Verfahren der Signalverarbeitung für Query-by-Humming-Systeme

Jan-Mark Batke

Fachgebiet Nachrichtenübertragung, Technische Universität Berlin, Deutschland, Email: batke@nue.tu-berlin.de

Query by Humming

Suchmaschinen nehmen im Internet immer größere Bedeutung ein [9]. In diesem Beitrag werden Verfahren der Signalverarbeitung speziell für Musiksuchmaschinen vorgestellt. Ein Query-by-Humming (QBH) System ermöglicht es, über das Vorsummen einer Melodie eine Datenbanksuche zu starten. Der Nutzer des Systems summt eine Melodie in ein Mikrofon, das QBH-System verwendet diese Anfrage, um in einer angeschlossenen Datenbank nach einem Musikstück mit dieser Melodie zu suchen. Das Rechercheergebnis wird schließlich dem Nutzer präsentiert. Bei QBH-Systemen wird üblicherweise eine *symbolische* Melodiedarstellung verwendet, welche den gespeicherten Musiktitel in der Melodiedatenbank repräsentiert. Die Suchanfrage wird vom QBH-System in eine ebensolche symbolische Melodiedarstellung umgewandelt und kann dann mit den Melodien der Datenbank verglichen werden.

Abbildung 1 zeigt die Architektur des QBH-Systems »Queryhammer« [1]. Die Nutzeranfrage wird von einem Mikrofon in ein PCM-Signal (PCM: Pulse-Code-Modulation) umgewandelt, dieses wird in der monophonen Transkriptionsstufen in eine symbolische Melodiedarstellung, in diesem Fall ein MPEG-7-konformes Format, überführt. Die Melodiedatenbank enthält ebenfalls MPEG-7-Melodiedarstellungen; diese werden aus PCM- oder MIDI-Dateien (MIDI: Musical Instrument Digital Interface) extrahiert oder direkt eingegeben. Nach einem Vergleich der Anfrage mit den Melodien der Melodiedatenbank erfolgt die Ausgabe des QBH-Systems in Form einer Liste mit den besten Treffern.

Musikdatenbanken

Die in Abbildung 1 gezeigten Datenbanken unterscheiden sich durch die verwendete *Musikrepräsentation*. Die Repräsentation von Musik kann technisch gesehen in zwei Kategorien unterschieden werden: in der **Audio-Repräsentation** werden Aufnahmen von Konzerten oder Studioproduktion festgehalten, die zum Beispiel als WAV- oder MP3-Datei auf einem Rechner digital gespeichert werden können. Die **symbolische Repräsentation** von Musik hingegen meint die Notenschrift oder technische Formate für Musiknotation wie *Guido*, die Noteninformation enthalten [7]. Das Format *MIDI* kann man die Informationen zu Noten, wie sie auf einem elektronischen Musikinstrument gespielt werden, speichern. Speziell für QBH-Systeme können Melodiekonturen verwendet werden, die zum Beispiel in Form des *Parsons*-Codes [13] oder als MPEG-7 *MelodyContour* [1] dargestellt werden. Der Vorgang der Transkription meint die Umwandlung von einer Audiorepräsentation in eine symbolische Repräsentation.

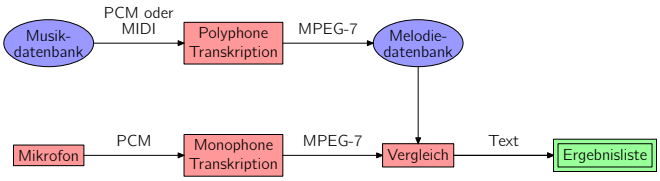

Abbildung 1: Das Blockschaltbild eines QBH-Systems.

Monophone Transkription

Die Transkription von Gesangseingaben in eine symbolische Darstellung ist ein notwendiger Schritt in jedem QBH-System. Viele Publikationen beschäftigen sich speziell mit dieser Problematik [4]. In bestehenden Systemen ist dieser Teil oft als Java-Programm ausgeführt, siehe zum Beispiel [11] oder [12]. Daraus ergibt sich die Anforderung an die verwendeten Algorithmen, möglichst kostensparend implementiert werden zu können. Da durch die Bereitstellung als Java-Programm auch fast jeder beliebige Rechnerarbeitsplatz in Frage kommt, ist eine gute Robustheit gegen akustische Störungen erforderlich.

Abbildung 2 zeigt die Verarbeitungsschritte aus dem System »Queryhammer« [1]. Die Anfrage wird nach einer Bandpassfilterung einer Grundfrequenzanalyse (GFA) unterzogen. Aus dem Grundfrequenzverlauf werden in der Rhythmuserkennung die Anfänge der einzelnen Noten ermittelt. Das ist möglich, sofern auf »na« oder »da« gesungen wird, andernfalls sind aufwändigere Verfahren notwendig, wie sie zum Beispiel in [14] vorgestellt werden. Die Auswertung der erkannten Noten führt zu der gewünschten Melodiedarstellung.

Von besonderer Bedeutung ist das GFA-Verfahren. Die Anzahl existierender GFA-Verfahren ist unüberschaubar groß [6], an dieser Stelle sollen einige Beispiele genannt werden. In [1] wird die Autokorrelationsmethode verwendet, die aus dem Sprachanalyseprogramm »Praat« stammt [2]. Im Verfahren »YIN« [3] wird die Betragsdifferenzfunktion (average magintude difference function, AMDF) verwendet. Das Verfahren in [4] arbeitet aufwändiger und verwendet ein Gehörmodell. Die Erfolgsquote für solcher Verfahren ist gut, es werden bei gesungenen Silben Erkennungsraten bis 85 % erreicht [10].

Polyphone Transkription

Der Block *Polyphone Transkription* in Abbildung 1 ist kein notwendiger Teil eines QBH-Systems, jedoch erforderlich zur Erzeugung einer Melodiedatenbank. Enthält die Musikdatenbank MIDI-Dateien, so lässt sich die gewünschte Melodiedarstellung technisch einfach erstellen. Wesentlicher schwieriger und derzeit technisch noch nicht beherrscht ist es, die Melodien aus Audio-Informationen zu erstellen. Ein wichtiger Schritt bei der

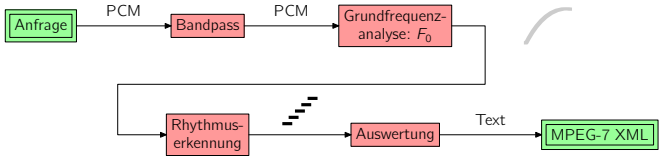

Abbildung 2: Transkription monophoner Signale im Beispielsystem ≫Queryhammer≪ [1].

polyphonen Transkription ist die Aufgabe der Mehrfachgrundfrequenzanalyse (MGFA). Eine gute Übersicht zu diesem Thema findet sich in [8].

Besonders die Arbeiten von Goto [5] sind interessant für QBH-Anwendungen, da dort herkömmliche CD-Aufnahmen untersucht und mit gutem Erfolg transkribiert werden. Für die Suche im Signalgemisch wird ein Signalmodell bestehend aus Grundton und mitschwingenden Harmonischen angenommen. Der untersuchte Frequenzbereich wird durch eine Filterbank in Melodie- und Baßbereich unterteilt. Einschränkung dieses Verfahrens ist jedoch, dass die Melodie bezüglich ihrer Lautstärke sehr klar im Vordergrund stehen muss, um als solche erkannt zu werden.

Klapuri [8] beschreibt ein automatisches Transkriptionssystem für polyphone Audiosignale, das auf einem Mehrfachgrundfrequenz-Erkenner basiert. Es wird nach Klängen gesucht, die ein harmonisches Obertonspektrum besitzen. Frequenzkomponenten, die einem harmonischen Klang zugeordnet werden können, werden sukzessive aus dem Signal entfernt. Weiterhin wird in [8] ein ebenfalls iteratives Verfahren vorgestellt, das auf einem Modell des menschlichen Gehörs basiert. Dieses Verfahren ist bezüglich der Analyse von Musiksignalen besonders interessant, da sich Gehörmodelle bei der auditiven Szenenanalyse gut bewähren.

Die Erfolgsquote für polyphone Transkription hängt sehr stark vom Audiomaterial und von den Bewertungskriterien ab; beim ≫Melody Extraction Contest≪ der ISMIR 2004 (siehe `http://ismir2004.ismir.net/ISMIR_Contest.html`) wurden im Schnitt 50 % erreicht.

Zusammenfassung

QBH-Systeme ermöglichen die Suche nach Melodien durch Summen, zur Darstellung und zum Vergleich von Melodien werden intern Melodiekonturen verwendet. Die Transkription von gesummten, monophone Anfragen in eine Melodiekontur ist technisch beherrschbar, die Transkription von Melodien aus polyphonen Musiksignalen hingegen schwierig. Während für monophone Signale die möglichst effiziente Verarbeitung Gegenstand weiterer Entwicklungen ist, sind für polyphone Signale noch grundlegende Forschungsarbeiten notwendig. Besonders hoffnungsvoll sind hier die Ansätze unter Verwendung von Gehörmodellen.

Literatur

[1] Batke, Jan-Mark ; Eisenberg, Gunnar ; Weishaupt, Philipp ; Sikora, Thomas: A Query by Humming system using MPEG-7 Descriptors. In: *Proc. of the 116th AES Convention*. Berlin : AES, Mai 2004

[2] Boersma, Paul: Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-To-Noise Ratio of a Sampled Sound. In: *IFA Proceedings 17*, 1993

[3] Cheveigné, Alain de: YIN, a fundamental frequency estimator for speech and music. In: *J. Acoust. Soc. Am.* 111 (2002), April, Nr. 4

[4] Clarisse, L. P. ; Martens, J. P. ; Lesaffre, M. ; Baets, B. D. ; Meyer, H. D. ; Leman, M.: An Auditory Model Based Transcriber of Singing Sequences. In: *Proceedings of the ISMIR*, 2002, S. 116–123

[5] Goto, Masataka: A Robust Predominant-F0 Estimation Method for Real-Time Detection of Melody and Bass Lines in CD Recordings. In: *Proc. ICASSP*, 2000, S. 757–760

[6] Hess, Wolfgang ; Schroeder, Manfred R. (Hrsg.): *Springer Series in Information Sciences*. Bd. 3: *Pitch Determination of Speech Signals*. Berlin : Springer-Verlag, 1983

[7] Hoos, Holger H. ; Renz, Kai ; Görg, Marko: GUIDO/MIR — an Experimental Musical Information Retrieval System based on GUIDO Music Notation. In: *Proceedings of the Second Annual International Symposium on Music Information Retrieval*, 2001

[8] Klapuri, Anssi: *Signal Processing Methods for the Automatic Transcription of Music*, Tampere University of Technology, Diss., 2004

[9] Lehmann, Kai (Hrsg.) ; Schetsche, Michael (Hrsg.): *Die Google-Gesellschaft*. transcript-Verlag, 2005

[10] Mulder, Tom D. ; Martens, Jean-Peiree ; Lesaffre, Micheline ; Leman, Marc ; Baets, Bernard D. ; Meyer, Hans D.: An Auditory Model Based Transcriber of Vocal Queries. In: *Proc. of the ISMIR*, 2003

[11] *Musicline: Die ganze Musik im Internet*. `http://www.musicline.de`. – QBH-System der phononet GmbH

[12] Musipedia: *Musipedia, the Open Music Encyclopedia*. `www.musipedia.org`, 2004

[13] Prechelt, Lutz ; Typke, Rainer: An interface for melody input. In: *ACM Transactions on Computer-Human Interaction* 8 (2001), Nr. 2, S. 133–149

[14] Viitaniemi ; Klapuri ; Eronen: A probabilistic model for the transcription of single-voice melodies. In: *Finnish Signal Processing Symposium, FINSIG*. Tampere University of Technology, Mai 2003, S. 59–63