

1 EINLEITUNG

Durch die rasante Entwicklung im Bereich der Netzwerk- und Speichertechnologien sind heutzutage große Datenarchive und weit verzweigte Netze zum Alltag geworden. Allen voran steht hier das World Wide Web (WWW) mit seiner äußerst heterogenen Netzwerk- und Speicherstruktur. Das WWW ermöglicht es einer großen Anzahl von unterschiedlichen Nutzern, gleichzeitig Daten abzurufen sowie dem Netzwerk hinzuzufügen. Diese dynamische Struktur macht es unmöglich, ein statisches Verzeichnis der Inhalte zu erstellen, da dieses bereits zum Zeitpunkt des Erscheinens überholt wäre. Die Katalogisierung von Webinhalten wird aktuell durch schlagwortbasierte Suchmaschinen wie Google oder Yahoo vorgenommen, die permanent den Textinhalt des WWW durchsuchen und in Datenbanken indizierte Verweise speichern [Goog08, Yaho08]. Nutzer des WWW können nach bestimmten Inhalten suchen, indem sie diese Datenbanken durchsuchen und den gespeicherten Verweisen zum eigentlichen Inhalt folgen. Es besteht jedoch keine Garantie, dass der Inhalt zum Zeitpunkt des Nutzerzugriffs noch vorhanden ist.

1.1 INHALTSBEZOGENE DATENSUCHE

Anfang des Jahres 2006 wurden mindestens 10 Milliarden Webseiten mit einem Umfang von ca. 200 TB in Suchmaschinen gelistet, wobei der Gesamtdatenumfang des WWW aufgrund der heterogenen Netzwerkstruktur und der volatilen Inhalte schwer zu bestimmen ist. Der nach außen hin sichtbare Teil des Netzes, das so genannte Surface Web, hat einen mindestens 20 mal größeren Datenbestand als die größte Biblio-

thek der Welt, die Library of Congress. Der Umfang des WWW wächst seit der Gründung im Jahre 1989 exponentiell und verdoppelt sich ungefähr alle 24 Monate [ShPr01, Tant06].

Bei den Inhalten des WWW handelt es sich heutzutage neben reinen Textdaten auch um große Sammlungen multimedialer Inhalte wie Musik-, Video- und Bilddaten. Diese sind entweder in Webseiten eingebettet oder in eigenständigen, dem WWW angegliederten Archiven verfügbar. Um multimediale Daten über klassische, schlagwortbasierte Suchmaschinen finden zu können, ist es erforderlich, dass zusätzlich zu den eigentlichen Inhalten Metainformationen verfügbar sind, die den Inhalt in Textform beschreiben und von Suchmaschinen abgefragt werden können. Bei in Webseiten eingebetteten Daten müssen zusätzlich Schlagwörter oder beschreibender Text vorhanden sein, bei Archiven müssen Schlagwörter oder beschreibender Text separat gespeichert werden. Inhaltsbeschreibende Metadaten sind im Zusammenhang mit Musikstücken beispielsweise der Interpret, der Titel, das Genre, das Tempo, die Tonart, die Notendarstellung oder, dem Thema dieser Arbeit entsprechend, Informationen über die in dem Musikstück enthaltenen Instrumente.

Im einfachsten Fall können Metadaten ohne spezielle Formatierung als Text oder als Aufzählung von Schlagwörtern gespeichert werden, bei größeren Datenbeständen bietet sich eine Auszeichnungssprache wie die Extensible Markup Language (XML) oder eine Datenbanklösung an, was die Möglichkeiten einer gezielten Suche vergrößert. Im Jahre 2002 wurde in diesem Zusammenhang der Standard MPEG-7 [ISO15938] veröffentlicht, der unter anderem eine einheitliche Annotation von multimedialen Daten mit Metadaten spezifiziert und so eine erhöhte Vergleichbarkeit und erweiterte Suchfunktionen ermöglicht.

Mit textbasierten Metadaten ist es möglich, die hoch entwickelte Technologie der vorhandenen, schlagwortbasierten Suchmaschinen direkt weiter zu verwenden, inklusive der vorhandenen Fehlerkorrektur bei Tippfehlern und der Möglichkeit von Alternativvorschlägen bei wenigen gefundenen Suchergebnissen. Typische Anfragen für Musikstücke könnten beispielsweise lauten: „langsames Tempo, Michael Jackson, Gitarre“ oder „Klavier, männlicher Sänger, Pop“.

Metadaten müssen heutzutage größtenteils noch manuell annotiert werden, oftmals sind deshalb, gerade wegen der anwenderoffenen Struktur des WWW, Metadaten nicht oder nur unzureichend vorhanden. Somit ist eine große Menge an multimedialen Daten im WWW gespeichert, deren Inhalte von außen jedoch nicht zu erkennen sind und einem suchenden Benutzer verborgen bleiben. Der Umfang dieses nicht durch textbasierte Suchmaschinen erfassten, so genannten Deep Web wird etwa 50 bis 500 mal größer als das gelistete Surface Web geschätzt [ShPr01, Harr06, Tant06].

Damit die inhaltsbeschreibenden Metadaten nicht zu jeder Datei manuell erstellt werden müssen, konzentriert sich die Forschung derzeit auf die Entwicklung von Algorithmen, mit denen diese inhaltsbeschreibenden Metadaten automatisch aus den zu beschreibenden Dateien extrahiert werden können. Durch die automatische Extraktion von Metadaten wird es Suchmaschinen ermöglicht, multimediale Inhalte auf semantischer Ebene zu durchsuchen, ohne vollständig auf eine vorhandene korrekte Annotation des Inhalts angewiesen zu sein. Weiterhin werden durch die automatische Extraktion von Metadaten inhaltliche Vergleiche von multimedialen Inhalten möglich. Somit kann der Nutzer einer Suchmaschine neben schlagwortbasierten Anfragen auch Ähnlichkeitsabfragen formulieren. In Query by Humming Systemen (QBH, Anfrage durch Summen) kann ein Nutzer beispielsweise als Anfrage die Melodie eines gesuchten Musikstücks summen [BEW+04a, Batk06], Query by Tapping Systeme (QBT, Anfrage durch Klopfen) erlauben die Formulierung einer Suchanfrage über das Klopfen der Rhythmik des Stücks und sind auch auf Endgeräten ohne Mikrofon implementierbar [EiBS04a, EiBS04b]. Der Nutzer könnte als weitere Ähnlichkeitsabfrage auch ein Musikstück angeben, das eine große klangliche Ähnlichkeit mit einem von ihm gesuchten Musikstück aufweist [MaSS02, TzEC02, WeCr03].

Für Ähnlichkeitsabfragen ist neben der Qualität des verwendeten Algorithmus zur Extraktion der Metadaten der Detailgrad der Beschreibung von entscheidender Bedeutung. Bei einer zu detaillierten Beschreibung wird es schwierig, Ähnlichkeiten bei geringen Abweichungen noch zu erkennen, ist die Beschreibung hingegen zu ungenau, wird eine zu große Ähnlichkeit bei unterschiedlichen Inhalten erzeugt. Für die Verarbeitungsdauer einer Ähnlichkeitsabfrage ist es von entscheidender Bedeu-

tung, wie schnell und effizient sich die einzelnen Metadatenätze vergleichen lassen.

1.2 MUSIKINSTRUMENTENERKENNUNG

Im Zusammenhang mit der automatischen Extraktion von Metadaten aus Musikstücken wurden im Rahmen dieser Arbeit verschiedene Algorithmen umfangreich untersucht, die automatisch die in Musikstücken auftretenden Musikinstrumentenklänge und Geräusche identifizieren und klassifizieren. Diese Aufgabe lässt sich in die zwei Teilgebiete der Merkmalsextraktion und der Klassifikation unterteilen. Algorithmen der Merkmalsextraktion gewinnen aus Musikstücken oder einzelnen Musikinstrumentenklängen Informationen, die bestimmte Aspekte des analysierten Signals widerspiegeln. Die Klassifikation beruht auf Algorithmen der Mustererkennung und ordnet bestimmte Merkmalsmuster bekannten Musikinstrumentenklassen zu. Somit wird eine Entscheidung über die zu hörenden Musikinstrumente und Geräusche gefällt.

Im Rahmen dieser Arbeit wurden weiterhin drei Systeme zur automatischen Identifikation und Klassifikation von Musikinstrumentenklängen entwickelt, mit denen sich unterschiedliche theoretische und technische Aspekte erforschen und demonstrieren lassen. Unterschiedliche Musikinstrumentenklänge und Geräusche werden von den Systemen ihrem Ursprung entsprechend klassifiziert, wobei Klänge von gleichen Instrumentenarten jeweils in einer Klasse zusammengefasst werden. Die Systeme sind lernfähig und weisen ein Ordnungsverhalten auf, das dem eines musikalisch geschulten Menschen entspricht.

Das erste System ist ein monophones Experimentiersystem, das einzelne monophone Klänge oder Geräusche klassifizieren kann. Es ist modular aufgebaut, so dass sich die einzelnen Prozessstufen sehr einfach austauschen lassen. Somit lassen sich unterschiedlichste Algorithmen der Merkmalsextraktion, Merkmalsaufbereitung und Klassifikation sehr effizient sowohl einzeln als auch im Zusammenspiel testen. Dieses System bildet die Basis für die beiden anderen Systeme, die aufgrund der gewonnenen Erkenntnisse in unterschiedliche technische Richtungen weiterentwickelt und optimiert wurden.

Das zweite System ist ein monophones Echtzeitsystem, das ebenfalls einzelne aufeinander folgende monophone Klänge oder Geräusche klassifizieren kann, allerdings arbeitet es als echtzeitfähiger Prozess. Das System stellt eine Studie für den industriellen Gebrauch dar und wurde als Plugin für die im Bereich der Studiotechnik weit verbreitete VST-Schnittstelle [Ste108] implementiert. Es verwendet Algorithmen, die ein für den Echtzeitbetrieb nötiges, gutes Verhältnis zwischen effizienter Berechenbarkeit und guten Klassifikationsergebnissen haben.

Das dritte System ist ein polyphones Klassifikationssystem, das in der Lage ist, harmonische Musikinstrumentenklänge in komplexer polyphoner Musik zu identifizieren und zu klassifizieren. Dieses System benutzt Algorithmen, die sehr robust gegenüber Störeinflüssen durch überlappende oder gleichzeitig auftretende Klänge oder Geräusche sind. Das entwickelte System verwendet die Merkmale der Musikinstrumentenklänge nicht nur zur Analyse und Identifikation bzw. Klassifikation, sondern ist auch in der Lage, die Klänge durch ein auf dem Prinzip der Additiven Resynthese basierendes harmonisches Modell zu resynthetisieren.

1.3 STUDIOTECHNOLOGIE

Die entwickelten Systeme leisten neben den genannten Anwendungsmöglichkeiten im Bereich der semantischen Suche und automatischen Metadatenerzeugung zusätzlich einen wesentlichen Beitrag im Rahmen der Studiotechnologie. So können die extrahierten Informationen über die in einem Musikstück erkannten Musikinstrumente beispielsweise dazu genutzt werden, die Melodie des Stücks von einem ähnlich klingenden Instrument doppeln zu lassen. Die Auswahl eines geeigneten Instruments könnte hierbei nach gewissen Vorgaben automatisch durchgeführt werden. Datenbanken mit Synthesizerklängen, wie sie in Musikstudios vielfältig verwendet werden, könnten automatisch nach ihren klanglichen Eigenschaften in einer Datenbank angeordnet werden. Diese Datenbank könnte durch die Benutzung bestimmter klanglicher Eigenschaften und Ähnlichkeiten durchsucht werden.

Da sich im Bereich der musikalischen Metadatengewinnung viele Verfahren gegenseitig beeinflussen und eng miteinander verflochten sind, können die von den implementierten Systemen ermittelten Informatio-

nen vielfältige andere Aufgaben erleichtern. So können beispielsweise für Verfahren der Quellentrennung sowie der automatischen Transkription Signalmodelle verwendet werden, die auf die enthaltenen Musikinstrumente angepasst sind. Weiterhin ergeben sich Vereinfachungen für die in QBH- oder QBT-Systemen durchzuführenden Transkriptionen von komplexen Musikstücken in monophone Melodien oder Rhythmen [EiBS04a, EiBS04b, BEW+04b, Batk06].

1.4 KAPITELÜBERSICHT

Aus den für die vorliegende Arbeit betrachteten Aspekten und durchgeführten Forschungen ergibt sich thematisch die folgende Gliederung:

- **Musikterminologie:** Neben der Definition von Hör- und Klangeignissen und ihrer Parameter werden in Kapitel 2 alle für diese Arbeit relevanten musikalischen Phänomene sowie die Parameter der Musik erläutert. Weiterhin werden Tonarten und ihre Bedeutung im Zusammenhang mit der Musikinstrumentenidentifikation und Klassifikation beschrieben.
- **Musikinstrumente:** Die physikalischen Eigenschaften von Musikinstrumenten sowie die daraus resultierenden Klangeigenschaften werden in Kapitel 3 untersucht. Hierbei werden insbesondere charakteristische Eigenschaften der Dynamik und spektrale Eigenschaften betrachtet, die für die Identifikation von Musikinstrumentenklängen verwendet werden können.
- **Merkmalsextraktion:** Verfahren zur Extraktion relevanter Merkmale aus Musikstücken und den in ihnen vorkommenden Musikinstrumentenklängen werden in Kapitel 4 analysiert. Die vorgestellten Merkmale werden hinsichtlich ihrer Verwendbarkeit für die Beschreibung von Musikinstrumentenklängen und hinsichtlich ihrer effizienten Berechenbarkeit untersucht.
- **Klassifikation:** In Kapitel 5 werden verschiedene Verfahren der Mustererkennung vorgestellt und bezüglich ihrer Verwendbarkeit im Zusammenhang mit der Identifikation und Klassifikation von Musikinstrumentenklängen analysiert.