

# DEMONSTRATING THE MEERTENS TUNE COLLECTIONS: ANNOTATED CORPUS (MTC-ANN) VERSIONS 1.1 AND 2.0

Peter van Kranenburg, Berit Janssen

Meertens Institute

{peter.van.kranenburg,berit.janssen}@meertens.knaw.nl

Anja Volk

Utrecht University

a.volk@uu.nl

## ABSTRACT

We present two updates of our data set with songs and meta-data from Dutch oral culture: The Annotated Corpus (MTC-ANN) versions 1.1 and 2.0.

## 1. INTRODUCTION

With the Meertens Tune Collections the Meertens Institute provides a rich set of collections of melodic data for research purposes, such as musicological investigations or music information retrieval tasks.

At ISMIR 2014 we presented version 1.0 of the collections, which is described in detail in [1]. We now present two updates of the annotated corpus (MTC-ANN): versions 1.1 and 2.0.

## 2. MTC-ANN 1.1

The Annotated Corpus is a relative small set of 360 folk song strophes in 26 tune families [2]. These 360 songs were selected by a collection specialist of the Meertens Institute to form a relatively small subset of vocal songs that is representative for the collection as a whole with regard to the variations that occur between the melodies of a tune family.

Version 1.0, which has been released previously, contains the digitised melodies and lyrics as used in various publications. Most of the metadata were not yet included. Version 1.1, which we present now, adds a rich set of annotations, including:

- occurrences of c. 1,400 melodic motifs in c. 100 motif classes;
- similarity ratings of pairs of songs for contour, rhythm, motifs, and lyrics;
- similarity ratings of pairs of phrases for contour and rhythm;
- tune family membership;
- reference melodies;
- phrases repetitions.

## 3. MTC-ANN 2.0

In version 2.0, many improvements have been made both in the songs and in the metadata. The melodies from field-work recordings were carefully checked against the original recordings. In that process, many errors were corrected both in melodies and in lyrics.

One problem in version 1.0 is that songs from the same tune family could have been partitioned into phrases in different ways. For example, it occurs that one variant has twice the number of phrases as another variant, because a mid-cadence was not considered strong enough to count as phrase end in the latter. As a result phrases from these songs do not correspond with each other. In version 2.0 the partitioning of songs into phrases was revised such that phrases from different variants do correspond with each other.

Furthermore, many metadata have been revised and added:

- occurrences of c. 1,650 motif occurrences in c. 100 motif classes;
- phrase labels indicating phrase similarities within the tune family;
- pitch shifts to correct for key differences;
- time dilation factors to correct for meter and beat differences;
- recording place (longitude, latitude), and recording date;
- source information of songs from written sources;
- place and date of birth of the singers.

With these improvements and additions, we present a data set that has great potential to serve many research questions concerning melodic similarity, motivic analysis, oral variation, et cetera.

The data sets are available for download from: <http://www.liederenbank.nl/mtc>.

## 4. REFERENCES

- [1] Peter Van Kranenburg, Martine De Bruin, Louis P. Grijp, and Frans Wiering. The meertens tune collections. Meertens Online Reports 2014-1, Meertens Institute, Amsterdam, 2014.
- [2] Anja Volk and Peter Van Kranenburg. Melodic similarity among folk songs: An annotation study on similarity-based categorization in music. *Musicae Scientiae*, 16(3):317–339, 2012.



© Peter van Kranenburg, Berit Janssen, Anja Volk. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** Peter van Kranenburg, Berit Janssen, Anja Volk. “Demonstrating the Meertens Tune Collections: Annotated Corpus (MTC-ANN) versions 1.1 and 2.0”, Extended abstracts for the Late-Breaking Demo Session of the 16th International Society for Music Information Retrieval Conference, 2015.