

**Stefan Huber**

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

**Diplomarbeit**

## **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.

Dieses Werk sowie alle darin enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsschutz zugelassen ist, bedarf der vorherigen Zustimmung des Verlanges. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen, Auswertungen durch Datenbanken und für die Einspeicherung und Verarbeitung in elektronische Systeme. Alle Rechte, auch die des auszugsweisen Nachdrucks, der fotomechanischen Wiedergabe (einschließlich Mikrokopie) sowie der Auswertung durch Datenbanken oder ähnliche Einrichtungen, vorbehalten.

Copyright © 2003 Diplomica Verlag GmbH  
ISBN: 9783832476984

**Stefan Huber**

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

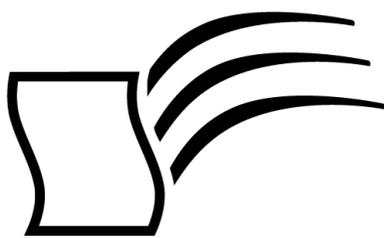


---

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

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

---

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtes.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Die Informationen in diesem Werk wurden mit Sorgfalt erarbeitet. Dennoch können Fehler nicht vollständig ausgeschlossen werden, und die Diplomarbeiten Agentur, die Autoren oder Übersetzer übernehmen keine juristische Verantwortung oder irgendeine Haftung für evtl. verbliebene fehlerhafte Angaben und deren Folgen.

Diplomica GmbH

<http://www.diplom.de>, Hamburg 2004

Printed in Germany

# Inhaltsverzeichnis

1.	Einleitung	6
1.1	Schall	7
1.2	Erzeugen von Sprache und Gesang	7
1.3	Wahrnehmung durch Hören	8
1.4	Aufbau von Tönen und Klängen	9
1.5	Klangeigenschaften	11
1.6	Auftretende Probleme bei der Grundfrequenzerkennung	11
2.	Technische Grundlagen	14
2.1	Digitales Audio	14
2.2	Filter	15
2.3	Das Dateiformat Wave	18
2.3.1	Das Dateiformat RIFF	18
2.3.2	Chunk Architektur	18
2.3.2.1	Wave- oder Header-Chunk	19
2.3.2.2	Format-Chunk	19
2.3.2.3	Daten-Chunk	20
2.3.2.4	Waveheader	20
2.3.2.5	Weitere Chunkarten	21
2.3.3	Speicheranordnung	22
2.4	Der Standard MIDI	22
2.4.1	Vergleich mit Waveformat	22
2.4.2	MIDI Hardware	23
2.4.3	MIDI Systeme	24
2.4.4	MIDI Sequenzer	25
2.4.5	MIDI Noten	26
2.4.6	MIDI Dateien	27
2.4.6.1	Header-Chunk	28
2.4.6.2	Zeitformat	28
2.4.6.3	Track-Chunk	28
2.4.7	MIDI Events	29
2.4.7.1	Status- und Datenbytes	29
2.4.7.2	Befehlskategorien	30
2.4.7.3	NoteOn Befehl	30
2.4.7.4	NoteOff Befehl	31
2.4.7.5	Programm Change Befehl	31
2.4.7.6	End of Track Befehl	33
2.4.7.7	MIDI Text und Copyright	33
2.4.7.8	Noten- und Pausenlängen	33
2.4.8	Delta-Time Kommando	33
2.4.9	MIDI Tempo	34
2.4.10	General MIDI	35
2.4.11	Little und Big Endian Datenformate	36
2.5	Faltung	37
2.6	Korrelation	38
2.7	Fourier-Transformation	39
2.7.1	Fourier-Integral	39
2.7.1.1	Definition	39
2.7.1.2	Veranschaulichung	40
2.7.1.3	Inverse Fourier-Transformation	42

2.7.1.4	Parseval'sches Theorem.....	42
2.7.2	Fourier-Reihe.....	43
2.7.3	Diskrete Fourier-Transformation, DFT.....	43
2.7.3.1	Definition .....	43
2.7.3.2	Komplexe Koeffizienten.....	44
2.7.3.3	Korrelation .....	45
2.7.3.4	Symmetrie .....	46
2.7.3.5	Zeit- und Frequenzauflösung.....	46
2.7.4	Inverse DFT.....	47
2.7.5	Fast Fourier-Transformation, FFT.....	47
2.7.5.1	Schnelle Berechnung .....	47
2.7.5.2	Zerlegung der Transformationslänge .....	48
2.7.5.3	Architekturen der FFT.....	49
2.7.5.4	Butterfly .....	50
2.8	Analyse .....	51
2.9	Energieberechnung .....	51
2.10	Fensterung aufgrund Spektralverbreiterung .....	52
3.	Verfahren zur Grundfrequenzbestimmung .....	54
3.1	Zeitbereich .....	54
3.1.1	Nulldurchgangsrate.....	54
3.1.2	Spektraltransformation mittels Bandpässen.....	56
3.1.2.1	Definition .....	56
3.1.2.2	Parallelschaltung .....	56
3.1.2.3	Umsetzung .....	56
3.1.2.4	Nachteile .....	57
3.1.2.5	Maximumerkennung .....	57
3.1.3	Autokorrelation.....	59
3.1.3.1	Einführung .....	59
3.1.3.2	Formel.....	59
3.1.3.3	Berechnung .....	59
3.1.3.4	Bestimmung der Grundperiode .....	60
3.1.4	Betragsdifferenzfunktion AMDF .....	62
3.2	Zeit- und Frequenzbereich kombiniert.....	63
3.2.1	Autokorrelation im Zeit- und Frequenzbereich.....	63
3.2.1.1	Vorteile .....	64
3.2.1.2	Nachteile .....	64
3.2.1.3	Autokorrelation mit Center-Clipping .....	65
3.2.2	Cepstrum.....	66
3.3	Frequenzbereich.....	68
3.3.1	Spektrale Kompression.....	68
3.3.2	Spektrale Autokorrelation .....	71
4.	Pitch2Midi Konverter .....	73
4.1	Philosophie .....	73
4.2	Dialogumgebung und Klassenhierarchie .....	73
4.3	Benutzerführung .....	74
4.4	Definitionen.h.....	77
4.5	Die Klasse Wave .....	78
4.5.1	Headerdatei.....	78
4.5.2	Funktionen.....	79
4.5.2.1	Wavedatei öffnen und lesen.....	80
4.5.2.2	Wavedatei normalisieren .....	81

4.5.2.3	Dateipfad speichern.....	82
4.5.2.4	Wavedatei erstellen.....	82
4.5.2.5	Wavedatei prüfen.....	83
4.5.2.6	FOURCC Feld kopieren.....	83
4.5.2.7	Tiefpass-Filterung.....	83
4.5.2.8	Fensterfunktionen.....	86
4.6	Die Klasse Midi.....	87
4.6.1	Headerdatei Midi.h.....	88
4.6.2	MIDI Notenberechnung.....	88
4.6.3	Initialisierung im Konstruktor.....	89
4.6.4	Funktionen.....	90
4.6.4.1	Word Bit-Reverse.....	90
4.6.4.2	DoubleWord Bit-Reverse.....	90
4.6.4.3	Deltatime schreiben.....	91
4.6.4.4	Zeit- und Frequenzauflösung setzen.....	92
4.6.4.5	MIDI-Datei schreiben.....	92
4.7	Die Klasse FourTrans.....	93
4.7.1	Komplexe Arithmetik.....	93
4.7.2	Koeffizienten in vorberechneten Arrays.....	94
4.7.3	Die Funktion Bit-Reverse.....	95
4.7.4	Die Funktionen FFT und InverseFFT.....	95
4.8	Spektrale Gewichtung.....	96
4.8.1	Erklärung.....	96
4.8.2	Implementation.....	97
4.9	Klassen zur Grundfrequenzbestimmung.....	98
4.9.1	MIDI Konvertierung.....	98
4.9.2	Die Klasse AMDF.....	101
4.9.3	Die Klasse Autokorrelation.....	102
4.9.3.1	MIDI Konvertierung.....	102
4.9.3.2	PufferProzess der schnellen AKF.....	102
4.9.3.3	PufferProzesse der geklippten AKF.....	103
4.9.3.4	PufferProzess der AKF im Zeitbereich.....	104
4.9.3.5	PufferProzess der AKF im Frequenzbereich.....	104
4.9.4	Die Klasse Cepstrum.....	105
4.9.5	Die Klasse ErstesMaxima.....	105
4.9.6	Die Klasse FrequenzAbstand.....	106
4.9.7	Die Klasse SpektraleKompression.....	107
4.9.8	Die Klasse SpektralerIntervall.....	107
4.10	Die Pitch2Midi Dialog-Klasse.....	107
4.10.1	OnKonvertieren().....	107
4.10.2	Kombinierer.....	110
4.11	MFC-spezifische Hinweise.....	111
4.11.1	Schrift und Hintergrund in Dialogen färben.....	111
4.11.2	Mehrere Registerkarten.....	111
4.11.3	Bitmap Buttons.....	114
4.11.4	Textausgabe.....	115
4.11.5	CFileDialog.....	115
4.11.5.1	Wave öffnen.....	115
4.11.5.2	Speichern unter.....	116
4.11.6	Globaler Zugriff.....	116
4.11.7	Normalisierungs-Dialog.....	117
5.	Auswertung.....	117

6. ....	Zusammenfassung.....	120
7. ....	Literaturverzeichnis .....	121

## Abbildungsverzeichnis

Abbildung 1.2	Menschliches Ohr .....	8
Abbildung 1.3	Komplexe Schwingung .....	10
Abbildung 2.4.2	MIDI Verbindung.....	23
Abbildung 2.4.3 a)	Minimales MIDI System.....	24
Abbildung 2.4.3 b)	PC und MIDI.....	24
Abbildung 2.4.3 c)	Live MIDI System.....	25
Abbildung 2.7.1 b)	Eine Frequenzkomponente.....	40
Abbildung 2.7.1 d)	Drei Frequenzkomponenten.....	41
Abbildung 2.7.5.3. a)	Decimation in Time.....	49
Abbildung 2.7.5.3. b)	Decimation in Frequency.....	49
Abbildung 4.3 a)	Dominante Grundfrequenz.....	57
Abbildung 4.3 b)	Dominante Oberwellen.....	58
Abbildung 3.1.3.4. b)	Autokorrelations-Ergebnis .....	61
Abbildung 3.1.4	AMDF-Koeffizienten.....	62
Abbildung 3.2.1.3 a)	Abgetrenntes Audiosignal.....	65
Abbildung 3.2.1.3 b)	Geklipptes Autokorrelations-Ergebnis.....	65
Abbildung 3.2.2 b)	Cepstrumbereich, Wurzelbildung .....	68
Abbildung 3.3.1 a)	Fourierspektrum .....	69
Abbildung 3.3.1 b)	komprimiertes Spektrum.....	69
Abbildung 3.3.1 c)	verraushtes Fourierspektrum .....	70
Abbildung 3.3.1 d)	komprimiertes Spektrum verrauscht .....	70
Abbildung 3.3.2 a)	Fourierspektrum .....	72
Abbildung 3.3.2 b)	Autokorrelation im Frequenzbereich.....	72
Abbildung 4.4	Linienpektrum.....	96
Abbildung 4.9.1	PAP.....	99

# 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, hundertprozentig akkurat und zuverlässig zu arbeiten. Bis heute existiert keine Formel, kein Modell und kein universeller Algorithmus, der die wahrgenommene 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 entstand aus der alltäglichen Praxis heraus. Beim Komponieren von Musik hat man schnell einen einfachen Rythmus 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 Notenzahl 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 Rythmusstrukturen 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.

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

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

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

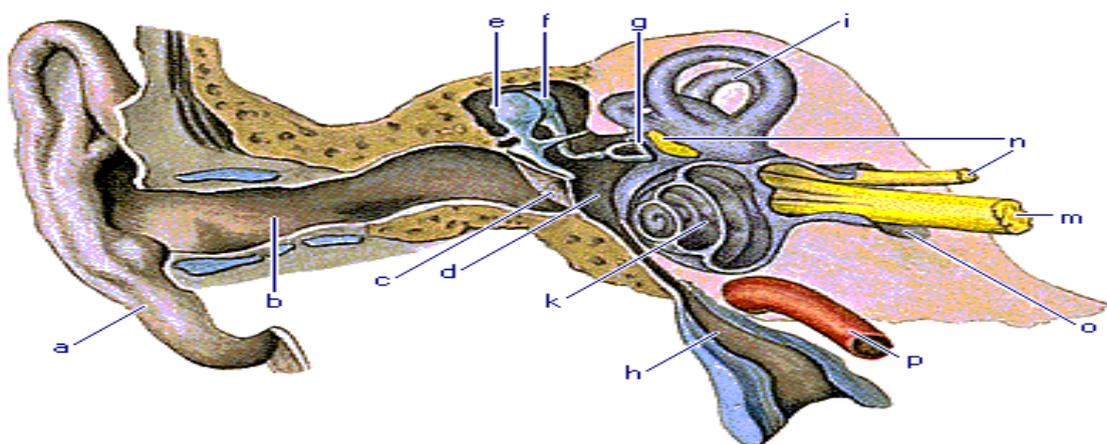
Das Ergebnis aus Luftstrom und Filterung durch den Vokaltrakt ist ein nichtstationärer Prozess, da beide Parameter jederzeit ihr Verhalten ändern können. Dies führt zu den starken Veränderungen der zeitlichen Struktur eines Sprachsignals. Nur für sehr kurze Zeitintervalle kann ein quasistationärer Zustand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