

# Praktikumsbericht

## zum Fachpraktikum

Beginn 01.08.2007

Ende 15.01.2008

Name Jakob Abeßer

Adresse Am Stollen 1  
98693 Ilmenau

Telefon +49 (0) 3677 / 622033

E-Mail Jakob.Abesser@stud.tu-ilmenau.de

Matrikelnummer 35081

Matrikel 2002

Studiengang Ingenieurinformatik

Vertiefung Telekommunikations- &  
Messtechnik

Fakultät Infomatik & Automatisierung

Universität TU-Ilmenau

Firma	Fraunhofer Institut für Digitale Medientechnologie (IDMT)
Anschrift	Ehrenbergstraße 29 98693 Ilmenau
Abteilung	Metadaten
Anschrift	Langewiesener Straße 22 98693 Ilmenau
Betreuer	Dipl.-Ing. (FH) Christian Dittmar
Telefon	+49 (0) 36 77/4 67-2 06
E-Mail	dmr@idmt.fraunhofer.de

Bestätigung der Berichtseinsicht

Datum

Unterschrift

Anwesenheitstage	104
Krankheitstage	0

# Inhaltsverzeichnis

<b>1</b>	<b>Vorstellung des Unternehmens</b>	<b>2</b>
1.1	Fraunhofer Gesellschaft . . . . .	2
1.2	Fraunhofer Institut für digitale Medientechnologie . . . . .	2
1.2.1	Institutsprofil . . . . .	3
1.2.2	Aufbau & Struktur . . . . .	3
<b>2</b>	<b>Aufgaben und Ziele</b>	<b>4</b>
2.1	Offizielle Aufgabenstellung . . . . .	4
2.2	Ziele . . . . .	5
<b>3</b>	<b>Bestehendes System</b>	<b>6</b>
3.1	Grafischer und funktioneller Aufbau . . . . .	7
3.2	Programmstruktur . . . . .	13
<b>4</b>	<b>Ablauf des Praktikums</b>	<b>15</b>
4.1	Nutzerinteraktion mit dem System . . . . .	16
4.2	Das Expert-Mode-Fenster . . . . .	17
4.3	Implementierung . . . . .	21
<b>5</b>	<b>Fazit des Praktikums</b>	<b>23</b>
5.1	Programmieren . . . . .	23
5.2	Ausblick . . . . .	23
<b>A</b>	<b>Arbeitsgrundlage - Liste an geplanten neuen <i>Features</i> für die <i>Transcription Toolbox</i></b>	<b>24</b>
<b>B</b>	<b>Abkürzungsverzeichnis</b>	<b>26</b>

# **1 Vorstellung des Unternehmens**

## **1.1 Fraunhofer Gesellschaft**

Die Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e.V. (FhG) [1] ist die europaweit größte Forschungsorganisation und hat ihren Sitz in München. Sie wurde 1949 mit dem Ziel gegründet, anwendungsorientierte Forschung zum direkten Nutzen von Unternehmen und zum Vorteil der Gesellschaft zu betreiben. Namensgeber der Organisation ist Joseph von Fraunhofer (1787-1826), der seinerzeit sowohl als Forscher und Erfinder sowie als Unternehmer erfolgreich tätig war. Die Fraunhofer Gesellschaft besitzt heute deutschlandweit 56 Institute und betreibt insgesamt mehr als 80 Forschungseinrichtungen, in denen etwa 12500 Angestellte arbeiten. Der Großteil des jährlichen Forschungsvolumens von etwa 1,2 Milliarden Euro wird durch Auftragsforschung aus der Industrie und durch öffentlich finanzierte Forschungsprojekte erwirtschaftet.

Die internationale Zusammenarbeit mit anderen Firmen und Forschungseinrichtungen wird durch Niederlassungen in den USA, Asien, im nahen Osten und in anderen europäischen Ländern gefördert.

## **1.2 Fraunhofer Institut für digitale Medientechnologie**

Das Fraunhofer Institut für digitale Medientechnologie (IDMT) [2] hat seinen Sitz in Ilmenau und wurde am 1. Januar 2004 als eigenständiges Institut gegründet. Vorausgegangen war die Gründung der Fraunhofer-Arbeitsgruppe für Elektronische Medientechnologie als Außenstelle des Fraunhofer-Instituts für Integrierte Schaltungen (IIS) in Erlangen im Mai 2000.

Am IDMT arbeiten z.Z. etwa 50 fest angestellte Mitarbeiter und 70 studentische Hilfskräfte, der Institutsleiter ist Prof. Dr.-Ing. Karlheinz Brandenburg.

### **1.2.1 Institutsprofil**

Die Forschungsgebiete am Fraunhofer Institut für Digitale Medientechnologie liegen hauptsächlich in der professionellen Audiotechnik und Unterhaltungselektronik. 2003 wurde das auf dem Prinzip der Klangfeldsynthese basierende *IOSONO*-System vorgestellt, für das seit dem neue Applikationen entwickelt werden. Weitere Forschungsschwerpunkte liegen im Bereich der Medizintechnik, in der Entwicklung von Vertriebsmodellen für digitale multimediale Inhalte, der Audiokodierung für spezielle Anforderungen und der semantischen Metadatenanalyse.

### **1.2.2 Aufbau & Struktur**

Das Institut ist in 7 Arbeitsbereiche aufgeteilt:

1. Akustik
2. Semantische Metadatensysteme
3. Media Distribution and Security
4. Audiocodierung für spezielle Anwendungen
5. Medizinische Audiotechnik
6. Data Representation and Interfaces
7. Autorensysteme

Im Arbeitsbereich für semantische Metadatensysteme, in dem ich mein Praktikum absolvierte, wird an Methoden zur automatisierten Erkennung und Beschreibung inhaltsbasierter Merkmale aus multimedialen audiovisuellen Inhalten geforscht. Eine wichtige Anwendung dieser Forschung ist die möglichst zielgerichtete Suche innerhalb großer multimedialer Datenbanken.

Neben der Bild- und Videoverarbeitung liegt der Forschungsschwerpunkt vor allem im Bereich der Audioverarbeitung. Dabei wird vor allem im Bereich der

Ähnlichkeitsanalyse von Musikstücken, der automatisierten Erstellung von Abspiellisten (*playlists*) aus großen Datenbanken und der Analyse und automatischen Extraktion von musikalischen Merkmalen wie Melodie- und Harmonieverlauf, Taktart und Tempo geforscht.

## 2 Aufgaben und Ziele

### 2.1 Offizielle Aufgabenstellung

Thema Nr.: 2007/701, erstellt am: 05.06.2007

*Konzeption und Implementierung von Nutzerinteraktion in ein System zur automatischen Transkription von Musikstücken*

In der Abteilung Metadaten am Fraunhofer IDMT wird intensiv an Verfahren zur automatischen Transkription von Musikstücken geforscht. Dabei ist das Ziel, mit möglichst hoher Präzision die gespielten Noten verschiedener Instrumente zu identifizieren und in eine symbolische Form – z.B. in Form des *MIDI*<sup>1</sup>-Formats – zu übertragen. Einerseits lassen sich diese Metadaten nutzen, um große Musikarchive zu annotieren aber auch um automatische Musikempfehlung und Suche per Melodie (*QbH*<sup>2</sup>) zu ermöglichen. Des Weiteren ist automatische Transkription interessant für den Bereich elektronische Musikinstrumente, Sequenzer, Audio-Editoren. Ein wichtiger Gesichtspunkt ist hierbei die Möglichkeit der Interaktion seitens des Anwenders, um mögliche Fehler in der automatischen Extraktion zu korrigieren.

Im Rahmen dieser Arbeit sollen Konzepte zur Einbeziehung von Nutzerinteraktion erstellt und in ein existierendes System zur Musiktranskription integriert werden. Dieses System ist derzeit in der Lage, Noten der Hauptmelodie, der Bassspur, der Schlagzeugspur, sowie die gespielten Akkorde zu identifizieren. Ziel des Prakti-

---

<sup>1</sup>Musical Instrument Digital Interface

<sup>2</sup>Query by Humming

kums ist es, dem Anwender für die verschiedenen Bereiche Möglichkeiten zum Eingriff in die Extraktion und zur Korrektur der Ergebnisse anzubieten. Beispielsweise kann der Nutzer durch Vorgabe der gespielten Tonart in mehrdeutigen Fällen die Fehlinterpretation von Noten verhindern. Weiterhin kann bei der Schlagzeugtranskription vorher festgelegt werden, welche Schlaginstrumente tatsächlich vorkommen oder welche von 3 Tempo-Hypothesen die zutreffende ist. Die dazu nötigen Konzepte und Voruntersuchungen werden in MATLAB implementiert, vielversprechende Ansätze werden dann im Laufe des Praktikums nach C++ portiert.

Voraussetzungen für die Arbeit sind gute Kenntnisse in C/C++, MATLAB, Musiktheorie und Signalverarbeitung.

Verantwortlicher Hochschullehrer:

Prof. Dr.-Ing. Karlheinz Brandenburg

Betreuer:

Dipl.-Ing. (FH) Christian Dittmar

Dipl.-Ing. Holger Grossmann

Kontakt:

Dipl.-Ing. (FH) Christian Dittmar

dmr@idmt.fraunhofer.de

## **2.2 Ziele**

Aus der im vorherigen Kapitel dargestellten allgemeinen Aufgabenstellung lassen sich folgende Ziele ableiten:

- Einarbeitung in das bestehende System zur automatischen Musiktranskription



- Entwicklung neuer Konzepte zur Implementierung von Nutzerinteraktion in Bezug auf eine nachfolgende Korrektur der Transkriptionsergebnisse
  - Konzeption und Umsetzung der entsprechenden Funktionalitäten
  - Entwurf einer Erweiterung der grafischen Oberfläche für das bestehende Programm
- Test aller Funktionalitäten

### 3 Bestehendes System

Das *Transcription Toolbox* - System wurde im Jahr 2006 von Dipl.-Ing. (FH) Christian Dittmar und Dipl.-Inf. Christoph Bastuck am Fraunhofer IDMT implementiert. Ziel war es dabei, alle bis dahin im Rahmen verschiedener Diplomarbeiten am IDMT entwickelten Technologien zur automatischen Transkription verschiedener Klassen von Musikinstrumenten in einer Software zu bündeln um diese als geschlossenes System vorführen zu können [5]. Hierbei können Schlagzeug- und Percussioninstrumente (DRUM TRANSCRIPTION), Bassinstrumente wie E- oder Kontrabass (BASS TRANSCRIPTION), Melodieinstrumente wie Saxofon oder auch Gesang (MELODY TRANSCRIPTION) und Harmonieinstrumente wie Klavier oder Gitarre (HARMONY TRANSCRIPTION) automatisch transkribiert werden. Dies bedeutet, dass aus einem real aufgenommenen Musikstück, in dem im Allgemeinen mehrere verschiedene Instrumente vertreten sind, die konkret gespielten Noten der einzelnen Instrumente mit ihren Parametern Tonhöhe, Tondauer, Tonstartzeitpunkt und Lautstärke automatisch extrahiert werden können.

Mögliche Einsatzgebiete für die automatische Musiktranskription finden sich z.B. beim Erlernen von Musikstücken. Hierbei kann das *Transcription-Toolbox* als zeitsparendes Hilfsmittel eingesetzt werden. Ebenfalls denkbar ist die automatisierte Transkription großer Mengen von Musikstücken aus Datenbanken und anschließender inhaltsbasierter Analyse musikalischer Merkmale wie Tonart oder Genre oder auch automatischer Segmentierung.

### 3.1 Grafischer und funktioneller Aufbau

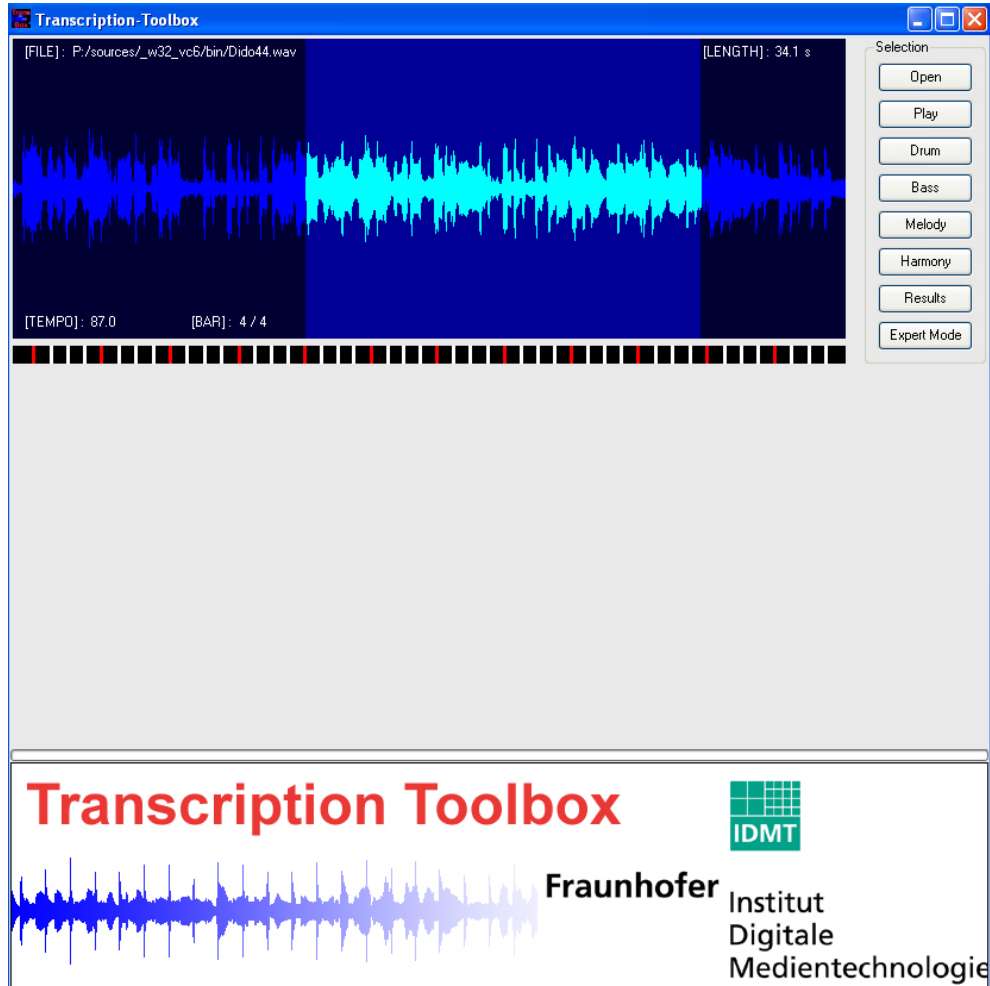


Abbildung 1: Transcription Toolbox - Startansicht

Die grafische Oberfläche der *Transcription Toolbox* ist in Abbildung 1 zu sehen. Über den Menüpunkt OPEN können Audiodateien im unkomprimierten WAV- oder im komprimierten MP3-Format mit einer Abtastrate von 44,1 kHz und einer Auflösung von 16bit (mono) eingelesen werden. Unmittelbar nach dem Einlesen der Abtastwerte erfolgt eine initiale Analyse des Musikstückes bezüglich seiner Tonart, seines Tempos und seiner Taktart. Hierfür wird auf verschiedene Funk-

tionalitäten der ebenfalls im Fraunhofer IDMT entwickelten *XProEx*<sup>3</sup>-Technologie zurückgegriffen, die eine Vielzahl von Modulen für die Audio- und Videosignalverarbeitung bereitstellt, welche je nach Anwendung unterschiedlich verknüpft werden können.

Die gewonnenen Parameter werden neben der Wellenform des Stückes im oberen Frame des Programmfensters grafisch dargestellt und sollen dem Benutzer als Hypothesen für eine erste grobe musikalische Beschreibung und Einordnung des Stückes dienen. Über den Button PLAY kann das Audiostück abgespielt werden.

Über die Menüpunkte DRUM, BASS, MELODY und HARMONY gelangt man in die Programmebenen der vier verschiedenen Transkriptionsalgorithmen, welche mittels APPLY gestartet werden. Über eine Statusanzeige unterhalb des oberen Frames kann der aktuelle Fortschritt der Transkription verfolgt werden.

Nachdem die Ergebnisse der Transkription vorliegen, werden im mittleren Frame des Hauptfensters die detektierten Notenereignisse entsprechend ihrer Auftretenszeitpunkte dargestellt (siehe Abbildung 2, 3 und 4). So wird ein direkter visueller Bezug zum zeitlichen Verlauf des Originalsignals innerhalb des oberen Frames ermöglicht.

Für die Melodietranskription wurde eine Darstellung implementiert, in der sowohl die erkannten Notenevents (blau) als auch der Verlauf der Fundamentalfrequenzen (orange) des erkannten Melodieinstruments dargestellt sind (siehe Abbildung 3).

Für die Schlagzeugtranskription werden die Noten als vertikale Linien dargestellt, da die Dauer einzelner Schläge eines Percussioninstruments in Bezug auf ihren zeitlichen Abstand sehr gering ist. In Abbildung 2 sind die Ergebnisse der Basstranskription und in Abbildung 4 die der Schlagzeugtranskription exemplarisch für das Lied „White Flag“ der Sängerin „Dido“ gegenübergestellt.

Über den Menüpunkt PLAY im mittleren Frame können die Resultate der Transkription in Form eines synthetisierten Audiosignals angehört werden.

Für die Wiedergabe der Ergebnisse der Schlagzeugtranskription wird dazu auf in-

---

<sup>3</sup>Extensible Processor Extractor

tern gespeicherte *Samples* – also reale Aufnahmen der entsprechenden Instrumente – zurückgegriffen. Diese werden entsprechend der erkannten Startzeitpunkte (*onset time*) zu einem Audiosignal zusammengemischt, welches die vom System transkribierte Schlagzeugspur repräsentiert. Für die Bass-, Melodie- und Harmonietranskription werden für jede erkannte Note neben dem Startzeitpunkt auch die Länge und der zur entsprechenden Tonhöhe äquivalente Notenwert gespeichert. Anhand dieser Informationen wird durch Überlagerung mehrerer sinusförmiger Signale der jeweiligen Frequenz und Dauer und einer anschließenden Nachbearbeitung (Verzerrung) ein synthetisches Audiosignal erzeugt.

Im Unterpunkt RESULTS des Hauptmenüs erhält der Nutzer die Möglichkeit, alle synthetischen Signale und das Orginalsignal gleichzeitig abzuspielen, wobei er über entsprechende Schieberegler direkten Einfluss auf die Lautstärken der einzelnen Spuren nehmen kann. So kann er über einen direkten Vergleich die Qualität der Transkriptionsergebnisse untersuchen.

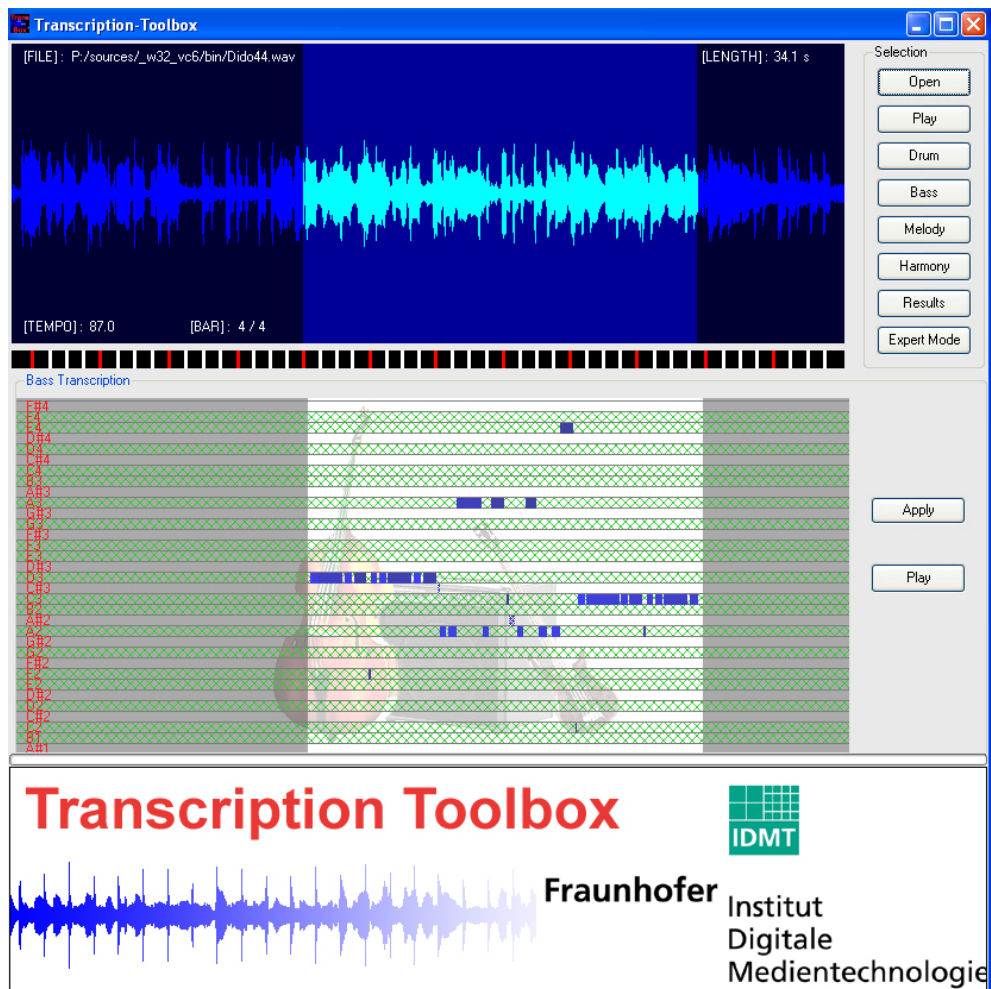


Abbildung 2: Transcription Toolbox - Ansicht nach Ausführen der BASS TRANSCRIPTION

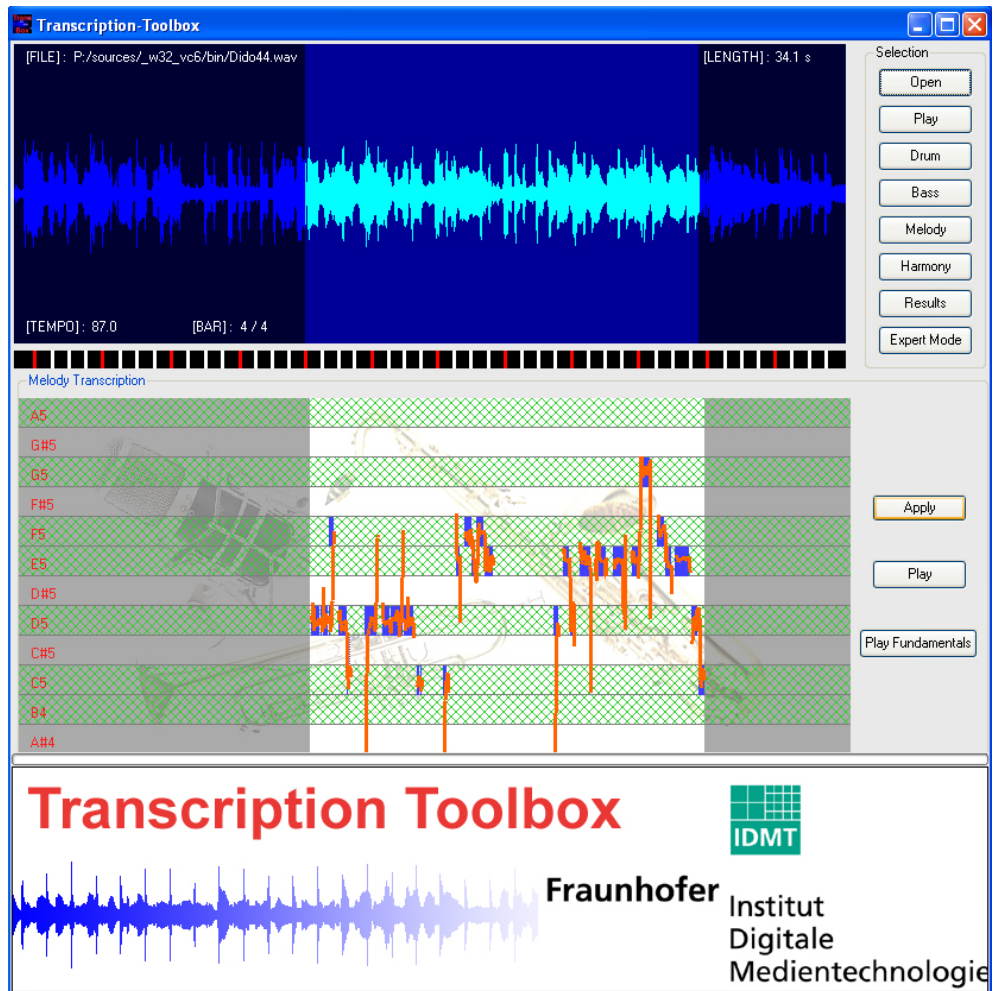


Abbildung 3: Transcription Toolbox - Ansicht nach Ausführen der MELODY TRANSCRIPTION

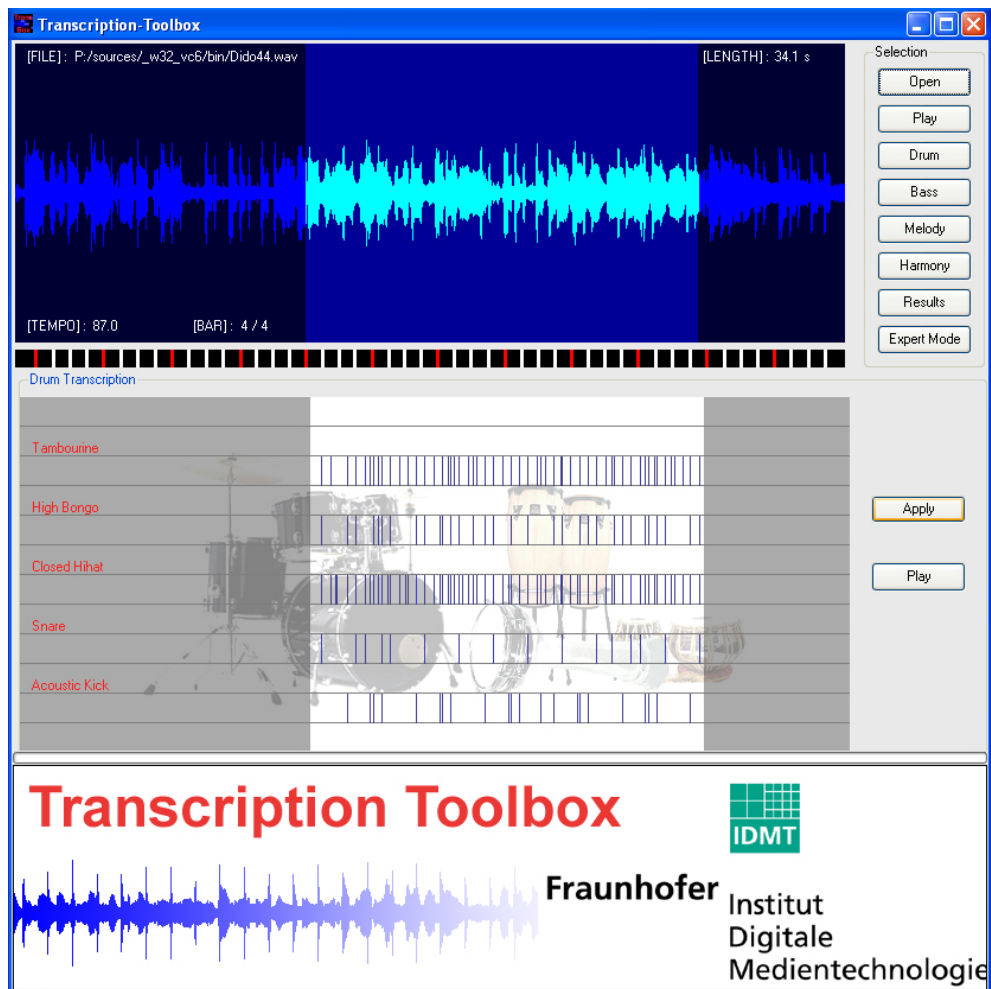


Abbildung 4: Transcription Toolbox - Ansicht nach Ausführen der DRUM TRANSCRIPTION

### 3.2 Programmstruktur

Die *Transcription Toolbox* ist in der objektorientierten Programmiersprache C++ implementiert, wobei für die Umsetzung der grafischen Oberfläche die Softwarebibliothek *QT* der Firma *Trolltech* [3] verwendet wird. Auf das Arbeiten mit C++ und QT wird im Kapitel 5.1 näher eingegangen.

Das *Transcription Toolbox* - System besteht aus einer Vielzahl von Klassen, die über vordefinierte Schnittstellen miteinander interagieren. Der prinzipielle hierarchische Aufbau ist in Abbildung 5 dargestellt.



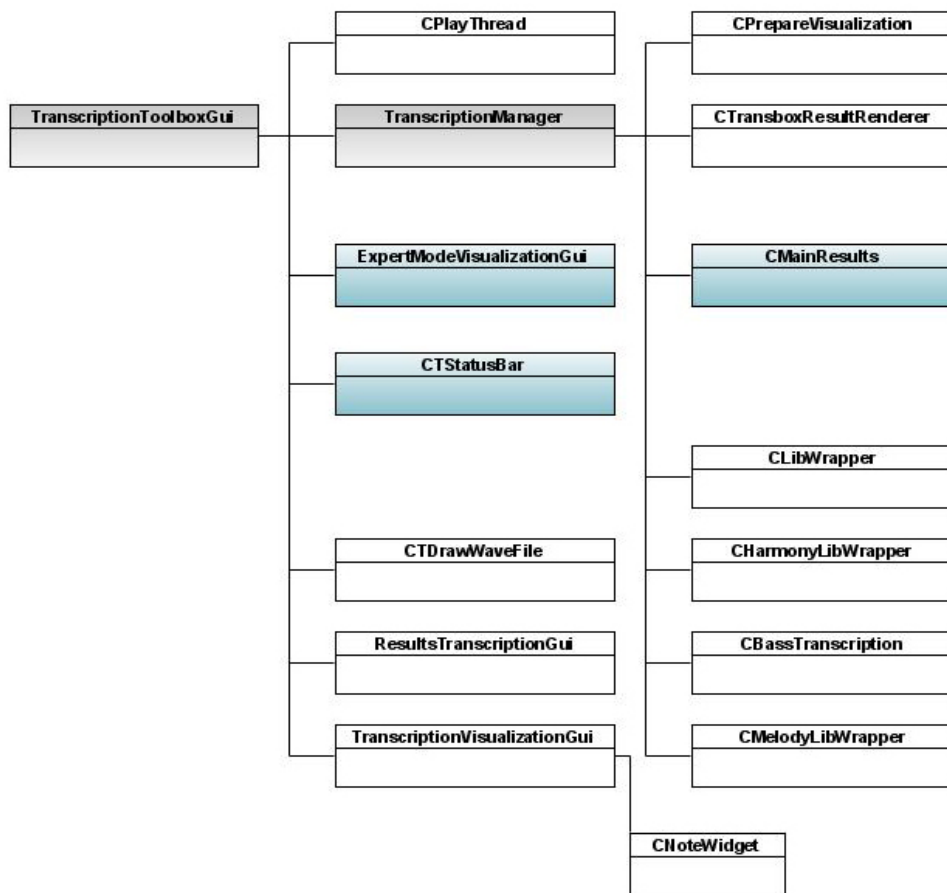


Abbildung 5: Transcription Toolbox - Prinzipieller Programmaufbau - Übersicht über die wichtigsten Klassen

Generell wurde bei der Entwicklung darauf Wert gelegt, dass alle Funktionalitäten für die grafischen Ausgaben und für die eigentliche Transkription und ihrer Auswertung in unterschiedlichen Klassen implementiert wurden.

Für die grafische Ausgabe der Wellenform des Stückes dient die Klasse CTDRAW-WAVEFILE, für die Darstellung der Transkriptionsergebnisse im Hauptfenster die Klassen RESULTSTRANSSCRIPTIONGUI, TRANSCRIPTIONVISUALIZATIONGUI bzw. CNOTEWIDGET für die konkreten Noten.

Die Schlagzeug-, Bass-, Melodie- und Harmonietranskription wurden in voneinander unabhängigen Modulen umgesetzt. Dies liegt darin begründet, dass die Tran-

skriptionsalgorithmen im Rahmen verschiedener Diplom- und Studienarbeiten innerhalb der letzten Jahre unabhängig voneinander entstanden sind und nun alle durch die Transcription Toolbox verwendet werden.

Über eine einheitliche Schnittstelle – die sogenannte *Wrapper*-Klasse – können die Transkriptionsalgorithmen mit den aktuellen Audiodaten aus dem Programm gestartet werden. Sie ist für alle 4 Transkriptionsmodule implementiert (CLIBWRAPPER für die Schlagzeugtranskription, CHARMONYLIBWRAPPER für die Harmonietranskription, CBASSTRANSSCRIPTION für die Basstranskription und CMELODYLIBWRAPPER für die Melodietranskription). Zur grafischen Ausgabe werden die Ergebnisse noch in der Klasse CPREPAREVISUALIZATION vorverarbeitet. Zur Audiowiedergabe mittels CPLAYTHREAD erfolgt analog eine Vorverarbeitung in CTRANSBOXRESULTRENDERER. Die in der Abbildung 5 blau hervorgehoben Klassen sind die, die im Laufe meines Praktikums komplett neu implementiert wurden, alle anderen wurden jeweils modifiziert.

Die Klasse CMAINRESULTS dient der Speicherung von Programmparametern (siehe Kap. 4.3), EXPERTMODEVISUALIZATIONGUI enthält die Implementierung des *Expert Mode* - Fensters (siehe Kap. 4.2) und CSTATUSBAR die Implementierung des Bereiches im Hauptfenster zur Darstellung des Metronoms (siehe Kap. 3.1).

## 4 Ablauf des Praktikums

Innerhalb der ersten 3 Wochen erarbeitete ich mir einen Überblick über das bisher implementierte System, über seinen Aufbau in verschiedene Klassen und ihre Abhängigkeiten untereinander (siehe auch Kapitel 3.2). Im Anschluss daran begann die Umsetzung konkreter Konzepte für eine Erweiterung der bestehenden grafischen Oberfläche. Hauptziel dabei war es, das System in Hinblick auf Nutzerinteraktion zu erweitern. Neben der in Kapitel 2.1 enthaltenen offiziellen Aufgabenstellung stellte eine Liste mit geplanten neuen Programmmerkmalen (*features*)

für das *Transcription Toolbox* - System (siehe Anhang A) in den darauf folgenden Monaten meine Arbeitsgrundlage dar.

Die Einarbeitung in die Softwarebibliothek *QT* (siehe auch Kapitel 4.3), welche zur Programmierung der grafischen Oberfläche notwendig war, erfolgte dank umfangreicher Dokumentation – u. a. auf der Website der Firma *Trolltech* [3] – relativ schnell, sodass ich Mitte September mit der konkreten Realisierung der neuen Programmmerkmale beginnen konnte.

Durch regelmäßige Rücksprache mit meinem Betreuer bezüglich einzelner Ideen, aber auch bezüglich des Gesamtkonzeptes des *Expert-Mode*-Fenster, das in den folgenden Abschnitten im Detail erläutert wird, konnten im Laufe der folgenden Monate Schritt für Schritt die einzelnen Erweiterungen am System umgesetzt werden.

Dabei erfolgten regelmäßige Tests, wobei die einzelnen Transkriptionsalgorithmen und die neu umgesetzten Programmmerkmale an konkreten Audiodateien getestet wurden, und auftretende Fehler lokalisiert und behoben wurden. Mit wachsender Komplexität der Software wurde dieser Schritt immer wichtiger, um einen fehlerfreien Betrieb sicherzustellen.

Zum Ende meines Praktikums erweiterte ich die bestehende englischsprachige Programmdokumentation [4], die Dipl.-Ing. (FH) Christian Dittmar 2006 geschrieben hatte, um aktuelle Abbildungen der Systemoberfläche (*screenshots*) des Programms und um das Kapitel „3.6. Expert Mode“, in dem die innerhalb meines Praktikums neu implementierten Programmfeatures im Detail erläutert werden.

## **4.1 Nutzerinteraktion mit dem System**

Die *Transcription Toolbox* soll dem Benutzer die Möglichkeit geben, verschiedene Instrumentalspuren aus einem von ihm gewählten Musikstück automatisch zu transkribieren und ihm damit konkrete rhythmische, melodische und harmonische Informationen über das Stück zurückgeben.

Die Tatsache, dass die einzelnen Transkriptionsalgorithmen eine Erkennungsquo-

te im Bereich zwischen 60% und 85% haben, führt zu der Notwendigkeit, eine Nachbearbeitung bzw. Korrektur der gewonnenen Ergebnisse zu ermöglichen. In den folgenden beiden Kapiteln werden die im Laufe des Praktikums umgesetzten Erweiterungen an der *Transcription Toolbox* beschrieben, im Kapitel 4.3 wird detaillierter auf ihre Implementierung eingegangen.

## **4.2 Das Expert-Mode-Fenster**

Um ein Nutzer-Feedback zu ermöglichen, wurde ein Layout für ein weiteres Fenster innerhalb des Programms entwickelt – den sogenannten *Expert-Mode*, also einem Experten-Modus zur Nachbearbeitung der Ergebnisse. Dieses lässt sich im Anschluss an die einzelnen Transkriptionen öffnen und erfüllt im Wesentlichen 2 Funktionen:

1. Darstellung von Informationen wie Tempo, Taktart und Tonart
2. Anbieten von verschiedenen Optionen zur Ausgabe und Nachbearbeitung der Transkriptionsergebnisse

Das *Expert-Mode* - Fenster ist in Abbildung 6 zu sehen – es ist in 5 Abschnitte untergliedert.



Alle Informationen, die sich aus der initialen Analyse des Musikstückes (siehe Kapitel 3.1) ergeben, werden im Bereich PRELIMINARY ANALYSIS dargestellt. Dazu gehört die Angabe des Tempos (TEMPO) in *beats per minute* (Schläge pro Minute), der Taktart (BAR MEASURE) und der Tonart (ROOT KEY).

Der Benutzer kann über die Checkbox METRONOME SWITCH zwischen 2 verschiedenen Metronomeinstellungen wählen. ARTIFICIAL bietet ein Metronom mit konstantem Tempo – also mit zueinander äquidistanten Zeitpunkten. Über die Auswahl REAL werden die im Verlauf der initialen Analyse detektierten Zeitpunkte der Taktzählzeiten innerhalb des Audiostückes als Referenz für das Metronom genutzt. Für Musikstücke, deren Tempi im Verlauf des analysierten Teils variiert, liefert dieses Metronom natürlich bessere Ergebnisse, da es flexibel auf nichtäquidistante Zeitpunkte der Zählzeiten reagieren kann. Beide Metronomeinstellungen können unabhängig voneinander konfiguriert werden.

Die Position der Hauptzählzeit – dem sogenannten *Down-Beat* – kann durch Veränderung der Taktart (BAR MEASURE) und eines entsprechenden Offsets (OFFSET : DOWN-BEAT) beeinflusst werden. Der Regler OFFSET : FINE-TUNING ermöglicht eine feinauflösende Positionskorrektur.

Über die beiden Optionen ENABLE METRONOME DURING PLAYBACK und DISPLAY OVERLAYED BEAT GRID kann der Nutzer das Abspielen des Metronoms während der Wiedergabe des Musikstückes im Hauptfenster bzw. die visuelle Darstellung des Metronoms unterhalb der Audiowellenform im oberen Frame des Hauptfensters (siehe Abbildung 1) aktivieren und deaktivieren.

Während der initialen Analyse werden durch eines der XProEx-Module aus dem Audiostück sogenannte *Chromavektoren* extrahiert, die Aussagen über die Tonhöhenverteilung und damit auch über die Häufigkeit bestimmter Töne im Verlauf des Stückes ermöglichen. Nach einer anschließenden Korrelation mit sogenannten *Tonartprofilen* kann allen möglichen Tonarten eine konkrete Auftretenswahrscheinlichkeit zugeordnet werden. Die 6 Wahrscheinlichsten werden dem Nutzer im Menüpunkt KEY SUGGESTIONS als Tonarthypothesen angeboten.

Jede Tonart beschreibt eine Menge von 7 der 12 chromatischen Töne in der Musik. Daraus entstand die Idee, dem Nutzer für die Bass-, Melodie- und Harmonietranskription eine zusätzliche Visualisierung in Bezug auf die Tonartbezogenheit der erkannten Noten des jeweiligen Musikinstrumentes anzubieten. Im Hauptfenster werden zu diesem Zweck alle dargestellten Tonhöhen, die einem Ton innerhalb der aktuellen Tonart zugeordnet werden können, mit einer grünen Schattierung hinterlegt (siehe Abbildung 2).

Im Abschnitt DRUM TRANSCRIPTION werden alle erkannten Schlagzeuginstrumente aufgelistet. Wird ein Instrument falsch erkannt, so kann es der Nutzer durch ein anderes ersetzen oder die jeweilige Schlagzeugspur per MUTE stummschalten. In Zukunft soll es ermöglicht werden, separat für jedes erkanntes Schlagzeuginstrument eine Erkennungsempfindlichkeit (DETECTION SENSITIVITY) nachzustimmen, falls einige Noten der entsprechenden Spur nicht richtig oder gar nicht erkannt wurden. Diese soll bei einem erneuten Ausführen der Transkription direkt als Parameter berücksichtigt werden.

Für die 4 Transkriptionstypen werden in den Abschnitten DRUM-, BASS-, MELODY- und HARMONY TRANSCRIPTION Optionen zur nachträglichen Korrektur der Transkriptionsergebnisse angeboten. Werden einzelne Noten zeitlich versetzt erkannt, so kann der Nutzer über Aktivierung der Option SNAP NOTES TO RHYTHMICAL GRID und der konkreten Auswahl eines Rasters eine rhythmische Quantisierung der Noten initiieren. Das Raster bestimmt hierbei, ob die Noten in einer Achtelnoten- oder beispielsweise in einer Sechzehntelnotenauflösung in Bezug auf das detektierte Tempo quantisiert werden.

Für die Bass-, Melodie- und Harmonietranskription wurde außerdem die Möglichkeit umgesetzt, durch die Transkriptionsalgorithmen detektierte Töne, die nicht in der Haupttonart liegen, entsprechend zu korrigieren. Dies geschieht durch Aktivieren der Option ONLY NOTES IN ROOT KEY. Hierbei wird vereinfachend angenommen, dass sich innerhalb des untersuchten Bereiches des Musikstückes die Tonart nicht ändert.

### 4.3 Implementierung

C++ ist eine sogenannte *höhere Programmiersprache*, die in den 1980er Jahren als Erweiterung der maschinennahen Sprache *C* entwickelt wurde und u. a. das Konzept der *objektorientierten Programmierung* unterstützt. Diese Art der Programmierung ermöglicht es, einzelne Teilaufgaben einer Software in Form von Daten und Funktionen innerhalb einzelner Klassen zu *kapseln* und zu *modularisieren*. Zugriff und Kommunikation zwischen verschiedenen Klassen erfolgt nur noch über vom Programmierer definierte Schnittstellen. All diese Ansätze ermöglichen ein hohes Maß an Wiederverwendbarkeit von bereits implementierten Softwarekomponenten und werden deshalb im IDMT angewandt.

Die Softwarebibliothek *QT* ermöglicht es, plattformunabhängige oberflächenbasierte Systeme zu erstellen. Sie bietet eine Menge an grafischen Grundelementen – sogenannte *Widgets* – wie Textfelder, Schaltflächen (*Buttons*), Schieberegler (*Sliders*), Statusanzeigen (*Progress bars*) und viele andere. Über ein *Layout* werden diese Widgets entsprechend der benötigten Struktur angeordnet. Eine bestimmte Aktion (*Event*) kann mit dem Auslösen einer bestimmten Funktion verbunden werden. Dies geschieht in QT mit Hilfe des *Signals & Slots* - Prinzip.

Jedem Widget sind verschiedene Ereignisse zugeordnet, wie z.B. das Klicken mit der linken Maustaste auf einen Button. Wird ein solches Ereignis ausgelöst, so sendet die entsprechende Klasse ein Signal an eine höherliegende Klasse, die anschließend dieses Ereignis auswertet, z.B. mittels der Veränderung des Wertes einer Variable oder einer bestimmten grafischen Ausgabe.

Die beiden wichtigsten Programmkomponenten, die im Laufe meines Praktikums implementiert wurden, sind die Klassen `CMAINRESULTS` und `EXPERTMODEVISUALIZATIONGUI` (siehe Abbildung 5).

Alle grafischen Elemente des Expert Mode Fensters wurden in der Klasse `EXPERTMODEVISUALIZATIONGUI` umgesetzt. Diese beinhaltet die folgenden



Komponenten:

- Definitionen der einzelnen *QT-Widgets*
- Strukturierte Anordnung der Widgets in definierten Layouts (z.B. dem *Grid Layout*)
- Verbinden des entsprechenden *Signals* mit den zugehörigen *Slots*
- Implementierung verschiedener Funktionen und Slots mit folgenden Funktionen
  - Aktualisierung der Ausgabe konkreter Werte wie Tempo, Taktart oder Tonart innerhalb des Expert Mode Fensters
  - Weiterleitung entsprechender Signale ausgehend von den einzelnen Widgets

Die Klasse CMAINRESULTS dient der Verwaltung aller programminternen Parameter. Sie dient gleichzeitig als Schnittstelle, über die alle Klassen des Transcription Toolbox Systems auf diese Parameter zugreifen, sie abfragen und verändern können. Sie beinhaltet z.B.

- Globale musikalische Parameter
  - Metronomtyp, Tempo, Tempo Strech Faktor, Down-Beat Offset, Fine-tuning Offset, Taktart
  - Tonart
- Detektierte Schlagzeuginstrumente
- Alle auswählbaren Optionen im *Expert-Mode*-Fenster
  - Akustische Wiedergabe bzw. visuelle Darstellung des Metronoms
  - Typ des Metronoms

- Zeitliche Notenquantisierung innerhalb der 4 Transkriptionsalgorithmen
- Notenkorrektur entsprechend der Haupttonart

Außer den beiden genannten Klassen waren viele Erweiterungen und Anpassungen bestehender Klassen notwendig, um die neuen Programmmerkmale in das Transbox-System zu integrieren.

## 5 Fazit des Praktikums

### 5.1 Programmieren

Im Laufe meines Praktikums konnte ich anhand einer konkreten Aufgabe den typischen Ablauf einer Softwareentwicklung kennenlernen. Hierbei konnte ich meine Fähigkeiten in der Programmierung sowohl in *C++* als auch in *QT* erweitern. Außerdem erhielt ich einen Einblick in den Aufbau größerer objektorientiert programmierter Softwaresysteme.

Bei der Programmierung des *Expert-Mode*-Fensters wurden knapp 40 verschiedene QT-Widgets umgesetzt, daher war eine genaue Planung der programminternen Weiterleitung einzelner Ereignisse unabdingbar. Dabei versuchte ich, den Quelltext sowohl konsistent als auch übersichtlich zu gestalten.

Vor allem in der Endphase meines Praktikums konnte ich mein Wissen über das *debugging*, also die programmgestützte Fehlersuche innerhalb eines Programms erweitern.

### 5.2 Ausblick

Als Praktikant fühlte ich mich sehr gut in den täglichen Ablauf integriert, erhielt bei allen Mitarbeitern stets eine Antwort auf offene Fragen und konnte viel über den typischen Arbeitsablauf am IDMT kennenlernen. Im Laufe meines Praktikums festigte sich mein Plan, im direkten Anschluss daran auch meine Diplomarbeit im

Institut Metadaten des IDMT zu schreiben. Speziell im Bereich der automatisierten Analyse von Audiodaten gibt es in der Abteilung Metadaten viele Forschungsfelder, die mich sehr interessieren.

Während meines Praktikums konnte ich bereits einige wissenschaftliche Veröffentlichungen (vor allem Papers, Dissertationen und Diplomarbeiten) in diesem Fachgebiet studieren, um mir so einen Überblick zu verschaffen, welchen aktuellen Wissensstand es in einzelnen Forschungsfeldern gibt.

Sowohl auf diesem Weg, als auch über das direkte Gespräch mit Angestellten des Institutes, Diplomanden und Doktoranden konnte ich viel über den Ablauf und die Anforderungen, die an eine wissenschaftliche Veröffentlichungen gestellt werden, lernen. Ich bin sicher, dass dieses Wissen für meine weitere berufliche Zukunft sehr hilfreich sein wird.

Auch die Einarbeitung in verwaltungstechnische Aspekte, die im Fraunhofer Institut zu berücksichtigen sind, sowie das Kennenlernen vieler Mitarbeiter aus verschiedenen Arbeitsbereichen wird es mir einfacher machen, hier meine Diplomarbeit zu schreiben.

## **A   Arbeitsgrundlage - Liste an geplanten neuen *Features* für die *Transcription Toolbox***

Einleitende rhythmische Analyse während des Öffnens der Audiodatei

- Schätzung einer rhythmischen Grundfigur (*root-pattern*), des Taktes, des Tempos und der Grunds Schlagzeitpunkte
- Ermöglichen der akustischen Wiedergabe und der visuellen Darstellung des Metronoms (*beat grid*)
- Ermöglichen der Korrektur des Metronoms durch den Nutzer über eine Änderung des Tempos oder Verschiebung der Grunds Schlagzeitpunkte

Schlagzeugtranskription:

- Der Nutzer kann die von ihm erwarteten Schlagzeuginstrumente vor der eigentlichen Transkription auswählen
- Darstellung der erkannten Schlagzeuginstrumente
- Im Falle einer unvollständigen oder fehlerhaften Erkennung einzelner Schlagzeuginstrumenten kann der Nutzer über eine Erhöhung der jeweiligen Erkennungsempfindlichkeit das Klassifikationsergebnis korrigieren.

#### Basstranskription:

- Verbesserung der Bassnotenerkennung in bestimmten mehrdeutigen Fällen basierend auf einer automatischen Tonarterkennung oder einer Eingabe der Tonart durch den Nutzer
- Extraktion der Bassnoten der jeweiligen Akkorde neben der eigentlichen Basslinie

#### Melodietranskription:

- Verbesserung der Melodietranskription in bestimmten mehrdeutigen Fällen basierend auf einer automatischen Tonarterkennung oder einer Eingabe der Tonart durch den Nutzer
- Darstellung von Notenkandidaten zur manuellen Auswahl durch den Nutzer

#### Harmonietranskription:

- Automatische Schätzung der Tonart mit manueller Korrekturmöglichkeit durch den Nutzer
- Darstellung von Akkordkandidaten zur manuellen Korrektur durch den Nutzer

## **B Abkürzungsverzeichnis**

Fhg	Fraunhofer Gesellschaft
IDMT	Fraunhofer Institut für digitale Medientechnologie (Ilmenau)
IIS	Fraunhofer Institut für integrierte Schaltungen (Erlangen)
MATLAB	Matrix Laboratory
MIDI	Musical Instrument Digital Interface
MP3	MPEG-1 Audio Layer 3
QbH	Query by Humming
QT	Quasar toolkit
WAV	Wave Format
XProEx	Extensible Processor Extractor

## Literatur

- [1] Fraunhofer Gesellschaft. <http://www.fraunhofer.de/ueberuns/index.jsp>.
- [2] Fraunhofer Institut für Digitale Medientechnologie.  
<http://www.fraunhofer.de/ueberuns/index.jsp>.
- [3] Trolltech - QT - Online Dokumentation. <http://doc.trolltech.com/>.
- [4] C. Dittmar and J. Abeßer. Transcription toolbox demo documentation. 2007.
- [5] C. Dittmar, K. Dressler, and K. Rosenbauer. A toolbox for automatic transcription of polyphonic music. Audio Mostly, 2007.