **text
7  *staff1 *staff1
8  *M6/8 *
9  *MM120 *
10 *k[f#] *
11 *G: *
12 !! verse 0
13 {8d 't Was
14 8e op
15 8f# een
16 =1 =1
17 4g mooi-
18 8d -e
19 8e zo-
20 8d -mer-
21 8d -dag
22 =2 =2
23 8e zo-
24 8d -mer-
25 8d;} -dag
26 !! verse 1
27 {8g toen
28 8B ik
29 8c in 't
30 =3 =3
31 4d bos
32 16e in
33 16e de
34 4d scha-
35 8c -duw
36 =4 =4

```

⁸<http://extras.humdrum.org/>

```

37 4.B;} zat.
38 !! verse 2
39 {8d 't Was
40 8e op
41 8f# een
42 =5 =5
43 4g mooi-
44 8g -e
45 4a zo-
46 8a -mer-
47 =6 =6
48 (4b -dag
49 8dd);} .
50 !! verse 3
51 {8cc Toen
52 8b ik
53 8a in 't
54 =7 =7
55 4g bos
56 16b in
57 16b de
58 (8a scha-
59 8e) .
60 8f# -duw
61 =8 =8
62 4.g;} zat.
63 ==|! ==|!
64 *- *-
65 !! produced by wce2krn 1.64 (released on 7 June 2014)

```

5.2 Metadata Table MTC-FS

This metadata table contains metadata for both MTC-FS and MTC-INST. It is included in both collections. Therefore, not all fields contain values for each song. There are, for example, fields that relate to audio recordings, such as recording date. These are empty for songs from written sources.

filename the basename of the file the metadata record refers to (string).

songid the record number of the song in the Database of Dutch Songs (integer).

source_id identifier of the source of the song - lookup in the table with sources (integer).

serial_number the serial number of the song in the source (string).

page the page number of the song in the source (string).

singer_id_s identifiers of one or more singers (list of integers).

date_of_recording date of recording (DD-MM-YYYY).

place_of_recording place of recording, in most cases the name of the municipality (string).

latitude latitude of place of recording (float).

longitude longitude of place of recording (float).

title title of the song (string).

firstline first line of the lyrics (string).

textfamily_id identifier of the text family the song belongs to (integer).

tunefamily_id short identifier of the tune family (ASCII string).

tunefamily name of the tune family the song belongs to (string).

type type of the song (`{folksong, instrumental}`).

voice_strophe_number serial number of the strophe or voice that is in the file (integer).

voice_strophe whether the value of `voice_strophe_number` is a voice (in case of instrumental music) or a strophe (in case of folk songs) (`{voice, strophe}`).

image_filename_s list of filenames of images in MTC-OGLSCAN that belong to the song (list of strings).

audio_filename filename of the audio recording in MTC_OGLAUDIO that belongs to the song (string).

variation whether the tune should be considered a more distant variant within the tune family (`variant`) or a full member (`0`).

confidence the confidence with which the melody was assigned to the tune family (`{weak, neutral, strong, 0}`).

5.3 Metadata Table MTC-FS-sources

This table is also included in MTC-INST.

source_id identifier of the source (integer).

title title of the source (string).

author author(s) of the source (list of string).

place_publisher place(s) and publisher(s) of the source (list of strings). Place and publisher are separated by a colon ("`:`").

dating (approximate) dating of the source (string).

sorting_year year that can be used for sorting (integer).

type type of source (`{manuscript, print, audio}`).

copy_used library siglum of the copy that was used for digitization (string).

scan_url url of scanned images of the source, in case the source is publicly available online (string).

This table is also included in MTC-INST.

5.4 Metadata Table MTC-FS-singers

singer_id identifier of the singer (integer).

year_of_birth year of birth of the singer (integer).

place_of_birth place of birth of the singer, in most cases the name of the municipality (string).

latitude latitude of place of birth.

longitude longitude of place of birth.

The structure of this table is the same as the MTC-OGLAUDIO-singers (see Section 4.3), but here only the singers of transcribed songs are included.

Menuetto



Voortekening ontbreekt in de eerste regel.

1) Halve noot met punt in bron.

NLB177137_01 - <http://www.liederenbank.nl/liedpresentatie.php?zoek=177137>

Figure 6.1: Score for NLB177137_01, which is a “Menuetto” from a 18th century untitled music manuscript (The Hague, Nederlands Muziek Instituut, 4 G 86). In the footnote, editorial remarks are presented.

6 MTC-INST

The instrumental tune collection, MTC-INST, includes 2,368 popular melodies from 18th century Dutch sources. Most of the tunes have been printed or written in monophonic form, without accompaniment. The core of the collection consists of 996 tunes from the printed series *Oude en nieuwe Hollandse Boeren Lietjes en Contredansen* (Old and New Dutch Peasant Songs and Countrydances, Amsterdam: Estienne Roger 1701–1714, 13 vols.). Although many of these dance and song tunes may originate in low culture, others have been identified as melodies from operas by French court composers such as J.B. Lully. With his ‘Peasant Songs’ Roger must have aimed at well-to-do amateurs, playing the violin, flute or oboe. The same tunes are found in Dutch music manuscripts from the 18th century, but in slightly varied form. This may be explained from oral transmission. MTC-INST contains tunes from 18 printed books and 8 manuscripts. In general, the tunes are short and simple. Characteristic forms are menuet and marche. This repertoire may be regarded as the Dutch equivalent of John Playford’s *The English Dancing Master* (1651, many reprints up to 1728).

MTC-INST consists of **kern encodings, midi sequences, LilyPond source files, images of the scores (png and pdf), and two metadata tables: songs, and sources.

6.1 Content Files

The manual input of the melodies is contained in the witchcraft editor (wce) files (see Section 3.4). The Humdrum **kern files and LilyPond source files have been generated from the wce files. The midi files are generated from the **kern files using humdrum extra command `hum2mid`,⁹ and the pdf and png representations of the scores have been generated from the LilyPond source files.

The **kern files contain a subset of the manual encoding. Ornaments, non-standard bar lines, footnotes and structural indications (such as *volta*s) are not included in the **kern files. As an example, the following listing shows the **kern file for the melody that is shown in Figure 6.1. All phrase endings, including soft breaks, are indicated with a fermata sign (;), while only the ‘hard’ breaks result in a phrase indication in **kern using curly brackets and a global comment (!! *verse* x) indicating the number of the phrase.

```
1 !! Published by the Meertens Institute (http://www.meertens.knaw.nl)
2 !! Part of the Database of Dutch Songs (NLB) (http://www.liederenbank.nl)
3 !! NLB identifier: NLB177137_01
4 !! NLB record number: 177137
5 !! NLB strophe/voice number: 1
6 **kern
7 *staff1
8 *M3/4
9 *MM120
10 *k[b-]
11 *F:
```

⁹<http://extras.humdrum.org/>

```

12  !! verse 0
13  =1
14  {4f
15  8f
16  8g
17  8a
18  8b-
19  =2
20  4cc
21  8cc
22  8dd
23  8ee
24  8ff
25  =3
26  4gg
27  16bb-
28  16aa
29  16gg
30  16ff
31  16ee
32  16ff
33  16gg
34  16ee
35  =4
36  2ff;
37  4cc
38  =5
39  4dd
40  8dd
41  8cc
42  4b-
43  =6
44  2cc
45  (12a
46  12b-
47  12cc)
48  =7
49  4cc
50  8dd
51  8cc
52  8b-
53  8a
54  =8
55  2g;}
56  !! verse 1
57  {8g
58  8a
59  =9
60  4b-
61  4b-
62  4a
63  =10
64  4.cc
65  (8a
66  8b-
67  8cc)
68  =11
69  4cc
70  8dd
71  8cc
72  8b-
73  8a
74  =12
75  2g;
76  4gg
77  =13
78  4aa

```

```

79 4gg
80 4ff
81 =14
82 4ee
83 4cc
84 8gg
85 8aa
86 =15
87 8bb-
88 8aa
89 8gg
90 8ff
91 16ee
92 16ff
93 16gg
94 16ee
95 =16
96 2ff; }
97 ==|!
98 *-
99 !! produced by wce2krn 1.64 (released on 7 June 2014)

```

6.2 Metadata Table MTC-INST

This metadata table contains metadata for both MTC-FS and MTC-INST. It is included in both collections. Therefore, not all fields contain values for each song. There are, for example, fields that relate to audio recordings, such as recording date. These are empty for songs from written sources.

filename the basename of the file the metadata record refers to (string).

songid the record number of the song in the Database of Dutch Songs (integer).

source_id identifier of the source of the song - lookup in the table with sources (integer).

serial_number the serial number of the song in the source (string).

page the page number of the song in the source (string).

singer_id_s identifiers of one or more singers (list of integers).

date_of_recording date of recording (DD-MM-YYYY).

place_of_recording place of recording, in most cases the name of the municipality (string).

latitude latitude of place of recording (float).

longitude longitude of place of recording (float).

title title of the song (string).

firstline first line of the lyrics (string).

textfamily_id identifier of the text family the song belongs to (integer).

tunefamily_id short identifier of the tune family (ASCII string).

tunefamily name of the tune family the song belongs to (string).

type type of the song (`{folksong, instrumental}`).

voice_strophe_number serial number of the strophe or voice that is in the file (integer).

voice_strophe whether the value of `voice_strophe_number` is a voice (in case of instrumental music) or a strophe (in case of folk songs) (`{voice, strophe}`).

image_filename_s list of filenames of images in MTC-OGLSCAN that belong to the song (list of strings).

audio_filename filename of the audio recording in MTC_OGLAUDIO that belongs to the song (string).

variation whether the tune should be considered a more distant variant within the tune family (*variant*) or a full member (\emptyset).

confidence the confidence with which the melody was assigned to the tune family (*weak, neutral, strong, \emptyset*).

6.3 Metadata Table MTC-INST-sources

This table is also included in MTC-FS.

source_id identifier of the source (integer).

title title of the source (string).

author author(s) of the source (list of string).

place_publisher place(s) and publisher(s) of the source (list of strings). Place and publisher are separated by a colon (“:”).

dating (approximate) dating of the source (string).

sorting_year year that can be used for sorting (integer).

type type of source (*{“manuscript”, “print”, “audio”}*)

copy_used library siglum of the copy that was used for digitization (string).

scan_url url of scanned images of the source, in case the source is publicly available online (string).

7 MTC-ANN

The annotated corpus is a collection of 360 folk song melodies from 26 tune families. The selection and annotation procedures for this set are described in detail by Volk and Van Kranenburg (2012). MTC-ANN has been extensively used in various publications. Version 1.0 contains the set exactly as used in these studies. It does not include the annotations. These will be released in version 2.0. There are two metadata tables. One lists the tune family labels of the songs, and the other lists for each of the 26 tune families a reference melody. For each tune family a reference melody has been chosen by the collection specialists of the Meertens Institute. This reference melody can be considered the ‘best example’ or most representative melody for the tune family.

8 MTC-LC

MTC-LC contains a relatively large corpus of 4,830 melodies that has been employed in various studies as a ‘background’ corpus for the Annotated Corpus (MTC-ANN). The 360 songs from the Annotated Corpus are included in MTC-LC, but the 4,470 other songs are from different tune families than the songs in the Annotated Corpus. MTC-LC consist of both instrumental and vocal songs. It has one metadata table, which lists the tune family labels of the songs.

9 License and Attribution



Meertens Tune Collections by Meertens Institute is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Based on a work at www.liederenbank.nl/mtc. If you use the data for a research paper, please cite this report:

Van Kranenburg, Peter, Martine de Bruin, Louis P. Grijp, Frans Wiering (2014). The Meertens Tune Collections. *Meertens Online Reports* 2014-1, Meertens Institute, Amsterdam.

To employ this data for commercial purposes, explicit permission should be obtained from the Meertens Institute.

Acknowledgements

We thank Jörg Garbers for initial work on the editor and file formats, Ellen van der Grijn for sharing her extensive knowledge of the contents of the data sets, Anja Volk for valuable advice, Marieke Klein, Mariet Kaptein, Frederik Advocaat, Simon Plantinga and various generations of students for entering many songs.

References

- Bayard, S. (1950). Prolegomena to a study of the principal melodic families of british-american folk song. *Journal of American Folklore* 63(247), 1–44.
- Doornbosch, A. (1987). Inleiding. In A. Dekker, M. van Dijk, and H. Kuijer (Eds.), *Onder de groene linde: Verhalende liederen uit de mondelinge overlevering*, Volume 1. Amsterdam: Uitgeverij Uniepers.
- Downie, J. S. (2003). Music information retrieval. *Annual Review of Information Science and Technology* 37, 295–340.
- Grijp, L. P. (2008). Introduction. In L. P. Grijp and I. van Beersum (Eds.), *Under the Green Linden — 163 Dutch Ballads from the Oral Tradition*, pp. 18–27. Amsterdam: Meertens Institute + Music & Words.
- Pollmann, J. and P. Tiggers (1941). *Nederland's volkslied*. Haarlem: De Toorts.
- Suppan, W. and W. Stief (Eds.) (1976). *Melodietypen des Deutschen Volksgesanges*. Tutzing: Hans Schneider.
- Van Duyse, F. (1903–1908). *Het oude Nederlandsche lied*. Den Haag / Antwerpen: Martinus Nijhoff / De Nederlandsche Boekhandel.
- Volk, A. and P. Van Kranenburg (2012). Melodic similarity among folk songs: An annotation study on similarity-based categorization in music. *Musicae Scientiae* 16(3), 317–339.
- Wiering, F., R. C. Veltkamp, J. Garbers, A. Volk, and P. Van Kranenburg (2009). Modelling folksong melodies. *Interdisciplinary Science Reviews* 34(2–3), 154–171.
- Willems, J. F. (1848). *Oude Vlaemsche liederen*. Gent: F. en E. Gyselynck.