Stefan Huber

Untersuchung verschiedener Verfahren zur Grundfrequenzbestimmung mit Einstellung einer Applikation zur Midi-Konvertierung

Diplomarbeit

Diplom.de

Bibliografische Information der Deutschen Nationalbibliothek:

Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de/ abrufbar.

ISBN: 9783832476984

Stefan Huber

Untersuchung verschiedener Verfahren zur Grundfrequenzbestimmung mit Einstellung einer Applikation zur Midi-Konvertierung

Diplom.de

Stefan Huber

Untersuchung verschiedener Verfahren zur Grundfrequenzbestimmung mit Einstellung einer Applikation zur Midi-Konvertierung

Diplomarbeit
Fachhochschule für Technik und Wirtschaft Berlin
Fachbereich Ingenieurwissenschaften I
Abgabe Juni 2003

Diplom.de

Diplomica GmbH
Hermannstal 119k
22119 Hamburg

Fon: 040 / 655 99 20
Fax: 040 / 655 99 222

agentur@diplom.de
www.diplom.de

ID 7698

ID 7698
Huber, Stefan: Untersuchung verschiedener Verfahren zur Grundfrequenzbestimmung mit Einstellung einer Applikation zur Midi-Konvertierung
Hamburg: Diplomica GmbH, 2004
Zugl.: Fachhochschule für Technik und Wirtschaft Berlin, Fachhochschule für Wirtschaft und Technik, Diplomarbeit, 2003

Diplomica GmbH
http://www.diplom.de, Hamburg 2004
Printed in Germany

Inhaltsverzeichnis

Abbildungsverzeichnis

1. Einleitung

In den letzten Jahrzehnten wurden viele Verfahren zur Grundfrequenzbestimmung erfunden. Es wurden verschiedene Versuche unternommen, die Grundfrequenz eines periodischen bzw. pseudoperiodischen akustischen Signals zu bestimmen. Dabei kommen Algorithmen im Zeitbereich sowie im Frequenzbereich zur Transformation zum Einsatz.

Die physikalischen Ansätze der Algorithmen weisen Stärken aber auch Schwächen auf. Keiner davon ist in der Lage, hunderprozentig akkurat und zuverlässig zu arbeiten. Bis heute existiert keine Formel, kein Modell und kein universeller Algorithmus, der die wahrgenomme Tonhöhe eines komplexen Tonals genau und fehlerlos bestimmen kann.

In dieser Arbeit werden verschiedene Verfahren zur Grundfrequenzbestimmung untersucht und versucht, die gewünschten Eigenschaften aller unterschiedlichen Algorithmen zu kombinieren, um ein bestmögliches Resultat bei einer Wandlung von analogen bzw. digitalisierten Audio-Signalen in das komprimierende Audio-Format Midi zu erzielen.

Die Idee dazu enstandt aus der alltäglichen Praxis heraus. Beim Komponieren von Musik hat man schnell einen einfachen Ryhtmus aus Schlagzeug und Bass zusammengestellt, die Komposition komplexerer, bereits im Kopf vorgedachter Melodien ist jedoch schwierig und zeitaufwendig.

Bei Melodieverläufen, die aus mehr als fünf Noten bestehen, kann man mittels einer Gitarre oder einem Klavier meist schnell die ersten drei bis fünf Noten bestimmen. Bei höherer Notenanzahl hören sich die ersten Noten oft nach dem gewünschten Melodieverlauf an, weitere Noten stehen aber oftmals in Dissonanz zur anfänglichen Notensequenz. Der bereits erstellte Melodieverlauf aus den ersten Noten muss wieder verworfen werden, um die im Gehirn vorhandene Melodie komplett im Einklang aller Noten zueinander komponieren zu können.

So erhöht sich der Zeitaufwand beim musikalischen Komponieren exorbitant, die anfängliche Euphorie durch das im Kopf gespeicherte Musikstück verliert sich mit der Zeit im Komponieren der gewünschten Melodie. Da es dem Menschen um ein Vielfaches leichter fällt, die Melodie im Kopf zu singen, zu pfeifen oder zu summen, soll die Applikation zur Midi-Konvertierung dem Musiker helfen, schnell und einfach ohne unnötige Hindernisse sein Ziel zu erreichen.

Der Fokus der Diplomarbeit liegt bei der Verarbeitung monophoner Audiosignale, da sich die Erkennung von mehreren Melodieverläufen, zusätzlichen Ryhtmusstrukturen und anderen polyphonen Sounds weitaus schwieriger gestaltet. Die erstellte Anwendung Pitch2Midi ist auf menschlichen Gesang optimiert, kann aber auch zur Konvertierung von Instrumentenklängen angewandt werden.

1.1 Schall

Schall besteht aus Schwingungen, die sich in elastischen Medien wie Luft, Wasser oder Metall als Longitudinalwellen fortpflanzen und die die Moleküle der Medien beim Auftreten der Welle zusammenpressen. Je höher der Schalldruck, desto stärker werden die Moleküle zusammengedrückt.

Vibrieren Objekte, Gegenstände und Oberflächen, werden die Luftmoleküle in der Umgebung entsprechend der Stärke und der Frequenz der Vibration gepresst und wieder entlastet, wodurch Schallwellen in der Luft erzeugt werden. Die Luft wird durch die Schwingungen einer Gitarrenseite, von Blasinstrumenten oder den Oberflächen von Perkussionsinstrumenten angeregt. Blasinstrumente nutzen den Luftstrom der menschlichen Lunge und erzeugen durch Brechung, Reflexion, Beugung und Interferenz der Schallwellen ihr charakteristisches Klangbild.

Der Schalldruck wird in Dezibel gemessen und gibt den Druckunterschied zwischen zwei Molekülzuständen an. Die Gesamtleistung eines Schalls ist das Integral der Intensität über eine Oberfläche.

$$\text{Schall-Lautstärke L in dB} = 10 \cdot \log 10\left(\frac{l}{l_{REFERENZ}}\right)$$

$$\text{Schall-Leistung P in dB} = 20 \cdot \log 10\left(\frac{P}{P_{REFERENZ}}\right)$$

Die Referenz-Leistung gibt den atmosphärischen Schalldruck an, der eine Druckreferenz auf die gemessene Schall-Leistung gibt und bei $P_{REFERENZ}$ = 0,00002 Pa liegt. Ein Schallunterschied von 20 dB bedeutet zehnfachen Schalldruck, 6 dB entsprechen zweifachem Schalldruck und bei – 6 dB wird der Schalldruck halbiert. Typische Leistungen sind 10 µW für Sprache, 1 mW für eine Geige und 100 W für einen Lautsprecher.

1.2 Erzeugen von Sprache und Gesang

Die menschliche Sprachproduktion kann in zwei Teile unterschieden werden. Der Kehlkopf mit seinen Stimmbändern erzeugt aus dem Luftdruck der Lunge ein Signal, das vom nachfolgenden Vokaltrakt, also der Mund- und Nasenraum sowie die Abstrahlung über Zunge, Lippen und Mund, gefiltert wird.

Sprache und Gesang ist mathematisch die Faltung eines Anregungsimpulses mit den im Vokaltrakt gegebenen Filtern. Durch Luftdruck aus der Lunge werden diese akustischen Filter im Mund- und Nasenraum angeregt. Die Stimmbänder des Kehlkopfs werden durch den Luftstrom der Lunge in Schwingung versetzt und verwandeln diesen in ein regelmäßiges, periodisches Anregungssignal, das im nachfolgenden Vokaltrakt durch Resonanzen verändert wird, um die charakterlichen Merkmale eines Tones zu erzeugen.

Der Vokaltrakt des Menschen stellt in Äquivalenz zu einem Musikinstrument den Hohlraum zur Klangspektrenbildung durch Resonatoren dar. Durch Veränderung von Form und Volumen des Hohlraumes werden die Resonanzeigenschaften des Vokaltraktes gesteuert, sodass Frequenzbereiche abgesenkt oder angehoben werden.

Die gepresste Luft der Lunge führt zu einem Druckanstieg, der die durch die Muskelanspannung der Stimmbänder geschlossene Stimmritze öffnet. Die Luftentweichung vermindert den Druck und lässt die Stimmritze wieder schliessen. Dies führt zu quasiperiodischen Luftstromimpulsen im nachfolgenden Vokaltrakt in Abhängigkeit des Luftstromdruckes und Spannung und Lage der Stimmbänder. Diese Impulse bestimmen die einhüllende Kurve einer Tonhöhenperiode und verursachen die Grundfrequenz.

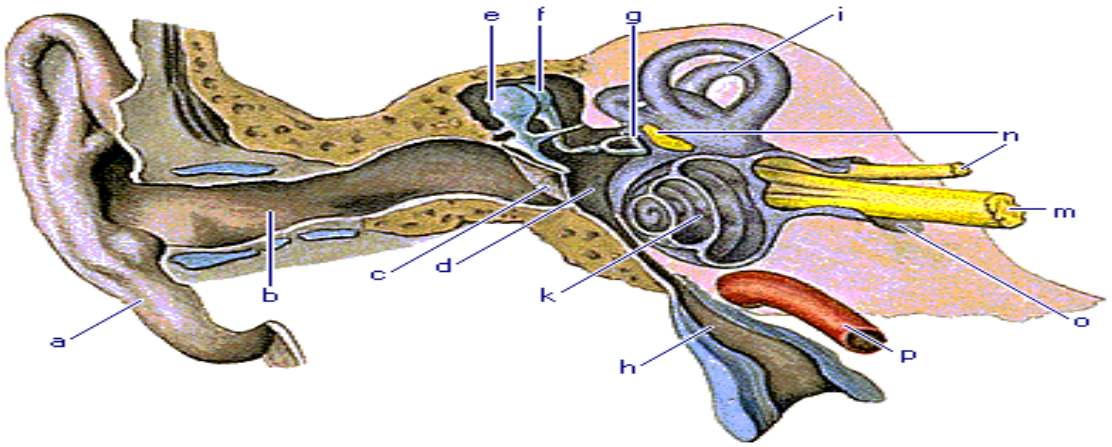
Das Ergebnis aus Luftstrom und Filterung durch den Vokaltrakt ist ein nichtstationärer Prozess, da beide Parameter jederzeit ihr Verhalten ändern können. Dies führt zu den starken Veränderungen der zeitlichen Struktur eines Sprachsignals. Nur für sehr kurze Zeitintervalle kann ein quasistationärer Zustand angenommen werden.

Zusammenfassend können die Stimmbänder des Kehlkopfes und der Luftstrom der Lunge als Impulsgenerator angesehen werden. Die Tonhöhe wird durch unterschiedliches Spannen der Stimmbänder im Kehlkopf bestimmt. Die Impulse werden von einer Anzahl an Filtern im Vokaltrakt bearbeitet. Die Ausgabe geschieht über Schallabstrahlung durch Mund und Nase.

1.3 Wahrnehmung durch Hören

Um im Gehirn gespeicherte Informationen mittels Sprache oder Gesang fehlerlos ausgeben zu können, bedarf es einer Rückkopplung der Ausgabe zum Gehirn, um dem kontrollierenden Organ mitzuteilen, ob die zuletzt ausgeführten Befehle entsprechend genau gesungen oder gesprochen wurden. Dies geschieht mittels dem Hörorgan Ohr, welches Schallwellen analysiert und dem Gehirn die ermittelten Ergebnisse mitteilt.

Das menschliche Ohr wird in drei gröbere Bereiche unterteilt, Aussenohr, Mittelohr und Innenohr. Den genaueren Aufbau verdeutlicht Abbildung 1.2.



Ohr: Schematische Darstellung des menschlichen Ohrs: a Ohrmuschel; b äußerer Gehörgang; c Trommelfell; d Paukenhöhle; e Hammer; f Amboß; g Steigbügel; h Ohrtrompete; i Bogengang; k Schnecke; m Gehör- und Gleichgewichtsnerv; n Gesichtsnerv; o innerer Gehörgang; p innere Kopfschlagader

Abbildung 1.2 Menschliches Ohr

Das äussere Ohr umfasst die Ohrmuschel a, den äußeren Gehörgang b und das Trommelfell c. Das Mittelohr, die Paukenhöhle d, besteht aus den drei Knochen Hammer e, Amboss f und Steigbügel g, und verbindet das Trommelfell mit dem Innenohr. Dieses setzt sich aus der Schnecke k und dem Gehörnerv m zusammen.