

Einleitung

1

Durch die Flut von digitalen Musikstücken im Internet und auf Datenspeichern gewinnt die Suche nach einzelnen Musikstücken gegenwärtig stark an Bedeutung. So ermöglicht etwa ein mobiles Wiedergabegerät wie der *Apple iPod* ohne weiteres die Speicherung von 5.000 Titeln [4]. Bislang können einzelne Musikstücke aus einer solchen Datenmenge nur durch die Eingabe von Titel oder Interpret ausgewählt werden. Ein typisches Problem ist es aber, dass dem Musiksuchenden diese Informationen nicht gegenwärtig sind, sondern er sich lediglich an die Melodie erinnern kann.

Technisch gesehen handelt es sich beim Auffinden einzelner Musikstücke aus einer großen Menge von Titeln um die Suche in einer Musikdatenbank. Wenn dem Suchenden nur die Melodie bekannt ist, muss folglich die Eingabe der Melodie in das Suchsystem möglich sein, um eine wie oben beschriebene Suchanfrage bedienen zu können. Diese Anforderung wird von einem Query-by-Humming-System (QBH-System) erfüllt: Die Melodie kann dem System vorgesummt werden und eine Anzahl ähnlicher Melodien der Musikdatenbank werden als Suchergebnis präsentiert. In Abbildung 1.1 ist das Benutzungsszenario eines solchen QBH-Systems dargestellt.

1.1 Aufgabenstellung

Für den Erfolg der Suche nach Melodien ist es von entscheidender Bedeutung, wie die Suchanfrage und der Datenbankbestand miteinander verglichen werden. Bei den meisten QBH-Systemen wird eine *symbolische* Melodiedarstellung verwendet, in welcher die Melodien der Musikstücke in der Datenbank gespeichert werden. Eine solche symbolische Darstellung kann zum Beispiel der PARSONS-Code sein, der den Verlauf der Melodie mit nur drei Buchstaben beschreibt: „U“ für aufwärts (up), „D“ für abwärts (down) und „R“ für gleichbleibend (repeat) [154]. Eine Darstellung wie der PARSONS-Code wird auch als *Melodiekontur* bezeichnet. Wird die Suchanfrage an das QBH-System in eine ebensolche Melodiekontur umgewandelt, kann sie mit dem Datenbankbestand verglichen werden. Ergebnis dieses Vergleichs ist schließlich eine Liste

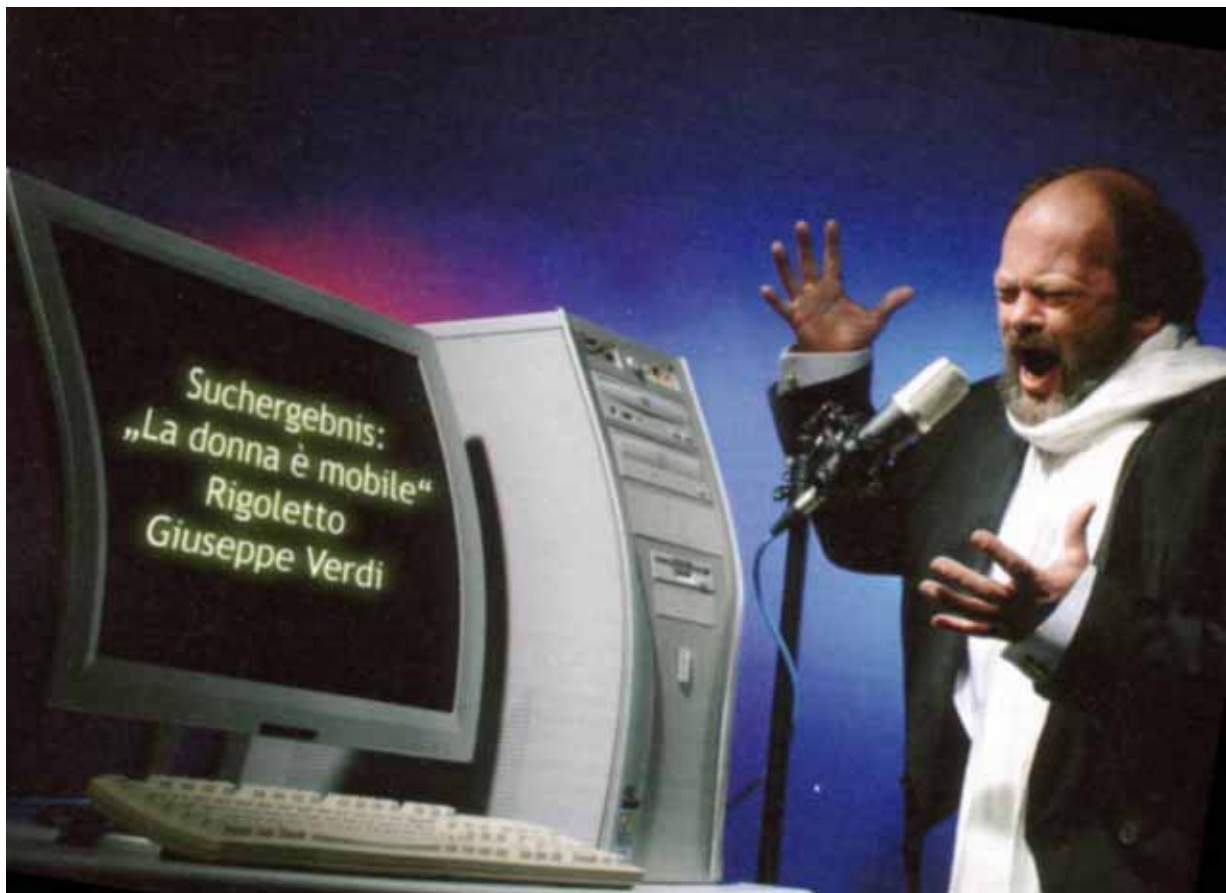


Abbildung 1.1: Ein Query-by-Humming-System in Anwendung: Der musikalische Vortrag des Nutzers wird vom QBH-System verarbeitet und die ähnlichste Melodie als Ergebnis einer Datenbanksuche angegeben. Quelle: [117]

mit den ähnlichsten Musikstücken. Die Musikstücke selbst können über Titel und Interpret, Noten oder als Audiodatei in der Datenbank gespeichert sein und dem Nutzer angeboten werden.

Für den Betrieb eines QBH-Systems sind damit im Wesentlichen zwei Aufgaben zu erfüllen: zuerst muss die gesummte Suchanfrage in eine symbolische Darstellung wie oben beschrieben gebracht werden. Dieser Vorgang ist die *Transkription* eines *monophonen* Gesangssignals in eine symbolische Darstellung. Die zweite Aufgabe besteht im *Vergleich* der erhaltenen Symbole mit dem Inhalt der Melodiedatenbank, um die ähnlichsten Melodien zu finden. Diese Aufgabe macht die Verwendung geeigneter *Ähnlichkeitsmaße* notwendig.

Bei beiden Aufgaben handelt es sich um komplexe Vorgänge. So ist die Transkription von gesummtten Signalen an Signalverarbeitungsschritte wie

eine Grundfrequenzanalyse und eine Rhythmuserkennung gebunden. Ähnlich wie zum Beispiel bei der Aufgabenstellung der automatischen Spracherkennung ist vieles, was für den Menschen leicht zu interpretieren ist, technisch nur durch aufwendige Verfahren zu analysieren. Der Vergleich von Melodien wird auf symbolischer Ebene durchgeführt. Die Darstellung der Melodie durch Symbole als Melodiekontur bedingt eine gewisse Ungenauigkeit, aber auch Verallgemeinerung der Repräsentation und führt daher auch zu einer Vergrößerung der Treffermenge. Diese größere Treffermenge wird zu Gunsten einer möglichst großen Freiheit beim Stellen der Suchanfrage bewusst in Kauf genommen. Beide Aufgaben, Melodietranskription und -vergleich, sind Gegenstand zahlreicher Untersuchungen in der Literatur [37, 41, 50, 69, 76, 88, 90, 101, 124, 132, 142, 153, 161, 165, 175, 198]. Bei allen Untersuchungen werden unterschiedliche Datenbestände und Bewertungskriterien herangezogen, damit ist ein Vergleich der Untersuchungen schwierig oder nicht sinnvoll. Ein Ziel dieser Arbeit ist es daher, möglichst flexibel einsetzbare Verfahren zur Bewertung von QBH-Systemen zu erarbeiten.

Um ein QBH-System anzubieten, ist der Aufbau einer Melodiedatenbank eine wichtige zu erfüllende Aufgabe. Audiodateien sind für Musikdatenbanken das am meisten verwendete Medium und liegen in großer Anzahl als Wellenform- oder MP3-Datei vor. So bietet zum Beispiel allein das Internetportal *mp3.de* über 150.000 frei erhältliche Titel an [5]. Beim kommerziellen Anbieter *Musicline* werden online über 1,8 Millionen Titel angeboten [11], über das Internetportal *iTunes* mehr als 2 Millionen Titel [4]. Das Portal *mp3.com* schließlich gibt 6 Millionen Titel an [2]. Um eine Melodiedatenbank aufzubauen, müssen symbolische Beschreibungen für die gespeicherten Musikstücke gefunden werden. Bei der Menge der verfügbaren Titel ist klar, dass dies auf automatischem Wege erfolgen sollte. Da es sich im Gegensatz zu gesummen Anfragen bei Musik meist um polyphone, d. h. mehrstimmige Signale handelt, muss zur Melodietranskription aus solchen Audiosignalen zusätzlich noch die richtige, also melodieführende Stimme ausgewählt werden.

Diese Aufgabe ist sehr schwierig und wird derzeit noch nicht beherrscht. Daher greifen bestehende QBH-Systeme meistens auf MIDI-Datenbanken zurück, bei denen sich die Melodie technisch einfach extrahieren lässt. MIDI steht für „music instrument digital interface“ – dieses Format enthält symbolische Informationen darüber, wie Noten auf einem elektronischen Musikinstrument gespielt werden. Daneben werden auch oft Melodien durch Musiker transkribiert und manuell in Melodiedatenbanken eingetragen. In jüngster Zeit hat es im Bereich der Forschung zur automatischen Transkription von polyphoner

Musik erhebliche Fortschritte gegeben [87,111]. Im Rahmen dieser Arbeit werden daher aktuelle Verfahren auf ihre Eignung für QBH-Systeme untersucht und bewertet.

Melodiedatenbanken und ihre Eigenschaften bestimmen in einem Suchsystem das Ergebnis und sind von zentraler Wichtigkeit für QBH-Systeme. Über die Eigenschaften von Melodiedatenbanken, die sich über statistische Methoden herleiten lassen, sind bislang keine Untersuchungen veröffentlicht worden. In der vorliegenden Arbeit sollen Melodiedatenbanken unter diesem Aspekt untersucht werden.

Der große Anteil existierender QBH-Systeme verwendet den oben beschriebenen PARSONS-Code zur symbolischen Darstellung von Melodien. Jüngere Arbeiten der *Moving Pictures Experts Group* (MPEG) bieten aber weitere, standardisierte Möglichkeiten zur Melodiebeschreibung im Standard MPEG-7. Darüberhinaus werden dort aber auch Definitionen angeboten, die zur Beschreibung von inhaltsbezogenen, abstrakten Daten geeignet sind. Damit lassen sich Systemschnittstellen wie die von QBH-Systemen repräsentieren. Im Rahmen dieser Arbeit wird die Anwendung des Multimedia-Standards MPEG-7 für QBH-Systeme ausführlich untersucht und diskutiert.

1.2 Fachgebiete

Die in einem QBH-System genutzten Technologien berühren ganz unterschiedliche Disziplinen und Fachgebiete. In Hinblick auf die Verarbeitung gesummter Melodien bzw. der Musiksignale für die Melodiedatenbank steht man vor Aufgaben der *Audiosignalverarbeitung*. Der dabei kardinale Begriff der Melodie und was sich damit verbindet, wird in den *Musikwissenschaften* beschrieben. Der Vergleich von symbolischen Melodiedarstellungen wird mit Methoden der *Informatik* vollzogen. Die Gewinnung von Daten aus Musiksignalen kann in diesem Zusammenhang den Disziplinen *Maschinenhören* und *Data-Mining* zugeordnet werden. Die Einbettung eines QBH-Systems in ein Netzwerk gehört gleichermaßen in das Aufgabengebiet der Informatik wie der *Elektrotechnik*. Es handelt es sich somit um eine *Multimedia*-Anwendung, die beispielsweise für *digitale Bibliotheken* genutzt werden kann.

Die Betrachtung von QBH-Systemen bringt also eine Fülle verschiedener Themen und Fachgebiete mit sich. Im Rahmen dieser Arbeit soll eine Darstellung der wichtigsten Grundlagen der genannten Fachgebiete erfolgen, um eine fundierte Bewertung von QBH-Systemen zu ermöglichen.

1.3 Wirtschaftliche Bedeutung

Nach dem Aufkommen von illegalen Musik-Tauschbörsen im Internet wie zum Beispiel „Napster“ um 1999 hat die Musikindustrie mittlerweile reagiert und bietet selbst Musik zum kostenpflichtigen Download im Internet an. Der große Ansturm auf solche Angebote steht Pressestimmen zufolge unmittelbar bevor [148]. Für 2005 prognostizierte der Bundesverband der phonographischen Wirtschaft 20 Millionen legale Downloads in Deutschland, nach acht Millionen im Vorjahr.

Feststellbar ist, dass der klassische Handel mit Tonträgern sich in Richtung Internethandel mit Dateien verschiebt. Die Firma *Apple* gibt für ihr Portal *iTunes* 35.000 verkaufte Titel pro Tag an [3], die Konkurrenz, u. a. auch *Microsoft*, folgt mit eigenen Angeboten. Auch unter diesem Gesichtspunkt kommt QBH-Systemen eine große Bedeutung zu.

1.4 Kapitelübersicht

Im Rahmen dieser Arbeit werden alle Teile eines QBH-Systems und Verfahren zu ihrer Funktionsprüfung untersucht. Zur praktischen Untersuchung wurde am Fachgebiet Nachrichtenübertragung ein eigenes System *Queryhammer* implementiert und ausführlich untersucht [27,28,66,91,180,195,196]. Die Komponenten des Systems werden in den einzelnen Kapiteln detailliert vorgestellt. Aus der Aufgabenstellung ergibt sich thematisch folgende Gliederung:

Musik und Melodie Melodiesuchsysteme sind Gegenstand dieser Arbeit, daher ist es notwendig, zunächst den Begriff der Melodie näher einzugrenzen. Damit verbunden werden in Kapitel 2 Begriffe der Musiktheorie erläutert, die zur Beschreibung von Musik und Melodien verwendet werden können. Ihre Eignung für QBH-Systeme wird im Einzelnen erörtert. Danach werden die technischen und musikalischen Aspekte der Melodiedarstellung erläutert. Weiterhin ist diesem Kapitel ein Abschnitt über das Singen von Melodien zugeordnet. Er enthält eine Übersicht von vorhandenen Untersuchungen in der Literatur, die das Singen von Melodien und speziell das Verhalten von Nutzern von QBH-Systemen behandeln.

Musiksuchmaschinen In Kapitel 3 wird der Stand der Technik anhand einiger Melodie- und Musiksuchsysteme dargestellt. Besondere Merkmale dieser Systeme wie verschiedene Eingabemöglichkeiten und an-

geschlossene Datenbanken werden diskutiert; die verwendeten Techniken werden kurz dargelegt. Die grundsätzlichen Anforderungen an ein QBH-System werden erläutert, darüberhinaus die Möglichkeiten der Anwendung in Netzwerken diskutiert. Abhängig vom Verwendungszweck wird erörtert, welche Anwender und Nutzergruppen ein Interesse an QBH-Systemen haben. Danach wird die technische Umgebung von QBH-Systemen mit Auswirkungen auf Hard- und Software erläutert. Aus diesen Betrachtungen werden schließlich die Zielbestimmungen für das Beispielsystem *Queryhammer* abgeleitet.

Multimedia-Standards Für die Beschreibung der Melodiekontur, aber auch weiterer Schnittstellen eines QBH-Systems eignen sich Multimedia-Standards, die in Kapitel 4 dargestellt werden. Während sich der Standard MPEG-7 vor allem auf die Beschreibung von Inhalten von Multimedia-daten konzentriert, können andere Standards wie MPEG-21 oder SMIL besonders für die Präsentation der Ergebnisse und Einbindung in Netzwerke benutzt werden. Die durch die Verwendung solcher Standards entstehenden Möglichkeiten werden erläutert und der besondere Bezug zu QBH-Systemen wird hergestellt.

Monophone Transkription Die Eingabe von gesummen Melodien und ihre Umwandlung in eine Melodiekontur bedeutet die Transkription eines monophonen Gesangssignals. Kapitel 5 klärt die verschiedenen Schritte der Signalverarbeitung, insbesondere Techniken der Grundfrequenzanalyse und Rhythmuserkennung. In eigenen Untersuchungen zur Transkription von gesummen Anfragen wird untersucht, welche Fehler auftreten und welche Ursachen sie haben. Weiterhin wird ein Verfahren zur möglichst objektiven Beurteilung von Transkriptionssystemen entwickelt.

Polyphone Transkription Kapitel 6 geht der Frage nach, wie Melodien in die Melodiedatenbank eingetragen werden können. Ein Überblick über bislang verwendete Methoden wie die Extraktion der Melodien aus MIDI-Dateien oder die Transkription durch Musiker diskutiert deren Vor- und Nachteile. Anschließend werden die aktuellen Möglichkeiten der Signalverarbeitung zur automatischen Extraktion von Melodien aus Musiksignalen dargelegt. Eigene Untersuchungen zeigen die Einsatzmöglichkeiten geeigneter Verfahren und wie weit damit die Transkription von Melodien aus polyphonen Signalen möglich ist.

Melodievergleich Liegen Anfrage und Melodie der Datenbank symbolisch vor, so kann der Vergleich dieser beiden Melodien erfolgen. Dieser Vergleich macht Ähnlichkeitsmaße notwendig, die eine anschließende Bewertung zulassen. Für QBH-Systeme, die mit Konturdarstellungen arbeiten, kommen Verfahren zur Zeichenkettensuche in Frage. Kapitel 7 erläutert häufig in QBH-Systemen verwendete Verfahren, ein besonderes Augenmerk gilt der Tauglichkeit der Verfahren für den Standard MPEG-7.

Melodiedatenbanken In Kapitel 8 wird untersucht, wie sich Eigenschaften der Melodiedatenbank auf Suchergebnisse in QBH-Systemen auswirken. Ausgehend von einigen statistischen Parametern wird ein Modell für Melodiedatenbanken vorgeschlagen. Danach werden die Ergebnisse aller vorangegangenen Kapitel zusammengefasst und anhand des Beispielsystems *Queryhammer* untersucht. Dazu wird erörtert, auf welchem Wege die Güte des Suchergebnisses eines QBH-Systems beurteilt werden kann. Für das Ergebnis einer Suchanfrage an ein QBH-System sind verschiedenste Parameter wie Größe und Inhalt der Datenbank, Qualität der Transkription und Art und Weise des Melodievergleichs ausschlaggebend. Die Abhängigkeit dieser Parameter voneinander wird untersucht und diskutiert.

Die **Zusammenfassung** der Arbeit gibt die wesentlichen Ergebnisse der Arbeit wieder, formuliert notwendige Konsequenzen für QBH-Systeme und gibt einen Ausblick auf weitere Themen.

*Musik wird oft nicht schön
gefunden,
weil sie stets mit Geräusch
verbunden.*

Wilhelm Busch

Im Vordergrund dieses Kapitels steht die Frage, *wonach* überhaupt gesucht wird, wenn man nach Musik sucht. Der Begriff „Musik“ leitet sich von „*musicé téchne*“ (griechisch: Kunst der Musen) ab und ist eine künstlerische Lebensäußerung des Menschen [199]. Besonderes Kennzeichen eines bestimmten Musikstücks ist oft die Melodie, vor allem in der westlichen Musik. Was unter dem Begriff Melodie allgemein verstanden werden kann und in welchem Sinn er für Query-by-Humming-Systeme (QBH-Systeme) verwendet werden soll, wird im nächsten Abschnitt dargelegt. Um Melodien beschreiben und notieren zu können, ist das Verständnis einiger Begriffe der Musiktheorie notwendig, die im darauf folgenden Abschnitt erklärt werden. Dies führt zur Frage, wie Musik technisch zu erfassen ist – Abschnitt 2.3 beschreibt daher verschiedene Formen der Musikrepräsentation. Schließlich wird das Singen von Melodien in Abschnitt 2.4 diskutiert; dies ist bedeutsam, da ein Mensch dabei natürlicherweise auch Einfluss auf die Melodie nimmt.

2.1 Der Melodiebegriff

Der Begriff der Melodie ist nicht klar definiert und bedarf näherer Erläuterung. In [53] findet man:

Melodie: die in der Zeit sich entfaltende selbständige Tonbewegung, die sich gegenüber weniger selbständigen Tonfolgen (Neben-, Begleit-, Füllstimmen) auszeichnet durch innere Folgerichtigkeit oder Gesanglichkeit oder leichtere Fasslichkeit oder durch Festigkeit und Geschlossenheit ihrer Gestalt und die als konkrete Erscheinung auch das rhythmische Element in sich enthält. [. . .]