

Tarsos

joren.six@hogent.be

11 februari 2011

Samenvatting

Dit artikel gaat over Tarsos: een softwareprogramma om het gebruik van toonhoogte in muziek te onderzoeken. Tarsos kan om het even welk muzikaal signaal annoteren met toonhoogte-informatie. Die annotaties worden verwerkt naar musicologisch interessante voorstellingen. Een van die voorstellingen is het toonhoogtehistogram, het suggereert de gebruikte toonladder. Een fijnmazig histogram kan meer dan enkel de twaalf westerse tonen bevatten. Die eigenschap maakt Tarsos goed bruikbaar voor de analyse van etnische muziek die vaak een volledig andere organisatie van toonmateriaal hanteert. Doordat de methode geautomatiseerd is, biedt Tarsos mogelijkheden tot het verwerken van toonhoogtegebruik in grote muziekbibliotheken. Een andere toepassing van Tarsos is real-time analyse van bijvoorbeeld gezongen intervallen.

Sleutelwoorden: Computational Ethnomusicology, Frequentie histogram, Toonladderherkenning

1 Inleiding

Onze ambitie is het ontwikkelen van een methode om een cultuuronafhankelijke kijk op muzikale parameters te verkrijgen. Meer concreet worden er technieken aangewend uit Music Information Retrieval (MIR) om toonhoogte, tempo en timbre te bestuderen. MIR is een verzamelnaam voor allerhande signaalverwerkingstechnieken die informatie uit muziek halen. Aanpassing van bestaande, meestal westers georiënteerde, MIR-methodes moet leiden tot een gestructureerde documentatie van verschillende klankkleuren, toonschalen, metrische verhoudingen en muzikale vormen. Die beschrijving kan dienen als inspiratie voor de ontwikkeling van een artistieke compositionele taal of kan gebruikt worden als bronmateriaal voor wetenschappelijk onderzoek rond etnische muziek. In de eerste fase van het onderzoek ligt de focus op één van de meer tastbare parameters: toonhoogte.

Op het eerste zicht is muzikaal toonhoogtegebruik goed gedefinieerd, gebaseerd op een lange historisch gegroeide traditie. Voor westerse muziek is dit inderdaad het geval. Een muziektheorie en notatie voor uitvoering en compositie ontstonden uit een praktische noodzaak. Een gelijkzwevende stemming,

een ander praktisch compromis, liet modulaties toe binnen een muziekstuk. Die verdeling van het octaaf in twaalf gelijke delen is uitgegroeid tot de basis voor bijna alle westerse muziek. Slechts enkelingen zoals Harry Partch en La Monte Young rukken zich los uit dit strakke keurslijf en spartelen rond in die oceaan van mogelijkheden met de naam microtonaliteit.

Niet-westerse klassieke muziek gebruikt vaak een theoretisch systeem met een andere kijk op toonhoogte. Een van de meer in het oog springende verschillen is dat niet alle intervallen even groot zijn. Dit kan zorgen voor een duidelijk aanwezige spanning in de muziek. Turkse muziek gebruikt bijvoorbeeld vaak kwarttonen. Een ander voorbeeld is de ongelijke octaafverdeling van de Gamelan Pélog toonladder.

Muziek van orale culturen baseert zich enkel op de uitvoeringspraktijk. Het gebrek aan een uitgeschreven muziektheorie met vaste regels in verband met toonhoogtegebruik zorgt voor een grote vrijheid en bepaalt de karakteristieke eigenschappen van de muziek zelf: minder harmonische structuren, instrumenten met variërende stemming, geen harmonische modulatie,...

Een niet te onderschatten probleem bij zo'n orale cultuur is het verlies aan authenticiteit. Door onder meer kolonisatie, missionarismissies, (muziekinstrumenten-) handel en massamedia is de impact van westerse muziek op etnische muziek niet te ontkennen. Zo is het bijvoorbeeld al bewezen dat toonladders van Afrikaanse muziek steeds meer en meer westerse intervallen gebruiken[10]. De muzikale traditie van orale culturen kan gezien worden als uniek maar bedreigd cultureel erfgoed. Om de specifieke kenmerken van die muziek beter te begrijpen is er nood aan een methode die net die kenmerken kan uitpueren, waarbij we ons in de eerste plaats op toonhoogte en toonladders concentreren.

Om onder meer toonladders uit muziek te extraheren en weer te geven werd het software platform *Tarsos* ontwikkeld. Met *Tarsos* is het mogelijk om automatische toonladderanalyses uit te voeren op een grote dataset of om manueel een gedetailleerde analyse te verkrijgen van enkele muziekstukken. De cultuuronafhankelijke analysemethode waarvan *Tarsos* gebruik maakt, kan even goed toegepast worden op Indonesische, Westerse of Afrikaanse muziek.

Na deze inleiding volgt een stukje met wat meer achtergrondinformatie over de gebruikte methode en vergelijkbaar onderzoek. Daarna volgt een omschrijving van de werking van *Tarsos* aan de hand van een voorbeeld. We besluiten, nogal voorspelbaar eigenlijk, met een kort besluit.

2 Methodologie & vergelijkbaar onderzoek

Dit stukje tekst geeft een erg kort overzicht gegeven van vergelijkbaar onderzoek en hoe we dankbaar de resultaten van die vergelijkbare studies gebruiken. Het schept een summier beeld van de manier waarop we te werk gaan.

De methode die we hanteren staat in de literatuur gekend onder verschillende namen: Constant-Q Profile[12], Harmonic Pitch Class Profile[3], Chromagram, Pitch Frequency Histogram of eenvoudigweg Frequency Histogram[13].

Het basisidee dat schuil gaat achter deze benamingen is eenvoudig: er wordt bijgehouden hoeveel keer bepaalde frequenties in het fysieke muzikale signaal voorkomen en die gegevens worden dan op een bruikbare manier voorgesteld. De implementatie van dit eenvoudige idee is echter niet triviaal.

Verschillende grondtonen uit een signaal halen is niet eenvoudig. Vroeger gebeurde dit op een erg arbeidsintensieve manier: alle aanwezige toonhoogtes werden vergeleken met een groot aantal bellen met een gekende stemming, bijvoorbeeld voor elke Hertz-waarde één. Hoewel er interessante resultaten te behalen vielen is het duidelijk dat deze werkwijze niet bruikbaar is op grotere schaal. Automatisatie was dan ook geen slecht idee.

Al eind jaren zestig[13] kwamen er onderzoekers op de proppen met een geautomatiseerd experiment om tooninformatie¹ uit monofone opnames te halen. Ze onderzochten de toonladder van een Zweeds blaasinstrument met de tot de verbeelding sprekende naam: de *spilåpipa* of, vrij vertaald, de speelpijp. De erg rudimentaire frequentieschatting van die eerste experimenten werd dankzij onderzoek naar spraakherkenning sterk verbeterd. In die context ontstonden er robuustere, maar nog steeds monofone, frequentiedetectoren zoals SHS of nog recenter YIN[2]. Die werden al toegepast op onderzoek naar muziek[4]. Er werden zelfs algoritmes van afgeleid specifiek voor de verwerking van muziek[9].

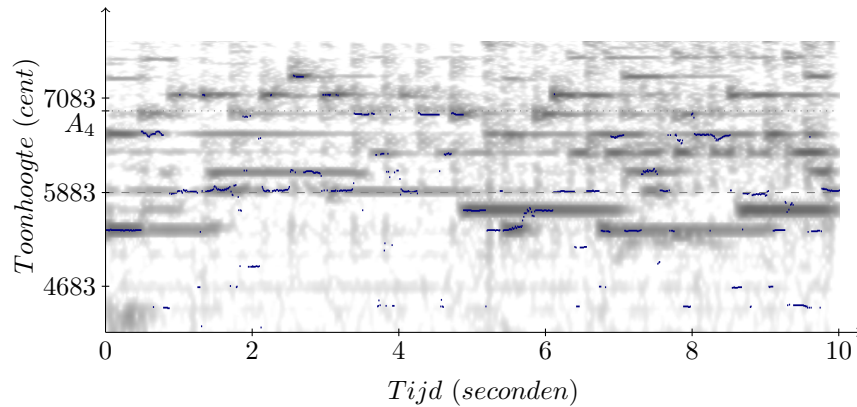
Grondtonen detecteren in polyfone muziek is nog steeds een onopgelost probleem. Hoewel er veel pogingen werden ondernomen [11, 5, 6] is er nog steeds geen universele oplossing voorhanden die even accuraat en flexibel is als een getraind menselijk oor[7]. Dit is hoofdzakelijk te wijten aan het feit dat harmonischen van zangers, instrumenten en percussie gereduceerd worden naar één signaal dat nadien niet meer te ontleden is. De grondtoon van het ene geluid overlapt vaak met een harmonische van een ander, wat grondtoondetectie erg bemoeilijkt[1].

Met de beperkingen van de huidige frequentiedetectoren in het achterhoofd kan er toch op een bruikbare, nuttige manier omgesprongen worden met de aangeleverde annotaties. De manier waarop wij dit doen is te lezen in onderstaande tekst.

3 Tarsos

In dit deel geven we een overzicht van de manier waarop Tarsos werkt. We beginnen op het allerlaagste niveau - het audiosignaal - en begeven ons stap voor stap naar een hoger, musicologisch interessanter, symbolisch niveau. Ter illustratie wordt er doorheen dit hoofdstuk gebruik gemaakt van een analyse op een Indonesisch muziekstuk.

¹(Grond)toon en (basis)frequentie worden hier door elkaar gebruikt omdat ze vaak hetzelfde zijn. Technisch gezien is dit echter niet zo. Meestal valt de grondtoon, de pitch samen met de basisfrequentie maar soms ook niet, voorbeelden zijn te vinden in [8]. Toon is een perceptuele ervaring frequentie een fysiek aantal trillingen per seconde.



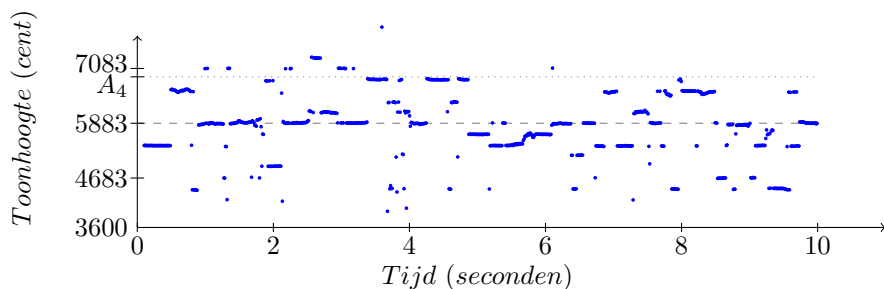
Figuur 1: Spectrogram van tien seconden van een Indonesisch muziekstuk: donkergrijze zones geven frequentiebanden aan met veel energie. De stippen geven toonhoogte-annotaties weer. Rond de stippellijn ter hoogte van 5883 cents zijn veel annotaties te vinden. Ter referentie werd een gestipte horizontale lijn toegevoegd met A4, 440Hz.

3.1 Signaalniveau

Hoe fijn de luistervaring ook kan zijn muziek is gezien als signaal, heel erg vaak een regelrechte ramp. Vooral de muziek - het signaal - dat onze interesse het meest wegdraagt: opnames van etnische muziek. Vaak zijn dit oude opnames gemaakt met heel wat achtergrondlawaai, opgenomen met krakkemikkige opnameapparatuur op fragiele dragers. Ook het digitalisatieproces is erg gevoelig voor fouten. Één en ander kan verholpen worden met bijvoorbeeld ruisonderdrukking, band-pass filtering of source separation. Na die optionele stap die het signaal bruikbaar maakt, belandt de audio in het systeem.

De audio wordt eerst en vooral omgezet naar een eenvoudig formaat (WAV) en meteen daarna volgt er een frequentieanalyse. Om het hoofd te bieden aan de eerder vernoemde problemen bij toonhoogtedetectie werd voor een modulaire aanpak gekozen: verschillende toonhoogtedetectoren kunnen gebruikt worden. De geavanceerde detectoren zijn soms moeilijk correct te installeren; daarom worden er ook twee eenvoudige meegeleverd: YIN[2] en MPM[9] werken altijd, zonder configuratie.

In figuur 1 is het spectrogram te zien van een tiental seconden van een Indonesisch muziekstuk. Het geeft weer waar de toonhoogte informatie te vinden is: op de donkerdere plaatsen is er meer energie aanwezig, op de witte plaatsen is geen informatie te vinden. De puntjes stellen toonhoogte-annotaties voor, hier werd een monofone detector gebruikt. Op de figuur is te zien hoe sommige annotaties niet in de donkere (waar frequenties duidelijk aanwezig zijn) maar in de lichtere, tot zelfs witte stukken voorkomen. Dit zijn ofwel octaaffouten, de detector schat de grondtoon dan een octaaf te laag, ofwel zijn de annotaties gewoon fout, bijvoorbeeld tijdens toon-overgangen. Dit toont aan dat de anno-



Figuur 2: Een piano-roll-achtige voorstelling. Annotaties van de eerste tien seconden van een Indonesisch Slendro muziekstuk zijn te zien. Het is duidelijk dat toon-informatie horizontaal is geclusterd, bijvoorbeeld rond 5883 cents, aangegeven door de gestreepte lijn. Ter referentie werd een gestipte horizontale lijn toegevoegd met A_4 , 440Hz.

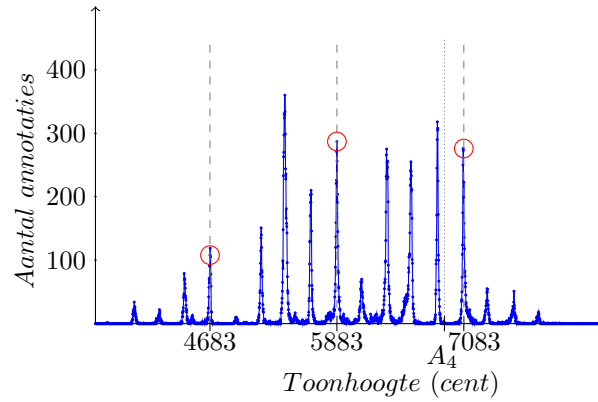
taties niet perfect zijn, misschien zijn betere annotaties mogelijk met een andere detector. Gelukkig kan de detector gekozen worden. Welke detector ook gekozen werd, het gebruik is volledig gelijklopend: de annotaties worden omgezet in een uniform formaat en bewaard voor verwerking op het symbolische niveau.

3.2 Symboolniveau

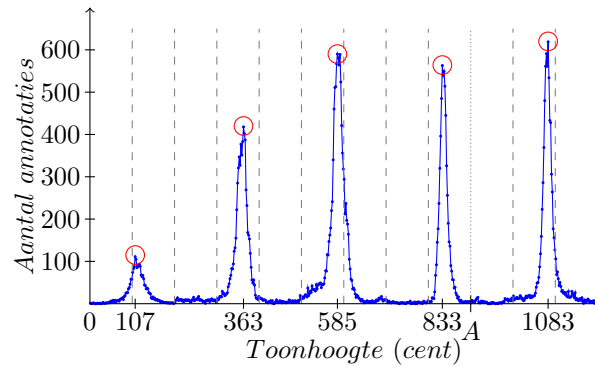
Eens de signaalverwerking achter de rug is, kunnen we aan de slag met de toonhoogte-annotaties. De meest recht-voor-de-raapse representatie van die annotaties lijkt er sterk op het spectrogram van figuur 1. In figuur 2 worden de annotaties in functie van de tijd geplaatst. In die voorstelling wordt de cent eenheid gebruikt. De cent verdeelt elk octaaf in 1200 gelijke stukjes. Het verschil tussen een A en $A\#$ is dan 100 cent. Om de cent eenheid te kunnen gebruiken moet er een referentiefrequentie vermeld worden, hier gebruiken we 8.176Hz - een erg lage C . Dit wil zeggen dat 8.176Hz gelijk is aan 0 cent, 16.352Hz aan 1200 cent enzovoort.

Wat ook al min of meer duidelijk was uit het spectrogram blijkt hier nog duidelijker. Toonhoogte-informatie is niet uniform aanwezig maar blijkt zich vooral horizontaal te clusteren. Dit is nog duidelijker zichtbaar in het toonhoogte histogram.

Het toonhoogtehistogram (pitch histogram) geeft weer hoeveel annotaties met een bepaalde toonhoogte in het gehele stuk voorkomen. Hoewel de tijdsinformatie verloren gaat, levert dit toch iets op: het wordt duidelijk welke tonen veel voorkomen. Figuur 3 gebruikt dezelfde gegevens als de voorgaande voorbeelden. De piek op 5883 cents is duidelijk aanwezig. De ambitus van een stuk kan met deze voorstelling ook bepaald worden: het is het verschil tussen de hoogste en laagste toon. In de figuur is een interessant patroon te ontdekken: het verschil tussen 4683, 5883 en 7083 is telkens 1200 cents, pieken komen elk octaaf terug. De volgende voorstelling maakt daar dankbaar gebruik van.



Figuur 3: Een toonhoogtehistogram van een Indonesisch Slendro muziekstuk. De cirkels geven de meest voorkomende tonen in verschillende octaven weer. Ter referentie werd een gestipte lijn met A_4 , 440Hz toegevoegd.



Figuur 4: Een tonenhistogram met een Indonesische Slendro toonladder. De cirkels geven de tonen weer. Ter referentie werden de gestreepte lijnen toegevoegd die de westerse gelijkzwevende stemming tonen. De A is aangegeven met een gestipte lijn.

Tonen (cent)	Toonintervallen (cent)
107	
	255
363	478
	222 725
585	470 976
	248 720
833	498
	251
1083	

Tabel 1: Toon en toonintervallen, beide in cents. Dezelfde pentatonische Indonesische slendro van figuur 4 werd gebruikt.

Het tonenhistogram (pitch class histogram) bevat dezelfde informatie als het toonhoogtehistogram maar alle gegevens worden gereduceerd naar één octaaf. De grootte van de piek geeft in deze voorstelling weer hoe vaak een toon (in tegenstelling tot een toonhoogte) voorkomt. In figuur 4 is te zien hoe de piek op 5883 cent in het toonhoogtehistogram bijdraagt tot de piek op 1083 cent in het tonenhistogram. Deze laatste voorstelling begint een toonladder waarneembaar te worden: de pieken, de cirkels in figuur 4, geven de tonen uit de toonladder.

Nu de tonen gekend zijn, kan een tabel opgesteld worden met daarin gegevens over de tonen en de intervallen. Uit tabel 1 blijkt dat een reine kwart - een frequentieverhouding $4/3$ of 498 cents - aanwezig is tussen 585 en 1083 cents. Omdat de voorstelling naar een octaaf gereduceerd is, is ook een frequentieverhouding van $\frac{2/1}{4/3} = 6/4 = 3/2$ of $1200 - 498 = 702$ cent aanwezig: de reine kwint.

3.3 Uitvoer- en gebruiksmogelijkheden

Tarsos kan de data na elke zopas omschreven stap wegschrijven: vanaf de toonhoogte-annotaties tot de toonintervallenmatrix. Voor toonladders is er een speciaal gestandaardiseerd formaat beschikbaar: het formaat dat gebruikt wordt door het programma Scala. Met Scala kunnen toonladders op allerlei manieren vergeleken worden. Het bevat ook een database met meer dan 3700 toonladders, dit gaat van historische klavecimbel stemmingen tot toonladders die gebruikt worden in hedendaags klassieke muziek.

MIDI biedt een volledig andere uitvoer- en gebruiksmogelijkheid. Met Tarsos kunnen er MIDI berichten verstuurd worden om synthesizers te stemmen in een willekeurige toonladder. Zo kan je bijvoorbeeld meespelen met een etnisch muziekstuk in de oorspronkelijke stemming.

Real-time toepassingen zijn met Tarsos ook mogelijk. Geluid afkomstig van een microfoon wordt dan meteen geanalyseerd en onmiddellijke feedback toont een gespeeld of gezongen interval. Het maakt kwarttonen of andere (ongewone)

intervallen visueel duidelijk. Tarsos kan zo gebruikt worden door zangers of strijkers die willen experimenteren met microtonaliteit. Ook kan het handig zijn voor etnomusicologisch veldwerk: bijvoorbeeld om kora (een Afrikaanse harp) toonladders te documenteren.

4 Besluit

Tarsos biedt een degelijke basis voor onderzoek naar toonhoogtegebruik in muziek van allerlei culturen. Het schept mogelijkheden voor automatische analyse van grote muziekbibliotheken en het ontgonnen materiaal kan gebruikt worden om evoluties, patronen te zoeken in toonhoogtegebruik. Naast deze wetenschappelijke toepassingen zijn er ook artistieke mogelijkheden. Een componist kan inspiratie opdoen en experimenteren met toonladders, een muzikant kan zijn of een zangeres kan haar intonatie verbeteren aan de hand van real-time feedback over toonhoogtegebruik.

Het gebruikte muziekvoorbeeld, de software en meer achtergrondinformatie is terug te vinden op de website van Tarsos: <http://tarsos.0110.be/tag/ARIP>.

Referenties

- [1] Michael Casey, Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney. Content-based mir: current directions and future challenges. *Proceedings of the IEEE*, 96(4):668–695, 2008.
- [2] Alain de Cheveigné and Kawahara Hideki. Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930, 2002.
- [3] Fujishima. Realtime chord recognition of musical sound: A system using common lisp music. In *Proc. Int. Comput. Music Conf*, pages 464–467, 1999.
- [4] Ali C. Gedik and Barış Bozkurt. Pitch-frequency histogram-based music information retrieval for turkish music. *Signal Processing*, 90(4):1049–1063, 2010.
- [5] Masataka Goto. A real-time music-scene-description system: predominant-F0 estimation for detecting melody and bass lines in real-world audio signals. *Speech Communication*, 43(4):311 – 329, 2004. Special Issue on the Recognition and Organization of Real-World Sound.
- [6] Anssi Klapuri. Auditory model-based methods for multiple fundamental frequency estimation. In *Signal Processing Methods for Music Transcription* [7], pages 229–265.
- [7] Anssi Klapuri and Manuel Davy. *Signal Processing Methods for Music Transcription*. Springer, 2006.
- [8] Philip McLeod. *Fast, accurate pitch detection tools for music analysis*. PhD thesis, University of Otago. Department of Computer Science, 2009.
- [9] Phillip McLeod and Geoff Wyvill. A smarter way to find pitch. In *Proceedings of International Computer Music Conference, ICMC*, 2005.
- [10] Dirk Moelants, Olmo Cornelis, and Marc Leman. Exploring african tone scales. In *Proceedings of 9th ISMIR Conference*, 2009.
- [11] Brian C. J. Moore and Brian R. Glasberg. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America*, 74(3):750–753, 1983.
- [12] Hendrik Purwins, Benjamin Blankertz, and Klaus Obermayer. Constant Q profiles for tracking modulations in audio data. In *International Computer Music Conference*, pages 407–410, 2001.
- [13] J Sundberg and P Tjernlund. Computer measurements of the tone scale in performed music by means of frequency histograms. *STL-QPS*, 10(2-3):33–35, 1969.