

# Een melodieënzoekmachine voor de Nederlandse liederenbank

In de tweede helft van de twintigste eeuw reisden de veldwerkers Will Scheepers en Ate Doornbosch met hun bandrecorders stad en land af om Nederlandse volksliederen vast te leggen die verloren dreigden te gaan. Het radioprogramma *Onder de groene linde* van Ate Doornbosch speelde daarbij een belangrijke rol. Reacties van luisteraars leidden tot talloze nieuwe opnames. De in totaal ongeveer 7.000 opnames zijn intussen gedigitaliseerd en worden bewaard in het Meertens Instituut in Amsterdam. Hiermee herbergt het Meertens Instituut een waardevolle collectie cultureel erfgoed. Sinds juni 2007 zijn de opnames publiek toegankelijk via de website van de Nederlandse liederenbank ([www.liederenbank.nl](http://www.liederenbank.nl)). Om een bepaald liedje te vinden kan in de liederenbank in de meta-data (zoals titel, zanger, opnamedatum, etc.) worden gezocht.

**Peter van  
Kranenburg**  
**Meertens  
Instituut**  
**Frans Wiering**  
**Universiteit  
Utrecht**

Een lang gekoesterde wens van het Meertens Instituut was om ook in staat te zijn op melodische inhoud te zoeken. In samenwerking met de Universiteit Utrecht is van 2006 tot 2010 aan dit probleem gewerkt in het WITCHCRAFT-project<sup>1</sup>. Dit artikel geeft een kort overzicht van de werkwijze en de resultaten.

volksliedmelodieën. Omdat deze mondeling worden overgedragen wil er nogal eens iets veranderen en soms kan de opeenvolging van veranderingen er zelfs tot leiden dat de oorspronkelijke melodie onherkenbaar is. Een belangrijk concept dat in het volksliedonderzoek wordt gebruikt om melodische verwantschap te beschrijven is *tune*

*family* (melodie-familie). Hiermee wordt een groep melodieën aangeduid die als variant van elkaar kunnen worden beschouwd. Een voorbeeld is afgebeeld in figuur 1.

Record 72103 - Strophe 1



Sol - daat kwam uit den oor - log en hoe - ra

Record 72288 - Strophe 1



Sol - daat kwam uit den oor - log weer en hoe - ra

Record 72284 - Strophe 1



Sol - daat kwam uit den oor - log en hoe - ra

Record 72285 - Strophe 1



Sol - daat die kwam hui - ten den oor - log en hoe - ra

Figuur 1 Voorbeeld van vier gerelateerde melodieën. Enkel de incipits ofwel de eerste woorden of openingslijn, zijn afgebeeld.

Een belangrijke voorwaarde voor het slagen van een dergelijk project is een interdisciplinaire aanpak. In Music Information Retrieval worden methoden ontwikkeld om op inhoud gebaseerd naar muziek te zoeken in grote verzamelingen, zoals op internet, of in digitale archieven. In de Muziekwetenschap, met name in het Volksliedonderzoek, wordt onderzoek gedaan naar verwantschappen tussen volksliedmelodieën. Beide disciplines zijn relevant binnen het WITCHCRAFT project.

Één van de eerste stappen was het begrijpen van de verwantschap tussen de

De domeinexperts aan het Meertens Instituut kennen de collectie *Onder de groene linde* door en door. Zij hebben – veelal impliciete – kennis over het soort relaties tussen de melodieën, die zij gebruiken om de melodieën in te delen in de juiste melodie-familie.

Om deze kennis expliciet te maken is een annotatiemethode ontwikkeld waarmee de domeinexperts de gelijkenisrelaties tussen een groot aantal melodieën hebben geannoteerd. Hieruit bleek dat aspecten als ritme en contour een rol spelen bij het herkennen van een melodie, maar dat de aanwezigheid van karakteristieke motiefjes in een melodie het belangrijkste is.

In het WITCHCRAFT project zijn verschillende benaderingen onderzocht. De methode die uiteindelijk het meest succesvol is gebleken, gebruikt toonhoogte, ritme en structuur. De methode is gebaseerd op uitlijning

van reeksen. Als voorbeeld een uitlijning van twee reeksen woorden:

**De ----- kat zit op het grijze paaltje**  
**De kleine hond zit op het ----- paaltje**

scores: 1 -1 0,5 1 1 1 -1 1  
 totaal: 3,5

De twee zinnen zijn uitgelijnd door op geschikte plaatsen gaten in te voegen en te accepteren dat twee gerelateerde woorden, zoals kat en hond, met elkaar uitgelijnd worden (substitutie).

In de computerwetenschap bestaat een algoritme dat voor twee reeksen symbolen de optimale uitlijning kan uitreken, dat wil zeggen, waar gaten ingevoegd moeten worden om de uitlijning met de hoogst mogelijke score te verkrijgen. De score wordt berekend door de scores van de individuele associaties bij elkaar op te tellen. In het voorbeeld hierboven geven we een exacte match score 1, een niet-exacte match 0,5 en een gat -1. De totaalscore voor de uitlijning is dus 3,5.

In het geval van de melodieën zijn de noten de symbolen die we uitlijnen. De berekening van de scores stelt ons in staat muzikale kennis te verwerken in het algoritme. Dat is een zeer belangrijke eigenschap. Hier vinden we het aanknopingspunt voor een interdisciplinaire benadering. In feite beantwoordt een individuele score de vraag: hoe graag willen we dat deze twee symbolen (noten) met elkaar uitgelijnd worden? Het antwoord is muzikaal bepaald en hangt af van een aantal factoren.

De volgende voorwaarden bleken in de meeste gevallen voldoende: de **toonhoogte** dient ongeveer overeen te komen, het **metrisch gewicht** (de 'belangrijkheid' van de noot) dient overeen te komen, en de **noten** dienen zich op ongeveer dezelfde plaats bin-

nen de zin te bevinden (bijvoorbeeld beide aan het begin van de zin, beide aan het eind, of beide in het midden). Een voorbeeld van zo'n uitlijning, inclusief scores, is afgebeeld in figuur 2.

Wanneer we een onbekende melodie hebben (de zoekvraag) en we willen gerelateerde melodieën in de collectie vinden, kunnen we als volgt te werk gaan. We berekenen de uitlijning (en daarmee de score) van de onbekende melodie met elk van de melodieën in de collectie. Vervolgens sorteren we deze volgens de scores van de uitlijningen. Dat levert een resultaatlijst op waarin de meest gelijkende melodie bovenaan staat. Bij een test met 360 zoekvragen en een collectie van bijna 5.000 melodieën blijkt dat in 90% van de gevallen een relevant zoekresultaat bovenaan de resultaatlijst staat en dat zich in 99% van de gevallen een relevant zoekresultaat bij de eerste 10 melodieën bevindt.

Met deze methode is een prototype van een melodieënzoekmachine voor de Nederlandse liederenbank gemaakt. Momenteel gewerkt aan een robuuste implementatie die publiek beschikbaar kan worden gesteld.

Hoewel het ontwikkelde model voldoende krachtig is om een zoekmachine voor de liederenbank te maken, nodigt het resultaat van het annotatie-experiment, namelijk dat motieven het belangrijkste zijn voor de herkenning van een melodie, uit tot nader onderzoek naar manieren om motieven te gebruiken voor het zoeken naar melodieën. Dit leidt tevens tot nieuwe onderzoeksvragen waarin relaties gelegd kunnen worden met de wijze waarop mensen melodieën onthouden of herkennen. In ieder geval zijn de huidige resultaten een goed voorbeeld van de wijze waarop muzikale kennis en computationele methoden gecombineerd kunnen worden in een succesvolle toepassing.

En daar reed eris een heer, en hij was wellegemoed En daar reed eris een heer, en hij was wellegemoed En hij droeg er d' r rozekrans on er zijn hoed En hij droeg er een ro - zekrans on erre zijn hoed.

Daar reed er een heer die was wel dideldon die was wel don da don didel on don Daar reed er een heer die was wel dergemoed

Figuur 2. Voorbeeld van een uitlijning van twee melodieën inclusief substitutiescores.