

# **Physik des menschlichen Stimmapparats-**

## **Computergestützte Experimente und Datenanalyse**

Schriftliche Hausarbeit im Rahmen der  
Ersten Staatsprüfung für das Lehramt für die  
Sekundarstufe I und II

dem Staatlichen Prüfungsamt für Erste Staatsprüfungen für Lehrämter an Schulen  
-Essen-

vorgelegt von:

**Andrea L. Moser**

Universität-GHS-Essen

Oktober 1996

Themensteller: Prof. Dr. H. J. Schlichting

Didaktik der Physik, Fachbereich 7

# Inhaltsverzeichnis

1 EINLEITUNG	3
2 FUNKTIONEN DES MENSCHLICHEN STIMMAPPARATS	7
2.1 Physikalische Voraussetzungen der Stimmerzeugung	7
2.2 Biophysikalische Voraussetzungen	15
3 MATHEMATISCHE ANALYSEVERFAHREN	24
3.1 Die Lorenzabbildung	25
3.2 Die Autokorrelationsfunktion	27
3.3 Die Fouriertransformation	32
4 VERWENDETE PROGRAMME	39
4.1 Aufnahmeprogramm	39
4.2 WAVcalc	40
4.3 Simple FFT	47
4.4 Spectrogram	51
4.5 Zusammenfassung	55
5 DURCHFÜHRUNG	57
5.1 Meßwertaufnahme	57
5.2 Meßabläufe	58
5.3 Prüfmessung	59
6 EXPERIMENTELLE ANALYSE DER GRUNDSCHALLFORMEN	62
6.1 Was ist eine Grundschaallform?	62
6.2 Die stumme Schallform	63
6.3 Der Explosionsschall	63
6.4 Das Friktionsrauschen	71
6.5 Der periodische Schall	77

---

7 UNTERSUCHUNGEN AM STIMMAPPARAT	99
7.1 Öffnen und Schließen des Mundes	99
7.2 Variation der Tonhöhe	105
7.3 Die Singstimme	109
7.4 Die Heliumstimme	113



---

8 SIMULATION DES STIMMAPPARATS	117
8.1 Aufbau des Experiments	117
8.2 Meßablauf	121
8.3 Meßergebnisse	123
9 ALTERNATIVE MEßMETHODE	127
10 ZUSAMMENFASSUNG UND AUSBLICK	132
11 ANHANG	135
11.1 Technische Hilfsmittel	135
11.2 Verwendete Programme	135
11.3 Nutzung der Disketten	136
11.4 Abbildungsverzeichnis	137
11.5 Literaturverzeichnis	141
11.6 Danksagung	143
11.7 Eidesstattliche Erklärung	143

## 1 Einleitung

Die menschliche Stimme war schon in der Antike ein beliebtes Forschungsgebiet für Gelehrte und andere, die sich mit der Redekunst auseinandersetzten. Cicero (106-43 v. Chr.), ein römischer Staatsmann und berühmter Rhetoriker, machte sich nicht nur Gedanken über den Inhalt der Reden, sondern legte viel Wert auf die Ausdrucksform und den Klang der gesprochenen Sprache. In seinem Werk „De Oratore“ beschrieb Cicero diesen Zusammenhang mit den folgenden Worten:

*„Wir müssen auch die Zunge, den Atem und sogar den Klang der Stimme regulieren ... Sie bieten sich dem Redner, wie dem Maler seine Farben, zur Abwechslung an.“ [1]*

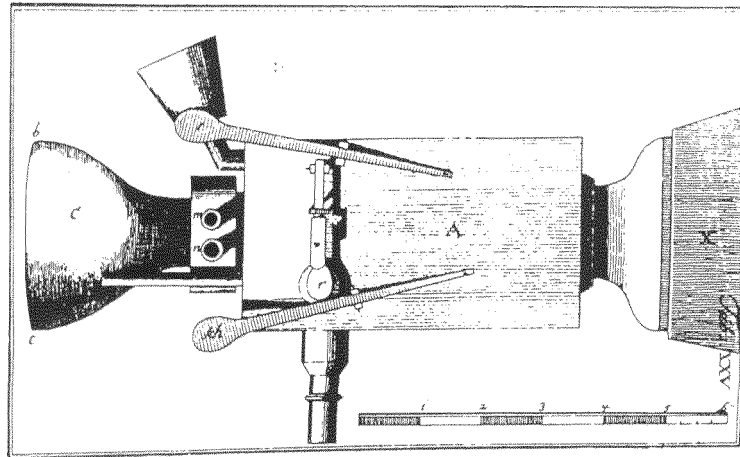
Cicero forderte die Schüler seiner Rednerschule auf, ihre Stimme wie Künstler zu benutzen. Die Stimme sollte so zu einem individuellen Ausdrucksmittel gestaltet werden.

Im Mittelalter stagnierte das Wissen um die menschliche Stimme und die gesprochene Sprache.

Erst in der Renaissance wurden die Forschungen auf dem Gebiet der Akustik und der Physiologie der Stimme wieder verstärkt. Besonders zu erwähnen ist Leonardo da Vinci (1452-1519), der eine Vielzahl von Zeichnungen des Kehlkopfes und die erste Darstellung des Ansatzrohrs schuf. Im Werk „De Corpore Humanis“ finden sich viele realistische Zeichnungen, die da Vinci durch viele Studien an seziierten Leichen erzielte. [2]

Durch das Aufkommen der exakten Naturwissenschaft in der Neuzeit wurde Ende des 18. Jahrhunderts zum ersten Mal eine genaue Beschreibung der Natur der Vokale erbracht. Mittels der akustischen Schwingungslehre gelang es Christian Gottlieb Kratzenstein, einem Schüler des Mathematikers Euler, diese Aufgabenstellung zu lösen, die als Preisfrage von Euler gestellt wurde. In seinem Buch über den „Mechanismus der menschlichen Stimme“ publizierte Johann Wolfgang von Kempelen (1734-1804) seine Erfahrungen, die er mit der

Bildung der Stimme und Sprache gemacht hatte. Er konstruierte eine Maschine, die ganze Äußerungen künstlich erzeugen konnte (Abbildung 1-1). [1,2]



**Abbildung 1-1** Teil einer Sprechmaschine von J. W. von Kempelen [1]

Hermann von Helmholtz veröffentlichte 1862 in seinem Buch über die „Lehre von den Tonempfindungen“ die akustischen Eigenschaften von Hohlkörpern. Zu dieser Zeit beschäftigten sich auch die englischen Physiker Willis und Wheatstone mit den akustischen Eigenschaften von Hohlkörpern. Mit diesen Erklärungsansätzen wurde versucht, die Entstehung der menschlichen Stimme physikalisch zu erläutern.

Die Untersuchung der menschlichen Stimme wurde in der zweiten Hälfte des 20. Jahrhunderts besonders durch Gunnar Fant geprägt. In seinen Büchern [3,4] beschreibt er die Entstehung der Stimme und der Formanten des russischen und schwedischen Vokalsystems. Weitere interessante Untersuchungen zur menschlichen Stimme sind im Buch von Flanagan [5] veröffentlicht.

Da dieser Themenkomplex in der Didaktik der Physik der Universität-GHS Essen noch nie erarbeitet wurde, war es eine Herausforderung, die Physik des menschlichen Stimmapparats zu untersuchen.

In dieser Arbeit soll nun ein Überblick über die Funktionsweisen des menschlichen Stimmapparats gegeben werden. Die Zusammenhänge werden anhand der verschiedenen Bereiche, wie zum Beispiel der Physik, Biologie, Deutsch und der Physiologie, dargelegt. Die Absicht dieser Arbeit ist es, die fächerübergreifende Thematik verständlich zu erörtern und anhand von zahlreichen Experimenten zu veranschaulichen. Das fächerübergreifende

Thema bietet mir eine Verbindung meiner beiden Unterrichtsfächer Physik und Deutsch. Außerdem ist im Rahmen dieser Arbeit ein Meßsystem für den Computer geschaffen worden. Mit diesem System können Experimente mit der Stimme von jedem durchgeführt und analysiert werden.

Die vorliegende Arbeit gliedert sich in zwei Abschnitte. Der erste Teil (Kapitel 2 bis 5) schafft die Voraussetzungen für dieses Thema, damit die Experimente, die den zweiten großen Teil der Arbeit ausmachen (Kapitel 6-9), deutbar werden. Die Abschnitte gliedern sich folgendermaßen:

Das zweite Kapitel beschäftigt sich mit den Funktionen des menschlichen Stimmapparats. Anhand von physikalischen und biophysikalischen Gesetzmäßigkeiten wird erklärt, wie die Stimme erzeugt wird. Im dritten Kapitel werden die mathematischen Grundlagen erläutert, die die Voraussetzungen für die in Kapitel 4 vorgestellten Programme sind. Die verwendeten Programme stellen ein komplettes Analysesystem dar, das auf den Stimmapparat zugeschnitten ist. Die Umsetzung der Mathematik in diese Programme wird durch kurze Listings dargelegt. Die Durchführung der Experimente wird in Kapitel 5 erläutert. Zuerst wird die Meßwertaufnahme beschrieben. Anschließend werden kurz die Meßabläufe aufgezeigt, um den Zusammenhang zwischen den einzelnen verwendeten Programmen zu verdeutlichen.

Nach diesem Kapitel beginnt der zweite große Teil dieser Arbeit. In Kapitel 6 werden die Schallformen, die der Mensch erzeugen kann, in unterschiedliche Kategorien aufgeteilt und auf Gemeinsamkeiten und Unterschiede untersucht. Ferner werden in Kapitel 7 die Auswirkungen analysiert, die sich ergeben, wenn gezielte Änderungen am Stimmapparat durchgeführt werden. Besonders interessant sind die Auswirkungen von Helium auf die Stimme und welche Unterschiede zwischen einer Sing- und einer Sprechstimme bestehen.

Im achten Kapitel wird der Versuch beschrieben, den Stimmapparat nachzubauen, um dort gezielte Experimente durchzuführen. Da der Meßablauf sich in diesem Kapitel etwas anders gestaltet, wird der Ablauf der Messung noch einmal explizit dargestellt. Es wird der Rachenraum nachgebildet und geprüft, ob sich Ähnlichkeiten zum Menschen feststellen lassen können. Den Abschluß dieses großen Teils bildet eine alternative Meßmethode zu dem benutzten Analysesystem. In Kapitel 9 wird das Cassy-System vorgestellt und mit dem eingesetzten System verglichen, damit die Vor- und Nachteile der Meßverfahren hervorgehoben werden können.

Ziel der Arbeit ist es, prinzipielle Aussagen über die Funktionsweise des Stimmapparats zu treffen. Wie genau die einzelnen Organe des Stimmapparats zusammenwirken, steht dabei im Vordergrund. Mit anderen Worten gesagt, wie und was der Mensch bewegt, um einen bestimmten Laut zu erzeugen. Gleichwohl muß man sich verdeutlichen, daß es zum Teil große Unterschiede zwischen verschiedenen Menschen, sei es durch das Geschlecht, den Körperbau, das Herkunftsland oder die Kultur, gibt. Somit ist es schwierig, genaue Konstanten oder Gleichungen zu ermitteln, die eine bestimmte Änderung des Tons einer bestimmten Bewegung im Stimmapparat zuordnet. Um exakte Zahlenwerte herauszufinden, müßte man entweder nach den obenerwähnten Merkmalen differenzieren oder aber eine sehr große Anzahl an Probanden analysieren. Möchte man aber die Funktionsweise des menschlichen Stimmapparats verstehen, so genügt es die Gemeinsamkeiten an wenigen Probanden aufzuzeigen.

## **2 Funktion des menschlichen Stimmapparats**

Wie kommt es dazu, daß sich Menschen durch ihre Stimme mit anderen Menschen unterhalten können? Welche Voraussetzungen müssen erfüllt sein, damit die Stimme funktioniert? Und welche Vorgänge laufen ab, wenn der Mensch redet?

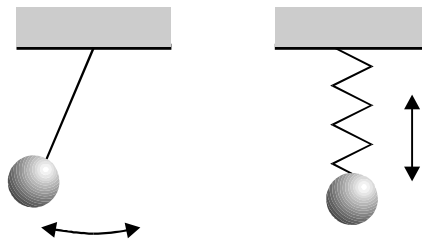
Bevor nun auf die biophysikalischen Voraussetzungen eingegangen wird, werden kurz die physikalischen Grundlagen angesprochen, denn diese sind zum Verständnis der Vorgänge bei der Sprachproduktion wichtig. Der physikalische Bereich der Akustik, der ein Teilgebiet der allgemeinen Schwingungslehre ist, gibt auf viele aufkommende Fragen eine Antwort.

## 2.1 Physikalische Voraussetzungen der Stimmerzeugung

### 2.1.1 Die harmonische Schwingung

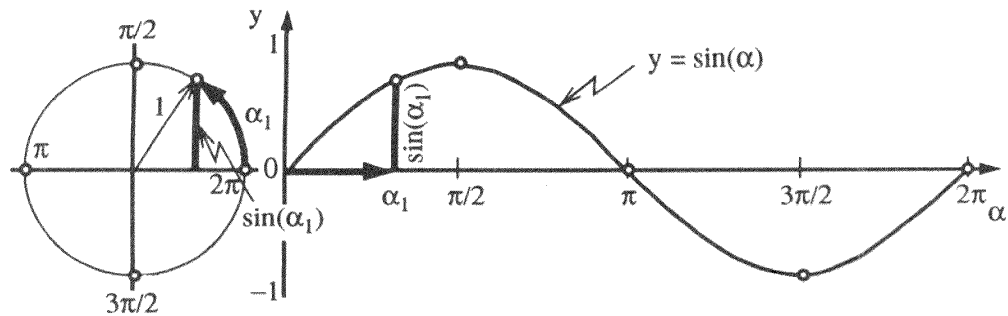
Die vermittelte akustische Schwingung, auch Sprachschall genannt, ist eine periodische Schwingung. Eine Schwingung ist ein Vorgang, bei dem ein Teilchen nach immer gleichen Zeitintervallen den gleichen Zustand annimmt. Ein Beispiel für eine solche Bewegung ist die eines Pendels, das zu schwingen beginnt, wenn es einmal aus der Ruhelage gebracht wurde. Um die Ruhelage herum ist die Bewegungsgeschwindigkeit maximal. An der Maximalauslenkung ist das Pendel dann auf eine Geschwindigkeit von Null abgebremst, kehrt um und bewegt sich durch die Ruhelage zum anderen Wendepunkt.

In Abbildung 2-1 werden zwei Beispiele für den eben beschriebenen Prozeß gegeben. Einmal handelt es sich um ein Fadenpendel, das aufgrund der Erdanziehung schwingt. Das andere Mal handelt es sich um ein Federpendel, das sich infolge der rücktreibenden Kraft einer Feder bewegt.



**Abbildung 2-1** Skizze eines Faden- und Federpendels

Den oben beschriebenen Bewegungsprozeß kann man durch geringfügige Abstraktionen an folgendem Graphen veranschaulichen:



**Abbildung 2-2** Illustration der Größen, die eine Sinusschwingung beschreiben nach Pompino-Marschall [2]

Um die in Abbildung 2-2 dargestellte harmonische Schwingung (Sinusschwingung) vollständig zu beschreiben, benötigt man drei Größen. Die erste Größe ist die maximale Auslenkung, die Amplitude  $A$ , die zweite die Frequenz  $f$ , die die Zahl der Durchläufe pro Sekunde angibt (1 Durchlauf pro Sekunde entspricht 1 Hz) und die dritte die Phase  $j$ , welche die Auslenkung zum Zeitpunkt  $t = 0$  beschreibt. In Gleichung 2-1 werden diese drei Größen miteinander verbunden:

$$y(t) = A \sin(2\pi f t + j) \quad \text{Gl. 2-1}$$

$y(t)$  ist die Auslenkung zur Zeit  $t$ .

### 2.1.2 Überlagerung von harmonischen Schwingungen

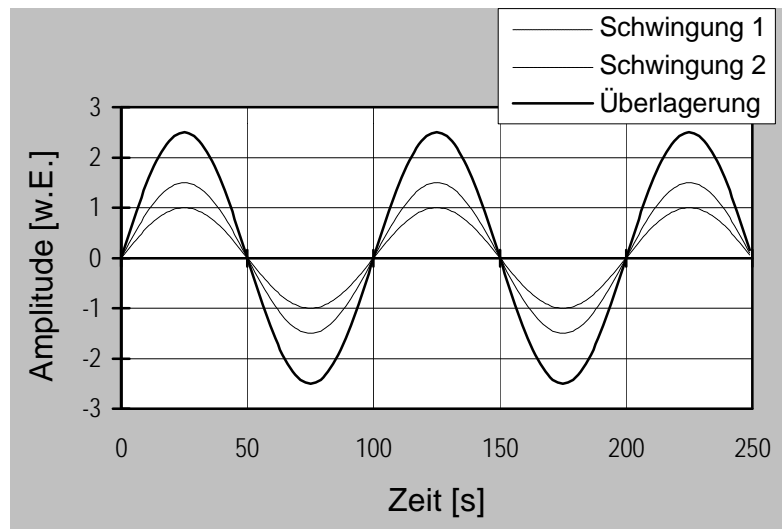
Ist ein Körper mehreren Sinusschwingungen gleichzeitig ausgesetzt, werden diese einzelnen Schwingungen zu einer Gesamtschwingung addiert.

$$y(t) = \sum_n y_n(t) \sum_n A_n \sin(2\pi f_n t + j_n) \quad \text{Gl. 2-2}$$

Gleichung 2-2 beschreibt die Summation („Superposition“) der einzelnen Schwingungen  $y_n$ .

Man kann nun Überlagerungen unterschiedlicher Schwingungen betrachten. Zur Vereinfachung werden in der weiteren Ausführung die Überlagerungen von zwei Sinusschwingungen untersucht.

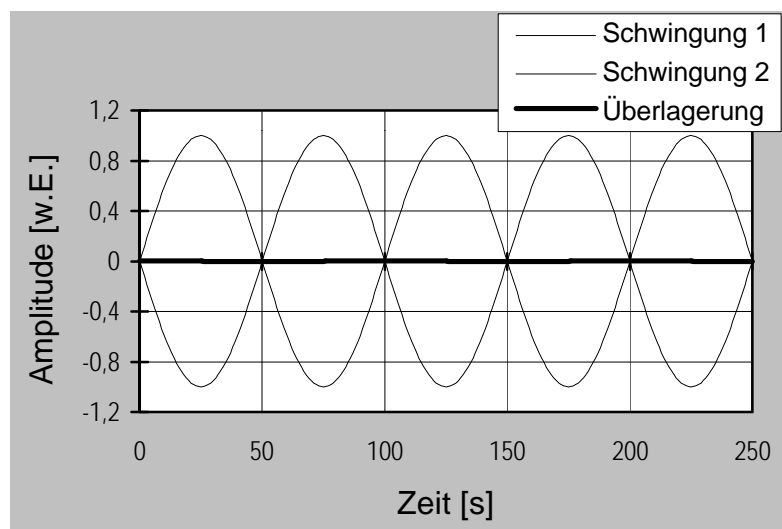
Es gibt die Superposition von zwei Schwingungen mit verschiedenen Amplituden, gleicher Frequenz und gleicher Phase.



**Abbildung 2-3** Überlagerung von Schwingungen verschiedener Amplituden

Abbildung 2-3 zeigt die resultierende Schwingung, die sich aus der Summation der Teilschwingungen zusammensetzt.

Besteht eine Phasendifferenz von  $180^\circ$  bei gleicher Amplitude und gleicher Frequenz, so ergibt die Summe der einzelnen Schwingungen zu jedem Zeitpunkt den Wert Null. Die folgende Abbildung 2-4 verdeutlicht diesen Zusammenhang.



**Abbildung 2-4** Überlagerung von Schwingungen mit Phasendifferenz

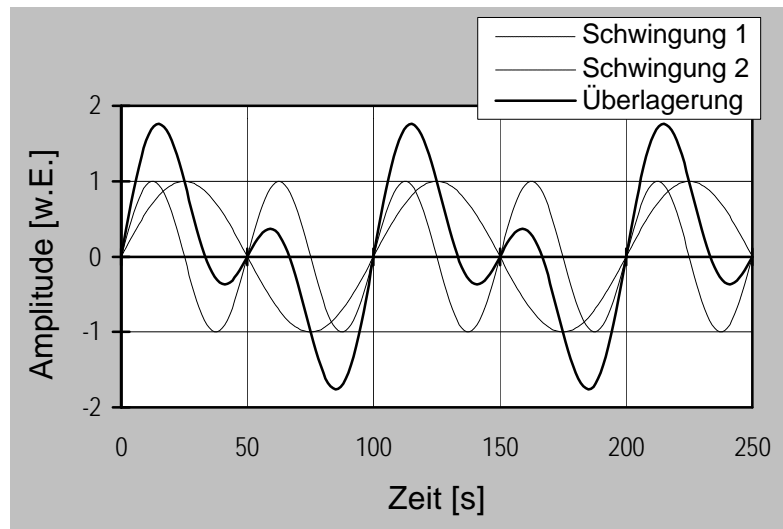
Werden zwei Schwingungen mit willkürlicher Phasendifferenz ausgesucht, so erhält man die resultierende Schwingung wieder durch Addition der



Partialschwingungen. Die Amplitude wächst zwischen Null und ihrem Maximalwert, wenn die Phasendifferenz zwischen Null und  $180^\circ$  variiert wird.

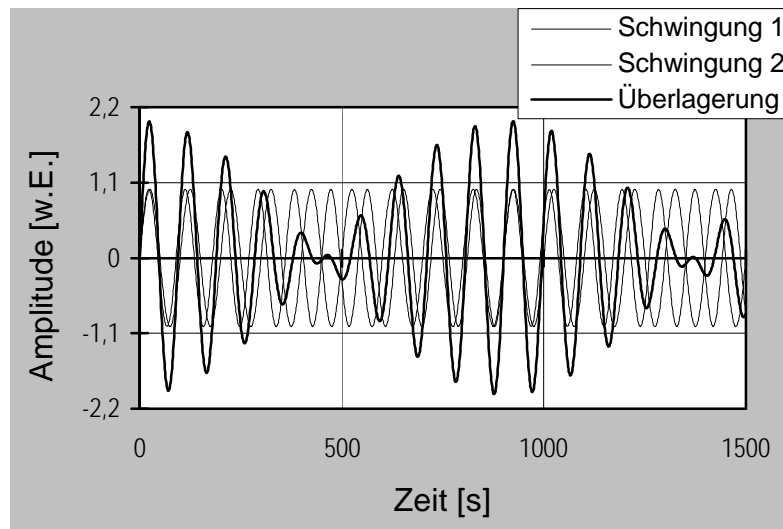
In allen oben gezeigten Beispielen sieht man, daß das Resultat immer eine harmonische Schwingung ist, wenn man die Amplitude und die Phase der einzelnen Schwingungen variiert und diese addiert.

Werden jedoch Partialschwingungen überlagert, wobei die eine genau die doppelte Frequenz der anderen Schwingung hat, so kann man sich die resultierende Schwingung durch die Abbildung 2-5 verdeutlichen.



**Abbildung 2-5** Überlagerung von Schwingungen mit doppeltem Frequenzunterschied

Überlagert man Partialschwingungen, die sich nur geringfügig in ihrer Frequenz unterscheiden, so entsteht eine „Schwebung“. Abbildung 2-6 zeigt diesen Fall:



**Abbildung 2-6** Schwebung

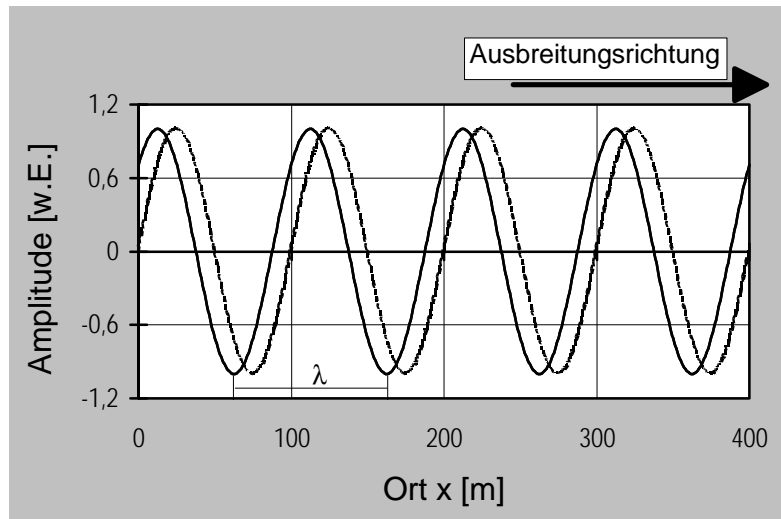
Die resultierende Schwingung ist harmonisch, jedoch schwankt ihre Amplitude ebenfalls harmonisch zwischen Null und einem Maximalwert.

Die Analyse aller Arten von Überlagerungen von Schwingungen werden in Kapitel 3 beschrieben.

### 2.1.3 Der Schall

Bevor man den Begriff des Schalls klärt, sollte man sich erst den Begriff der Welle erschließen. In der Natur ist man umgeben von den unterschiedlichsten Wellenformen. Die Schallwelle, Wasserwelle, Radiowelle und Lichtwelle sind nur einige der vorkommenden Wellenarten in der Natur. Aber was ist denn überhaupt eine Welle?

Wellen sind Schwingungen, die sich in Raum und Zeit ausbreiten. Vergleicht man allerdings die Wasserwelle mit der Schallwelle, so unterscheiden sie sich in einem Punkt. Die einzelnen Wasserteilchen vollführen eine Schwingung in einer Ebene senkrecht zur Ausbreitungsrichtung der Welle (Abbildung 2-7). Diese Art der Welle nennt man Transversalwelle.



**Abbildung 2-7** Bewegung einer Welle (Wasserwelle)

Die Luftteilchen der Schallwelle hingegen bewegen sich in Ausbreitungsrichtung, was eine Druckänderung zur Folge hat. Eine solche Welle bezeichnet man als Longitudinalwelle.

Der einzige Unterschied zwischen den beiden Wellenarten besteht also in der Richtung der Schwingung relativ zur Ausbreitungsrichtung. Ansonsten haben beide die gleichen Eigenschaften, wie zum Beispiel die Ausbreitungsgeschwindigkeit.

Die Geschwindigkeit  $c$  der Wasserwelle setzt sich zusammen aus der Wellenlänge  $l$ , welche den Abstand zweier Teilchen beschreibt, die die gleiche Auslenkung besitzen und der Zeit  $T$ , die ein Teilchen für eine Periode benötigt (Gleichung 2-3).

$$c = \frac{l}{T} \quad \text{oder} \quad c = l f \quad \text{Gl. 2-3}$$

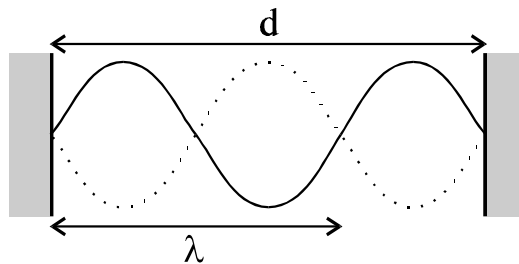
Im Jahre 1640 ermittelte der französische Mathematiker Mersenne zum ersten Mal die Ausbreitungsgeschwindigkeit  $c$  der Schallwellen. Er benutzte eine Kanone, von der er den Abstand zu sich kannte, und maß die Zeit zwischen dem Lichtblitz und dem Eintreffen des Knalls der Kanone. Es zeigte sich jedoch in den nächsten Jahren, daß sich Schallwellen in unterschiedlichen Medien und verschiedenen Temperaturen verschieden schnell ausbreiten. Je dichter das Material und je höher die Temperatur, desto größer ist die Schallgeschwindigkeit. Der Schall breitet sich laut Kuchling [6] in Luft bei 0°C mit  $332 \frac{\text{m}}{\text{s}}$  und bei 20°C mit  $344 \frac{\text{m}}{\text{s}}$  aus. In anderen Gasen, wie zum Beispiel

Helium, beträgt die Schallgeschwindigkeit  $1007 \frac{m}{s}$ . Die Auswirkung von unterschiedlichen Schallgeschwindigkeiten in verschiedenen Gasen wird in Kapitel 7.4 durch ein Experiment illustriert.

Zum Verständnis des Kapitels 2.2.3 wird im folgenden die Entstehung einer stehenden Welle erläutert. Eine stehende Welle entsteht durch Reflexion einer Welle an einem senkrechten Hindernis. Falls es zwei gleiche Enden (zwei freie oder zwei feste Enden) gibt, kann sich eine stehende Welle nur ausbilden, wenn deren Abstand ein ganzzahliges Vielfaches der halben Wellenlänge beträgt (Abbildung 2-8).

$$d = 3 \frac{\lambda}{2}$$

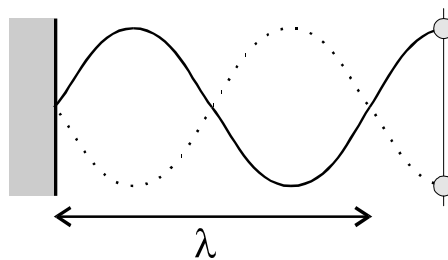
Gl. 2-4

**Abbildung 2-8** Stehende Welle mit festen Enden

Sind die Enden jedoch verschieden (ein festes und ein freies Ende), so muß deren Abstand ein ungeradzahliges Vielfaches von  $\frac{\lambda}{4}$  betragen.

$$d = 5 \frac{\lambda}{4}$$

Gl. 2-5

**Abbildung 2-9** Stehende Welle mit einem festen und einem losen Ende

Stehende Wellen breiten sich also, wie der Name schon sagt, nicht im Raum aus. An einer Stelle des Knotens in Abbildung 2-8 und Abbildung 2-9 bleibt stets ein Knoten. Die Teilchen bleiben an dieser Stelle also in Ruhe.

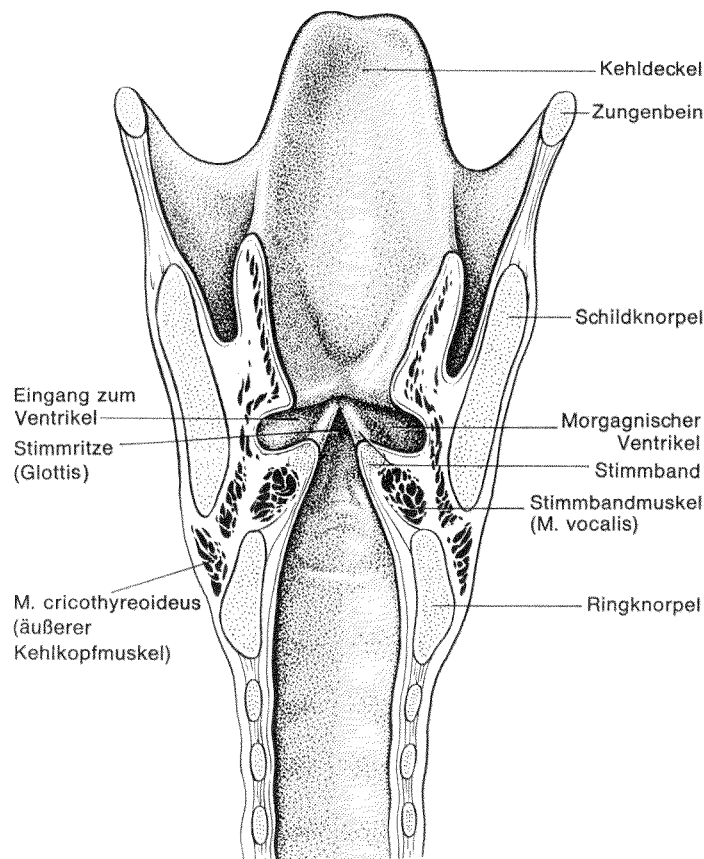
Für weitere Eigenschaften von Wellen, wie die Überlagerung, Brechung und Beugung von Wellen, verweise ich auf Alonso/Finn [7] oder andere Fachliteratur.

## 2.2 Biophysikalische Voraussetzungen

Durch ein kompliziertes Zusammenspiel verschiedener Organe kann der Mensch sprechen. Die Lunge, der Kehlkopf mit den Stimmlippen und das Ansatzrohr formen durch ihr Zusammenspiel die menschliche Stimme.

### 2.2.1 Der Kehlkopf als Klangerzeuger

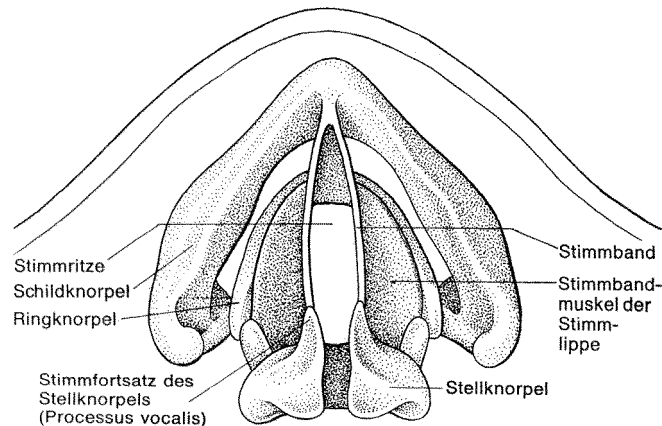
Die primäre Aufgabe des Kehlkopfes ist es, die Lunge vor Fremdkörpern zu schützen. Die Stimmlippen schließen sich und verhindern so das Eindringen von Fremdstoffen. In zweiter Linie ist der Kehlkopf ein Klanggenerator, der für die weiteren Betrachtungen von Interesse sein wird. Nach oben hin begrenzt der Kehlkopf die Luftröhre (Abbildung 2-10) und gehört zum Atemtrakt.



**Abbildung 2-10** Frontalschnitt des Kehlkopfes von hinten nach Habermann [8]

Der Kehlkopf baut sich aus etlichen Knorpeln auf, die gelenkig miteinander verbunden sind und durch Bänder zusammengehalten werden. Durch Muskeln ist es möglich, daß sich der Kehlkopf bewegen kann. Zwischen Schildknorpel

und Ringknorpel, im Inneren des Kehlkopfes, befinden sich die Stimmlippen (Abbildung 2-11). Sie bestehen im wesentlichen aus Muskelgewebe, dem „Musculus vocalis“, dessen innere Ränder in den Stimmbändern enden. Die Stimmbänder sind Membranen, die überwiegend aus elastischen Fasern bestehen. Zwischen den Stimmlippen liegt die Stimmritze, auch Glottis genannt. Der eigentliche Stimmklang entsteht in der Glottis.

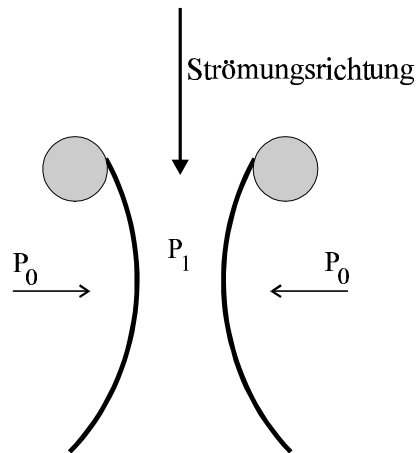


**Abbildung 2-11** Schematische Aufsicht auf die Stimmlippen (Habermann) [4]

Wie der Sprechton entsteht, kann durch die Stimmgebung (Phonation) beschrieben werden. Will man einen Sprechton erzeugen, so schließen sich die Stimmlippen mit dem Beginn des Ausatmens. Unterhalb der geschlossenen Glottis erhöht sich der statische Luftdruck. Wenn der Druck groß genug geworden ist, ist er in der Lage den Glottisverschluß zu „sprengen“. Zwischen den Stimmlippen entsteht eine Düse durch die die Luft sehr schnell entweichen kann. In der Düse besteht eine große Strömungsgeschwindigkeit, die einen Unterdruck zur Folge hat. Bedingt durch die „Sogkräfte“ des Unterdrucks (aerodynamischer Effekt) und durch Mithilfe der Muskel ziehen sich die Stimmlippen wieder zusammen. Dieser Vorgang wiederholt sich, wenn der Überdruck hinter der Stimmritze wieder angestiegen ist. Der Öffnungs- und Schließungsvorgang ist ein sich selbst steuernder und erhaltender zyklischer Prozeß. [2,8,9]

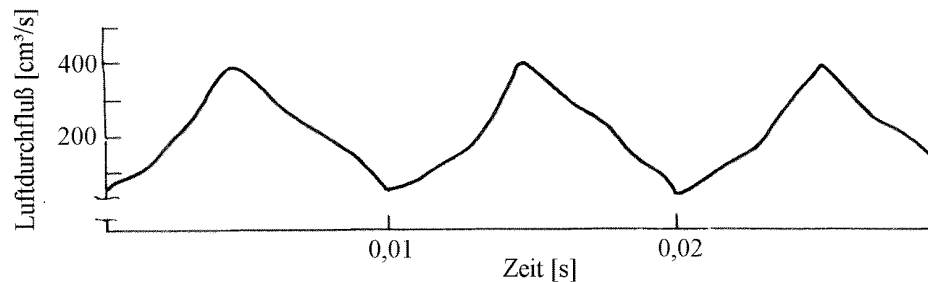
Daniel Bernoulli (1700-1782) beschreibt den Effekt der strömungsdynamischen Gesetze in seiner Strömungsgleichung.[10] Den Einfluß des Zusammenziehens der Stimmlippen bezeichnet man in der Physik als aerodynamisches Paradoxon. Man kann dieses einfach demonstrieren, indem man zwischen zwei Blättern hindurchbläst und beobachtet, wie sich die Blätter bewegen. Die Blätter gehen nicht, wie man zuerst annehmen würde, auseinander, sondern

ziehen sich aufgrund des Unterdrucks, welcher zwischen den Blättern besteht, zusammen. Abbildung 2-12 illustriert das geschilderte Beispiel. [11]



**Abbildung 2-12** Darstellung des aerodynamischen Paradoxons ( $P_1 < P_0$ )

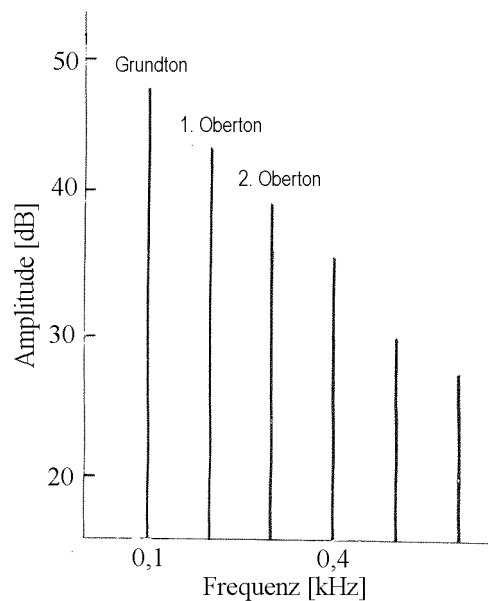
Abbildung 2-13 stellt den Luftdurchfluß durch die Glottis während der Phonation dar.



**Abbildung 2-13** Stimmbandschwingungen während des Luftdurchflusses [12]

Da man die Stimmbandschwingungen als Überlagerung von verschiedenen harmonischen Schwingungen (Kapitel 2.1.1) verstehen kann, kann man diese in einem sogenannten Frequenzspektrum (Kapitel 3.3.1) darstellen. Der entstandene Ton setzt sich aus dem Grundton und den Obertönen zusammen. In dem Frequenzspektrum (Abbildung 2-14) werden Tonhöhe und Stärke des Grundtons und der Obertöne aufgetragen.





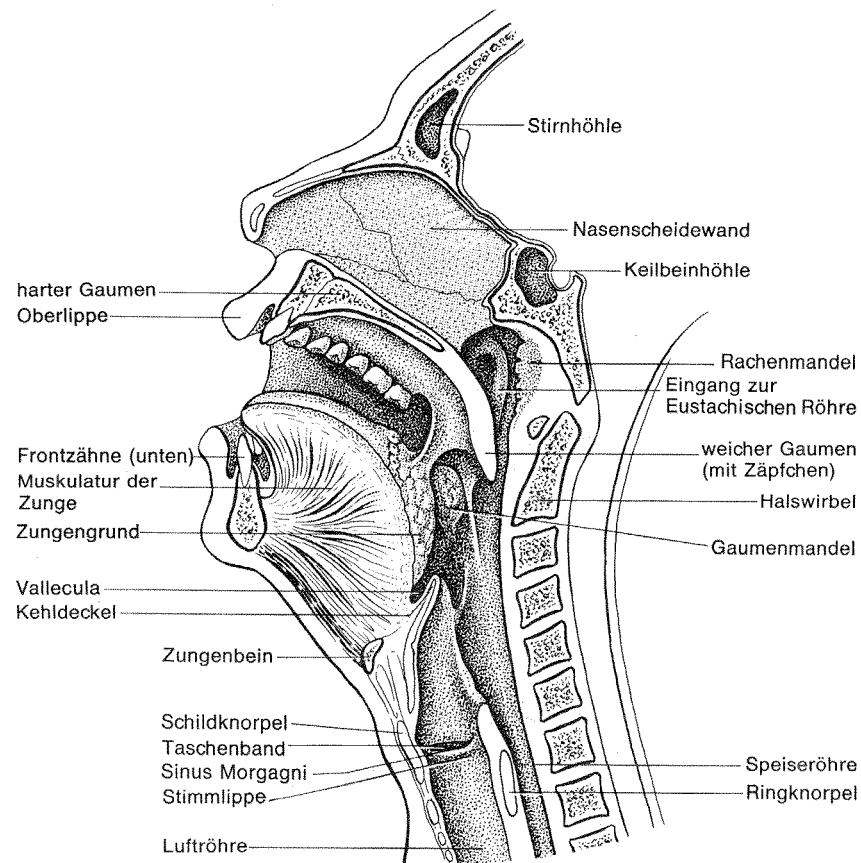
**Abbildung 2-14** Frequenzspektrum der Stimmbandschwingungen [12]

Die Tonhöhe hängt von der Länge der Stimmlippen, der Weite der Glottis und der Masse der Stimmlippen ab. Aus diesem Grund kann man Männerstimmen und Frauenstimmen recht gut unterscheiden. Bei Männern sind die Stimmlippen durch die Form des Schildknorpels („Adamsapfel“) länger als bei den Frauen, wodurch Männerstimmen eine niedrige Grundfrequenz besitzen.

Zum Erzeugen tiefer Töne sind die Stimmlippen entspannt und die Stimmritze wenig geöffnet. Infolgedessen schwingen die Stimmlippen langsam und ausholend. Steigt nun die Tonhöhe, so strecken sich die Stimmlippen mit Hilfe der Muskeln, die Glottis schließt sich allmählich und die Stimmlippen-schwingungen werden schneller. Der erzeugte Ton ähnelt einem undeutlichen „aaaaa“ und klingt nicht wie ein Sprachlaut.[1]

### 2.2.2 Das Ansatzrohr als Resonanzraum

In der Musik verwendet man den Begriff des „Ansatzrohrs“ für Blasinstrumente. Aufgrund der Ähnlichkeit zum menschlichen Vokaltrakt hat man den Begriff des Ansatzrohrs aus der Musik entlehnt [1,8]. Durch Änderung der Form und Größe des Ansatzrohrs und durch Verschluß- oder Engbildung einzelner „Teile“ im Ansatzrohr können Menschen Laute formen.



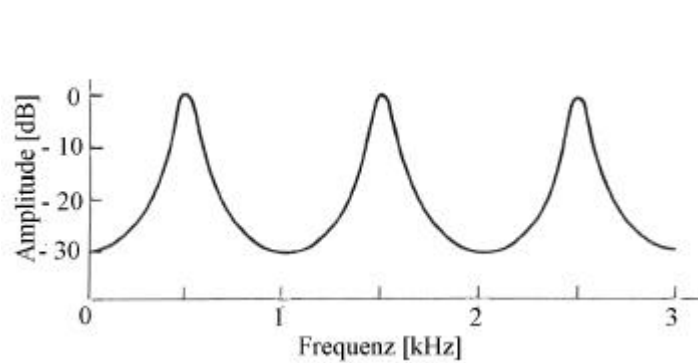
**Abbildung 2-15** Schematische Schnittzeichnung des Ansatzrohrs [8]

Das Ansatzrohr (Abbildung 2-15) erstreckt sich von den Stimmlippen bis zu den Mundlippen und der Nasenöffnung. Zum Ansatzrohr gehören der Rachenraum, die Mundhöhle und der längsgeteilte Nasenraum. Der Rachenraum dehnt sich von den Stimmlippen bis zum weichen Gaumen aus. Die daran anschließende Mundhöhle wird vom Gaumen und den Wangen sowie dem Mundboden und den Zahnreihen begrenzt. In der Mundhöhle befindet sich die Zunge, die einen großen Teil des Mundraums einnimmt. Durch ihre Beweglichkeit leistet sie den wesentlichen Beitrag zur menschlichen Artikulation. Der Nasenraum erstreckt sich vom weichen Gaumen bis zur Nasenöffnung. [1,8,13,14]

Der Nasenraum ist für die weiteren Betrachtungen nicht von Bedeutung, weil er bei der Filterfunktion des Ansatzrohrs keine große Rolle spielt, denn der weiche Gaumen trennt den Nasenraum vom Rest des Ansatzrohrs.

### 2.2.3 Filterfunktion des Ansatzrohrs - Vokalformanten

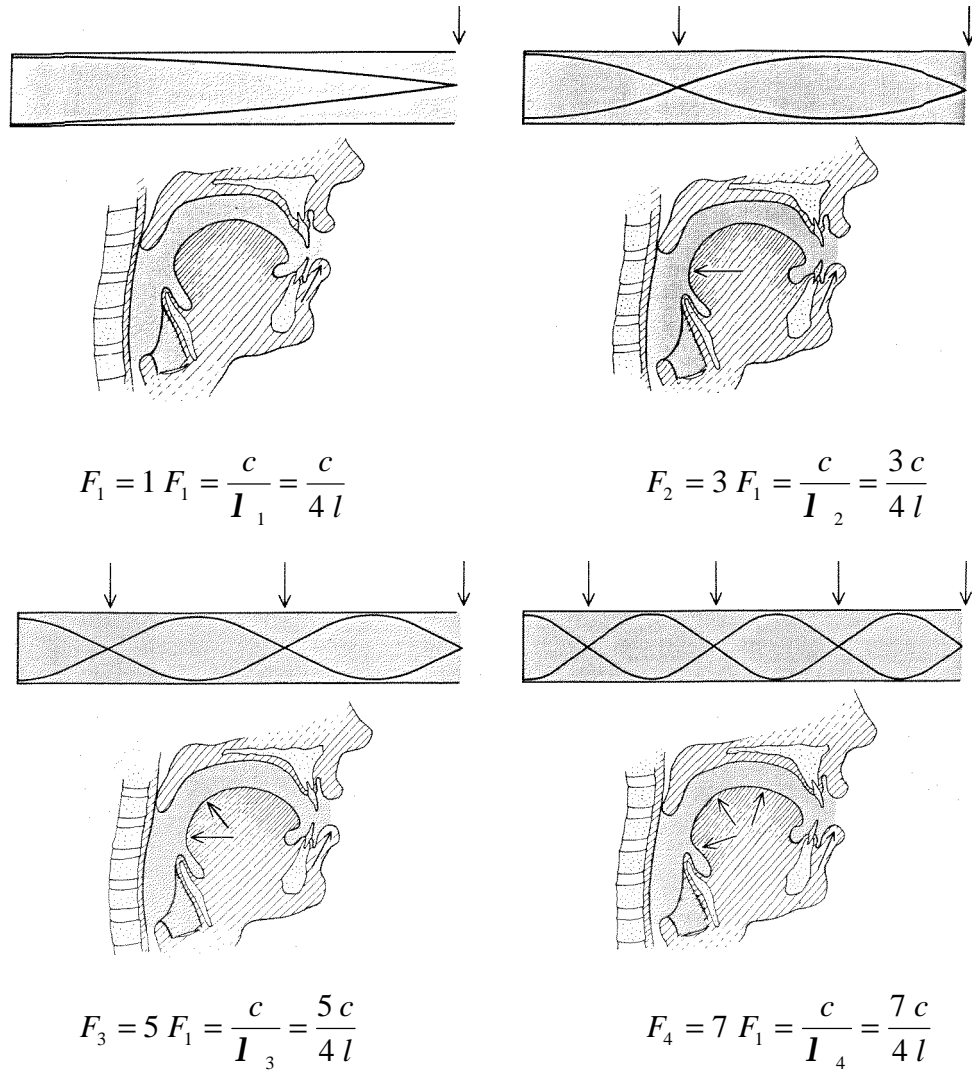
Das Ansatzrohr (Rachenraum und Mundraum) fungiert als akustischer Filter. Er läßt verschiedene Frequenzen unterschiedlich gut passieren. Der akustische Filter besitzt also verschiedene Eigenfrequenzen, die auch Eigenresonanzen genannt werden. Eigenresonanzen sind Frequenzbereiche, die den eingestrahltten Klang verstärken. Andere Frequenzen werden weniger verstärkt oder gar unterdrückt. Die idealisierte Frequenzkurve einer solchen Filterfunktion ist in Abbildung 2-16 dargestellt.



**Abbildung 2-16** Filterfunktion des Ansatzrohrs [12]

Möchte man nun einen Vergleich zum Ansatzrohr ziehen, so kommt das einseitig geschlossene Rohr in Betracht, wie zum Beispiel eine Orgelpfeife. In einem Rohr bilden sich durch mehrmalige Reflexion stehende Wellen aus (Kapitel 2.1.2). Diese sind die Ursachen für die Eigenresonanzen der Luftsäule im Rohr. Die niedrigste Eigenresonanz, der Grundton, wird durch die Wellenlänge bestimmt, von der ein Viertel in die Rohrlänge paßt. Zu höheren Resonanzen, den Obertönen, werden Wellenlängen benötigt, von denen jeweils ein ungeradzahliges Viertel in die Rohrlänge paßt.

In Abbildung 2-17 wird der Vergleich zwischen einem einseitig geschlossenem Rohr und dem menschlichen Ansatzrohrs dargestellt. [10]



**Abbildung 2-17** Analogie des Ansatzrohrs mit einem einseitig geschlossenen Rohr. Modell zur Erzeugung der Formanten [15]

Da die Wellenlängen der Eigenresonanzen von der Rohrlänge  $l$  abhängig sind, kann man sie durch die Beziehung der Frequenz  $f$ , der Wellenlänge  $l$  und der Schallgeschwindigkeit  $c$  beschreiben. Die allgemeine Gleichung für die Entstehung der Eigenresonanzen lautet:

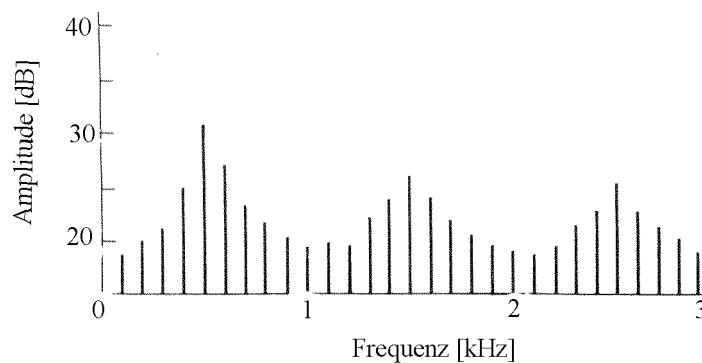
$$F_n = (2 n - 1) F_1 = \frac{c}{I_n} = \frac{c(2 n - 1)}{4 l} \quad \text{Gl. 2-6}$$

Wenn man von der menschlichen Stimme spricht, werden die Eigenresonanzen (Knotenpunkte in der Abbildung 2-17) Formanten. [1,4]

Aus der mittleren Länge des Ansatzrohrs (17,5 cm) eines Mannes lassen sich somit die mittleren Frequenzen der Ansatzrohrresonanzen berechnen. Die

Formanten liegen bei 500 Hz, 1500 Hz, 2500 Hz, 3500 Hz und so weiter. Diese Frequenzabfolge ergibt sich für den Idealfall der runden Röhre. Die Formanten von Frauen und Kindern liegen etwas höher als bei Männern, da die Ansatzrohrängen bei Frauen und Kindern meistens geringer sind. Durch Verformung des Mundbereiches, der Lippen und der Zunge kann man die Länge des Ansatzrohrs und damit die Amplitude des Grundtons und der Obertöne stark verändern. [3,4,15]

Das resultierende Frequenzspektrum (Abbildung 2-18), das der Mensch als Schall aussendet, setzt sich aus dem Spektrum der Stimmbandschwingungen (Abbildung 2-13) und der Filterfunktion des Rachens (Abbildung 2-16) zusammen.



**Abbildung 2-18** Idealisiertes Frequenzspektrum des gesprochenen Tons [8]

Wie die verschiedenen Mundstellungen der Vokalartikulation und deren Frequenzspektren aussehen, wird in Kapitel 6 erläutert.

Weiterführende Literatur, die zum Teil über die Erfordernisse dieser Arbeit hinausgehen, finden sich in der Sekundärliteratur bei Fant [3], Kent [16], Sataloff [17] und Sundberg [18].

### 3 Mathematische Analyseverfahren

Nachdem in den letzten Kapiteln auf die physikalischen und biophysikalischen Voraussetzungen der Spracherzeugung eingegangen wurde, werden in diesem Abschnitt die grundlegenden Analyseverfahren erörtert, die in der Sprachforschung verwendet werden. Dieses Kapitel dient in erster Linie zur Grundlage der in Kapitel 4 vorgestellten Programme.

Wesentlich für die Verarbeitung von Sprachsignalen ist es, den vom Sprecher ausgesandten Schall durch eine kleine Datenmenge zu beschreiben. Die Datenmenge muß aber das ursprüngliche Sprachsignal hinreichend genau durch Parameter beschreiben. Die Parameter sind

- die Energie des Sprachsignals, die die Lautstärke bestimmt und
- das Frequenzspektrum, woraus die Grundfrequenz und die Formanten bestimmt werden können.

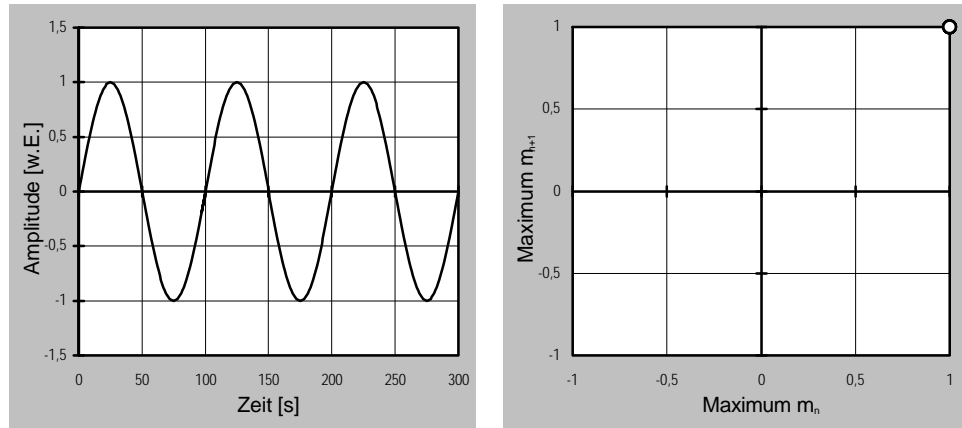
Im folgenden werden Analyseverfahren mathematisch beschrieben, die für die Bestimmung der akustischen Parameter häufig eingesetzt werden.

Das erste Analyseverfahren, die Lorenzabbildung, wird sehr häufig zur Analyse von regulären, stochastischen und chaotischen Signalklassen verwendet. Gleichwohl ist das Verfahren auch zum Anfertigen eines „sprecherspezifischen Fingerabdrucks“ und der Analyse von verschiedenen Schallformen geeignet, wie es sich im Laufe dieser Arbeit herausgestellt hat. Da man durch diese Darstellungsform sehr viel über die Struktur und Periodizität des Abtastsignals aussagen kann, wird dieses Analyseverfahren hier eingeführt. Die anderen Verfahren analysieren das ausgesandte Sprachsignal nach den obengenannten Parametern. Die Autokorrelationsfunktion und die Fouriertransformation sind Verfahren, die in den nächsten Kapiteln mathematisch erläutert werden.

### 3.1 Die Lorenzabbildung

Die Lorenzabbildung, die auf E. N. Lorenz zurückgeht, setzt jeweils aufeinanderfolgende Maxima zueinander in Bezug, indem man die Amplitude des  $n$ -ten Maximums als Abszisse und die des  $(n+1)$ -ten als Ordinate aufträgt. Man erhält durch diese verblüffend einfache Darstellungsart aussagekräftige Bilder. Betrachtet man eine periodische Funktion mit gleicher Amplitude, zum Beispiel eine Sinusfunktion, so ergibt sich in der Lorenzabbildung ein einziger Punkt, weil die Amplituden der Maxima übereinstimmen. [19]

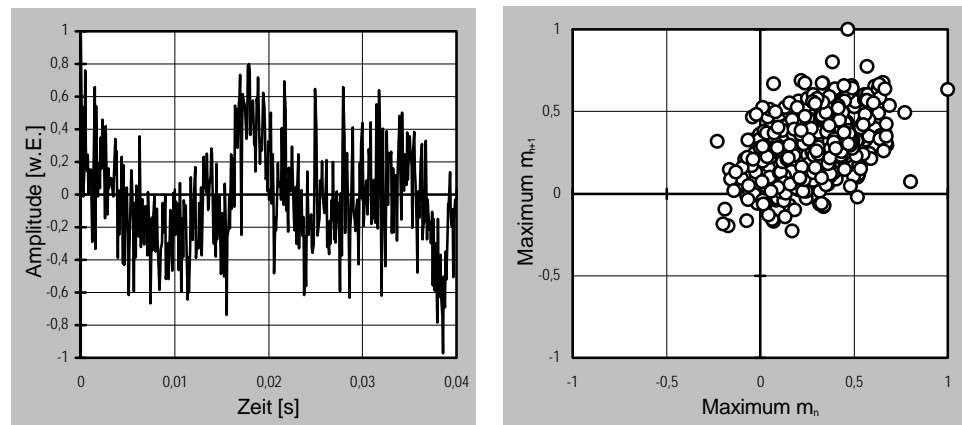
Abbildung 3-1 verdeutlicht diesen Zusammenhang.



**Abbildung 3-1** Sinusförmiges Signal und dessen Lorenzabbildung

Signale, die periodisch sind, weisen in der Lorenzabbildung genau die Punktzahl auf, die die Periodizität des Signals beinhaltet. Hat man eine Sinusschwingung mit  $n$  verschiedenen Maxima pro Periode als Ausgangssignal, so ergeben sich in der Lorenzabbildung exakt  $n$  Punkte.

Häufig wird dieses Verfahren in der Chaosforschung eingesetzt, um verschiedene Signalklassen zu unterscheiden. Zu diesen gehören reguläre (z.B. periodische Funktionen), stochastische (z.B. Rauschen) und chaotische Signalklassen. Die Untersuchung eines  $1/f$  Rauschsignals zeigt Abbildung 3-2.



**Abbildung 3-2**  $1/f$  Rauschsignal und dessen Lorenzabbildung

Die Lorenzabbildung des Rauschsignals zeigt eine Häufung der Punkte innerhalb des ersten Quadranten. Diese Anhäufung der Punkte weist darauf hin, daß die meisten Maxima des Ausgangssignals oberhalb des Nullpunkts liegen, sich aber keine Periodizität ausmachen läßt.

Während der Arbeit hat sich herausgestellt, daß die Lorenzabbildung ein nützliches Verfahren zur Erläuterung verschiedener Schallformen der

menschen Sprache ist. Überdies erweist sich die Lorenzabbildung als Möglichkeit einen „sprecherspezifischen Fingerabdruck“ zu erstellen.



## 3.2 Die Autokorrelationsfunktion

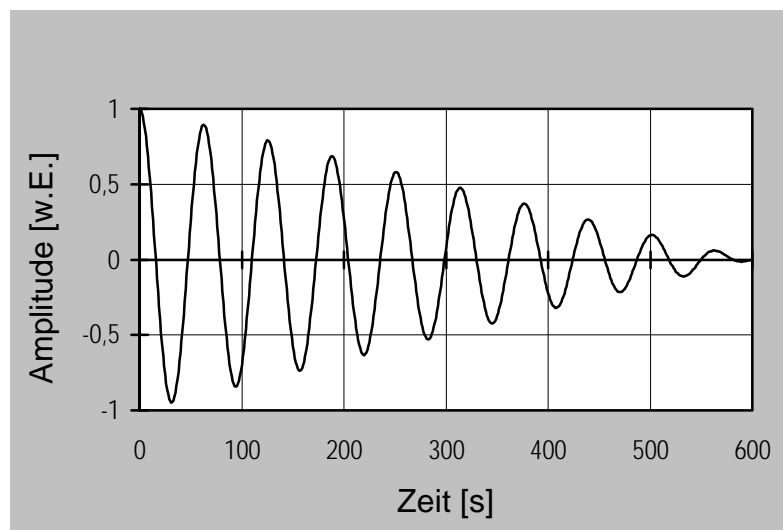
Die Autokorrelationsfunktion ist ein einfaches Verfahren zur Analyse von Rauschsignalen sowie von periodischen Signalen. Sie sucht nach Selbstähnlichkeiten („Korrelationen“) in der zeitlichen Abfolge einer Funktion. Ein beliebiger Prozeß kann Korrelationen aufzeigen, die nicht periodisch sind. Ein chaotisches Pendel führt zum Beispiel häufig ähnliche Bewegungen durch, ohne daß die Bewegung periodisch ist. Auch das Wetter ist stark chaotisch. Trotzdem weist es Selbstähnlichkeiten, wie die Jahreszeiten, auf.

### 3.2.1 Mathematische Erläuterung

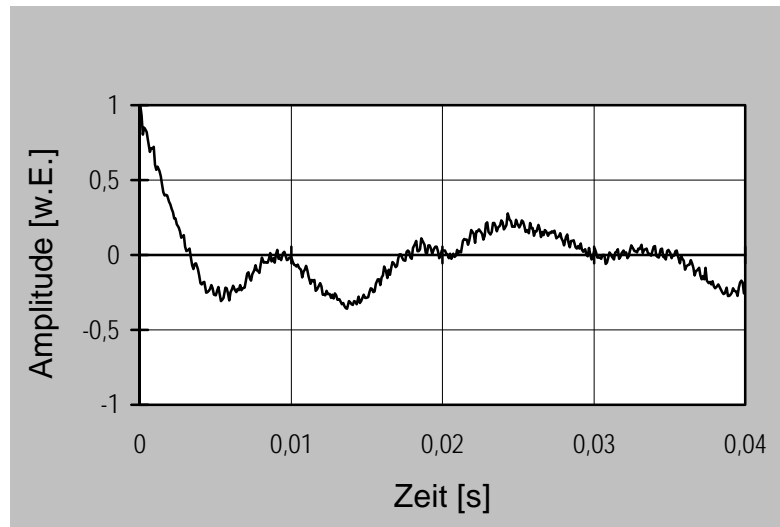
Die Autokorrelationsfunktion einer Funktion  $f(t)$  ist definiert als

$$R(l) = \int_{-\infty}^{\infty} f(t) f(t+l) dt . \quad \text{Gl. 3-1}$$

Das Argument  $l$  bezeichnet die zeitliche Verschiebung der Funktion gegen sich selbst. Die folgenden Abbildungen zeigen die Autokorrelationsfunktionen der Signalverläufe von Abbildung 3-1 und Abbildung 3-2:



**Abbildung 3-3** Autokorrelationsfunktion einer Sinusschwingung



**Abbildung 3-4** Autokorrelationsfunktion eines  $1/f$  Rauschens

Einmal handelt es sich um die Autokorrelationsfunktion eines periodischen Signals (Abbildung 3-3), der Sinusfunktion aus Abbildung 3-1. In Abbildung 3-4 ist die Autokorrelationsfunktion des rauschartigen Signals aus Abbildung 3-2 dargestellt.

An den Beispielen kann man einige Eigenschaften der Autokorrelationsfunktion zeigen, die wesentlich für die Verwendung als Erkennungsmerkmale sind:

- Sie hat ihr Maximum bei  $l=0$ . Der Koeffizient  $R(l)$  wird als Betragsmittel bezeichnet und ist proportional zur Energie.
- Sie hängt quadratisch von dem Absolutbetrag des Skalierungsfaktors ab, bei einer Skalierung aller Abtastwerte.
- Sie behält die periodische Struktur des Sprachsignals annähernd bei.
- Sie bleibt nahezu unabhängig von der Lage des Analysefensters, wenn es mehrere Abtastwerte nach rechts oder links verschoben wurde.
- Sie zeigt eine „fast regellose“ Struktur wie das Signal selbst (Abbildung 3-4).
- Sie kann hohe Werte annehmen, wenn die Funktion gut korreliert ist und nimmt niedrige Werte bei schlechter Korrelation an.
- Bei optimaler Korrelation nimmt die maximale Amplitude linear ab. Ist die Korrelation schlechter, sinkt die Amplitude schneller ab.

Möchte man nun die Autokorrelationsfunktion (Gleichung 3-1) für Spracherkennungszwecke verwenden, so hat sie den Nachteil, daß sie sehr empfindlich gegenüber Schwankungen der Abtastfunktion  $f(t)$  ist, da diese quadratisch eingeht. Zur Berechnung der Korrelation benötigt man das Produkt  $f(t)f(t+l)$ . Folglich ist die Autokorrelationsfunktion quadratisch in der Abtastfunktion  $f(t)$ . Dadurch werden große Korrelationen hervorgehoben und alle schwächeren Selbstähnlichkeiten unterdrückt.

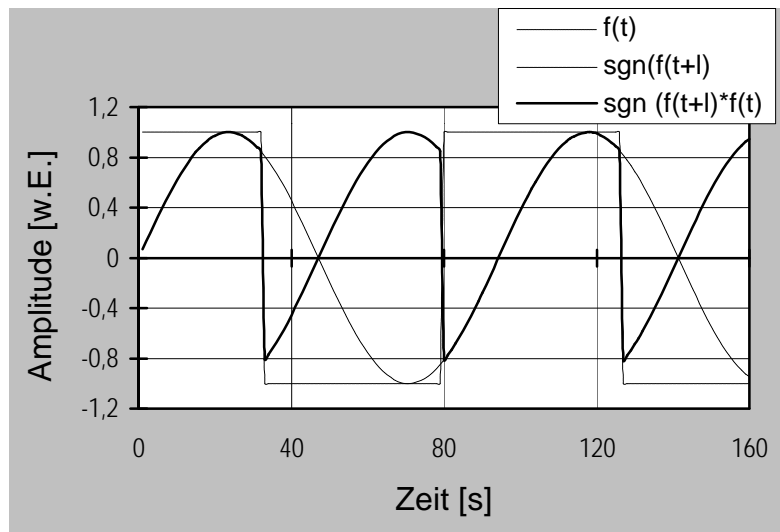
Diese Effekte kann man vermindern, indem man die geklippte Autokorrelationsfunktion einführt:

$$r(l) = \int_{-\infty}^{\infty} f(t) \operatorname{sgn}(f(t+l)) dt, \quad \text{Gl. 3-2}$$

wobei  $\operatorname{sgn}(a)$  als Vorzeichenfunktion definiert ist:

$$\operatorname{sgn}(a) = \begin{cases} +1 & a \geq 0, \\ -1 & a < 0. \end{cases}$$

Abbildung 3-5 zeigt  $x(i)$ , die um  $l$  verschobenen Vorzeichenfunktion  $\text{sgn}(x(i+l))$  und deren Produkt.



**Abbildung 3-5** Darstellung der Funktion für die geklippte Autokorrelationsfunktion

Nach Gleichung 3-2 wird die geklippte Autokorrelationsfunktion nur durch Additionen und Vorzeichenoperationen berechnet. Durch diese Operationen wird der Rechenaufwand des Computers herabgesetzt, denn es sind keine Multiplikationen mehr notwendig wie für die „normale“ Autokorrelationsfunktion. Aufgrund des linearen Zusammenhangs der geklippten Autokorrelationsfunktion werden nun auch schwache Effekte im Bild sichtbar. [20,21]

Um Meßdaten auf Autokorrelationen zu untersuchen, muß man diese diskret berechnen. Zu diesem Zweck ersetzt man die Funktion  $f(t)$  durch die diskreten Abtastwerte  $x(i)$ . Dadurch kann man das Integral durch eine Summe ersetzen, die sich statt von  $-\infty$  bis  $\infty$  über alle einzelnen Werte  $x(i)$  von  $i=1$  bis  $N$  erstreckt. Nach Ney [20] erhält man für die „normale“ Autokorrelationsfunktion:

$$R(l) = \sum_{i=0}^{N-1-l} x(i)x(i+l). \quad \text{Gl. 3-3}$$

Die geklippte Autokorrelationsfunktion errechnet sich aus:

$$r(l) = \sum_{i=0}^{N-1-l} x(i) \operatorname{sgn}(x(i+l)) \quad \text{Gl. 3-4}$$

Um den Rechenaufwand des Computers noch weiter herabzusetzen, kann die geklippte Autokorrelationsfunktion anders errechnet werden. Will man  $r(l+1)$  berechnen, so ist es möglich, die Vorzeichenfunktion um einen Punkt zu verschieben und dann den Wert neu zu berechnen. Um aber Rechenzeit zu sparen, muß man sich Abbildung 3-5 genauer betrachten und kann feststellen, daß sich  $r(l)$  und  $r(l+1)$  nur an den Stellen des Vorzeichenwechsels unterscheiden. Folglich ist es sinnvoller  $r(l+1)$  aus dem Unterschied zu  $r(l)$  zu berechnen. Dieser Algorithmus findet sich in keiner mir bekannten Literatur wieder, wurde aber von mir im Programm der Autokorrelationsfunktion als „Turbo“ verwirklicht (Kapitel 4.2).

### 3.3 Die Fouriertransformation

Wie im Kapitel 2.1.2 beschrieben, können sich Schwingungen durch Überlagerung zu einer im allgemeinen komplizierteren Schwingung zusammensetzen. In Umkehrung ist es auch möglich, eine komplizierte Schwingungsform in ihre Teilschwingungen zu zerlegen.

Jean Baptiste Fourier (1760-1830) zeigte, daß sich jede beliebige periodische Schwingung als eine Reihe von harmonischen Schwingungen unterschiedlicher Frequenzen zusammensetzen läßt. Wenn man aus der Überlagerung der einzelnen Schwingungen das ursprüngliche Signal erhält, findet man die Fourier-Transformierte des Ausgangssignals. Die Fourier-Transformation ist die Transformation einer zeitlichen Funktion vom Zeitraum in den Frequenz- und den Phasenraum.

#### 3.3.1 Mathematische Erläuterung

Mathematisch kann man die Fouriertransformation durch Gleichung 3-5 beschreiben. [5]

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \exp\{-i\omega t\} dt \quad \text{Gl. 3-5}$$

$f(t)$  ist das zu untersuchende Signal,  $F(\omega)$  ist die Fourier-Transformierte von  $f(t)$  und  $i$  die imaginäre Einheit.

Die inverse Fourier-Transformation, die die Transformation vom Frequenz- in den Zeitbereich beschreibt, findet sich in Gleichung 3-6.

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \exp\{i\omega t\} d\omega \quad \text{Gl. 3-6}$$

Da die Signale der Abtastfunktion nur numerisch vorliegen, ist die analytische Funktion  $f(t)$  nicht bekannt. Aus diesem Grund ist man gezwungen die Fourier-Transformierte numerisch zu berechnen. Man unterteilt das Signal in kleine Abschnitte und kann dadurch das Integral durch eine Summe ersetzen. Durch die Abtastung des kontinuierlichen Signals erhält man das zeitdiskrete Signal. Die folgende Gleichung gibt die Fourier-Transformation für zeitdiskrete Signale an.

$$F(\omega) = \sum_{n=-\infty}^{\infty} f(nT) \exp\{-i\omega nT\} \quad \text{Gl. 3-7}$$

Da man eine unendliche Summe nicht diskret ausrechnen kann, kann Gleichung 3-7 vereinfachen, indem die Grenzen der Summation auf eine endliche Zeit beschränkt werden. Man setzt die Zahl  $K$  der zu berechnenden Spektralwerte  $F(k)$  gleich der Zahl  $N$  der verwendeten Abtastwerte  $f(nT)$  und setzt  $T=1$ , dann erhält man Gleichung 3-8. [21]

$$F(k) = \sum_{n=0}^{N-1} f(n) \exp\left\{-i \frac{2\pi n k}{N}\right\} \quad \text{Gl. 3-8}$$

$$k=0,1,2,\dots, N-1$$

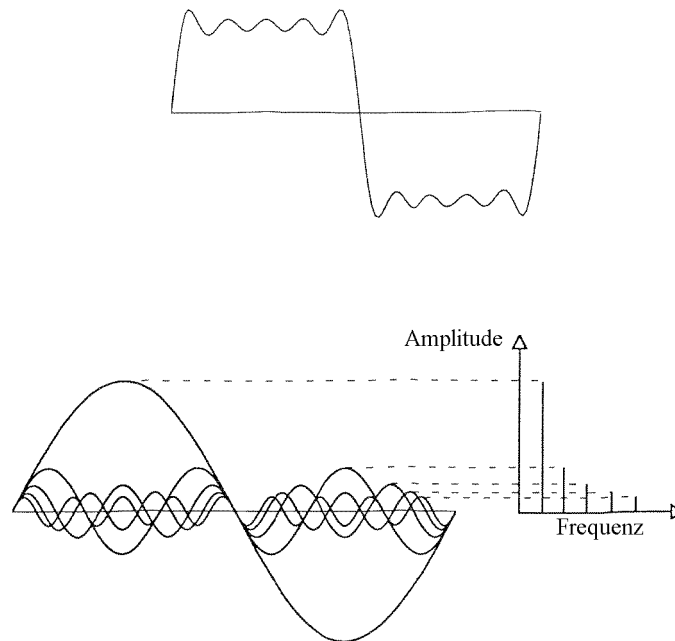
Gleichung 3-8 ist eine Definition der Fourier-Transformation diskreter Signale endlicher Dauer oder kurz gesagt die Diskrete-Fourier-Transformation (DFT).

Die inverse DFT (IDFT) geht, genauso wie die inverse Fourier-Transformation aus der Fourier-Transformation, aus der DFT hervor. Gleichung 3-9 zeigt diesen Zusammenhang.

$$f(n) = \frac{1}{N} \sum_{k=0}^{N-1} F(k) \exp\left\{i \frac{2\pi n k}{N}\right\} \quad \text{Gl. 3-9}$$

Gleichung 3-8 und 3-9 unterscheiden sich nur durch den Vorfaktor  $\frac{1}{N}$ .

Zur Verdeutlichung der mathematischen Berechnungen und des dazugehörigen Frequenzspektrums dient die folgende Abbildung 3-6.



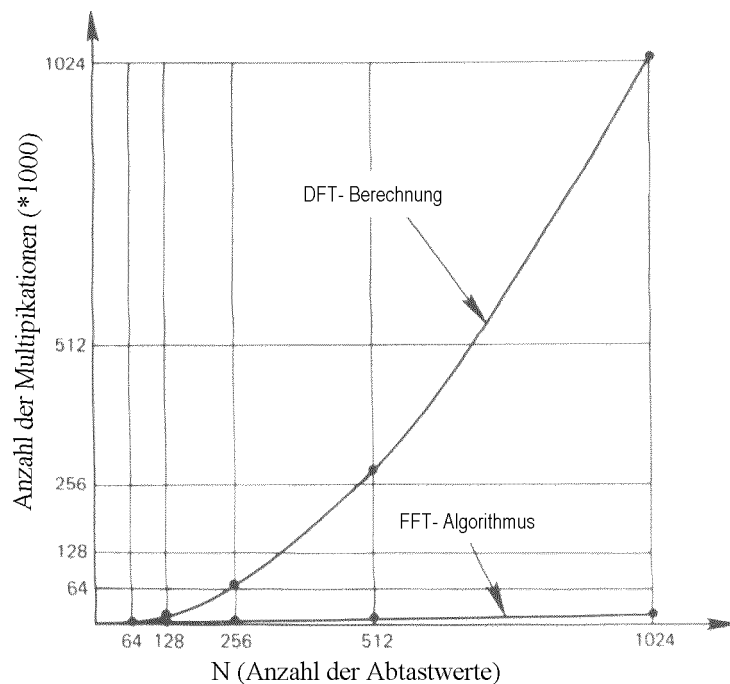
**Abbildung 3-6** Rechteckschwingung, deren Zerlegung in eine Summe von harmonischen Schwingungen und das dazugehörige Frequenzspektrum [1]

Die Amplituden der einzelnen periodischen Teilschwingungen der Rechteckschwingung, in Abbildung 3-6 oben, werden im Frequenz-Amplituden-Diagramm durch einen Punkt dargestellt. Von ihm aus fällt man das Lot auf die Frequenzachse und stellt diese als deutliche Linie dar. Diese Linien stellen durch ihre Lage und Länge die Frequenz und die Amplitude der Teilschwingungen dar. [9]

Neben der Diskreten-Fourier-Transformation gibt es auch die schnelle Fourier-Transformation (**Fast-Fourier-Transformation**). Sie ist durch die Verringerung der Rechenoperation wesentlich schneller als die DFT. Die FFT, entwickelt mit dem Cooley-Turkey-Algorithmus, zerlegt die DFT in zwei Berechnungen der Länge  $N' = N/2$  und überlagert die gewonnenen Teilergebnisse. Die Reduktion beläuft sich auf den Faktor zwei. Nun kann die Zerlegung der DFT-Berechnungen weiter fortgesetzt werden, solange  $N'$  durch 2 teilbar ist. Durch die Zerlegung der DFT erreicht man, daß die Gesamtzahl der nötigen arithmetischen Prozeduren nun nicht mehr quadratisch, sondern annähernd logarithmisch mit der Zahl der Datenpunkte wächst. Allerdings ist für dieses Verfahren notwendig, daß die Zahl  $N$  der Stützpunkte eine Potenz von zwei ist. [21,22]



Abbildung 3-7 veranschaulicht die Relation der Multiplikationen für den FFT-Algorithmus und den Berechnungsmethoden der DFT.



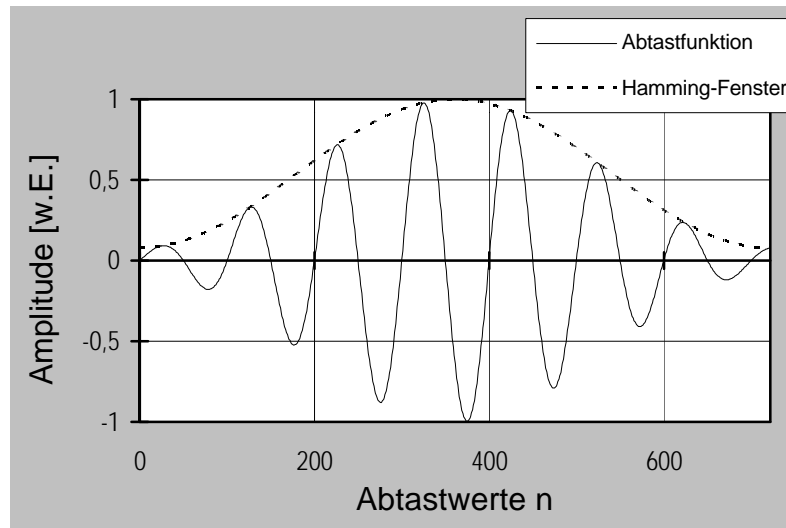
**Abbildung 3-7** Vergleich der Multiplikationen für die Berechnung der DFT und der FFT nach [22]

Da die FFT genau  $2^n$  Werte bearbeitet, schneidet sie Periodenabfolgen genau bei  $2^n$  Werten ab. Dies bringt einen großen Leckeffekt mit sich.[22] Die Fourier-Transformierte einer Sinusfunktion tritt nicht mehr als scharfe Spitze (Deltafunktion) auf, sondern wird durch weitere Frequenzkomponenten verbreitert. Das hat zur Folge, daß an allen diskreten Frequenzen der diskreten Transformierten die Frequenzkomponenten ungleich Null sind und somit die Spitzen ziemlich breit werden. Dieser Effekt wird stärker, wenn die Perioden abgeschnitten werden.

Der Leckeffekt tritt ebenfalls bei der Berechnung der Diskreten-Fourier-Transformation auf, doch kann man diesen gering halten, indem man die Beobachtungszeit gleich einem Vielfachen der Periode setzt. Um den Leckeffekt noch weiter abzuschwächen, benutzt man Fensterfunktionen. Eine ist das Hamming-Fenster. Nach Sickert [21] ist diese Fensterfunktion folgendermaßen definiert:

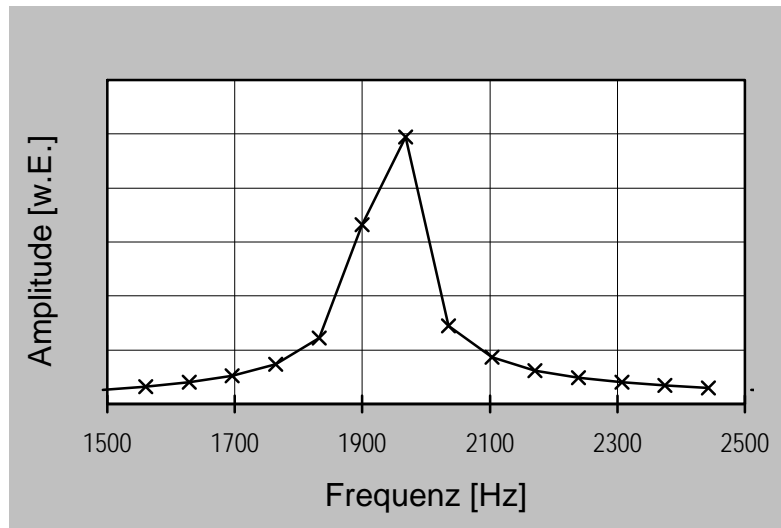
$$w(n) = 0,54 - 0,46 \cos\left(\frac{2\pi n}{N-1}\right) \quad \text{Gl. 3-10}$$

Multipliziert man die Fensterfunktion  $w(n)$  mit der Abtastfunktion  $f(n)$ , so vermindert man den Leckeffekt und die Amplitudenschwankungen werden verringert. Abbildung 3-8 zeigt das Hamming-Fenster als Einhüllende einer sinusförmigen Abtastfunktion.

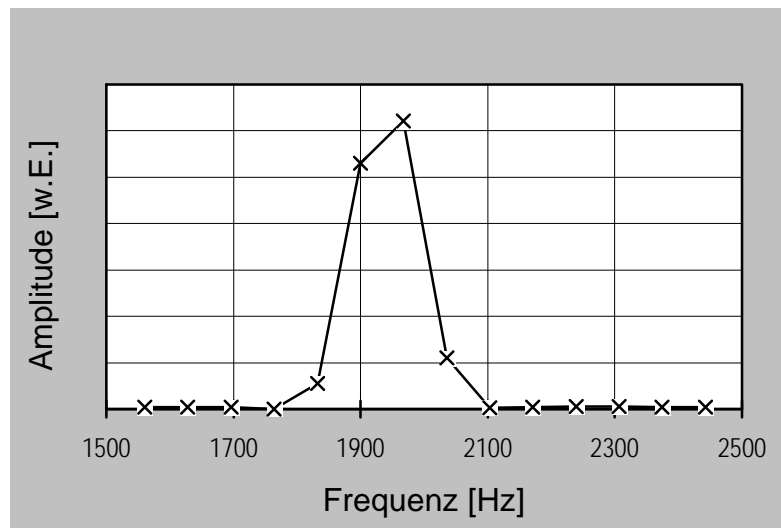


**Abbildung 3-8** Sinusfunktion überlagert mit Hamming-Fenster

Der Leckeffekt wird besonders deutlich, wenn man eine Sinusfunktion, die ungleich einem Vielfachen der Periode ist, einmal mit und einmal ohne Hamming-Fenster bearbeitet. Abbildung 3-9 und Abbildung 3-10 zeigen den Unterschied:



**Abbildung 3-9** DFT einer Sinusfunktion ohne Hamming-Fenster überlagert



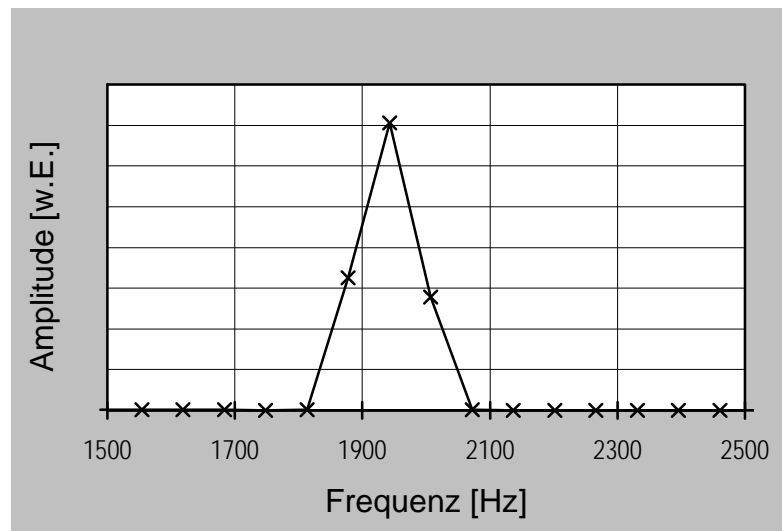
**Abbildung 3-10** DFT einer Sinusfunktion mit Hamming-Fenster überlagert

Die Spitze in Abbildung 3-9 hat noch einen recht breiten Fuß. Aufgrund des Hamming-Fensters ist in Abbildung 3-10 dieser nicht mehr zu sehen ist.

Um die zuvor beschriebenen Effekte zu minimieren, wurde die Diskrete-Fourier-Transformation benutzt und mit dem Hamming-Fenster gearbeitet. Es erwies sich als sinnvoll, für die zu analysierende Abtastfunktion eine

ganzzahlige Periodenanzahl (30-40 Perioden) zu nehmen, um den Fehler gering zu halten.

Abbildung 3-11 zeigt das Optimum einer Diskreten-Fourier-Transformation, die mit einem Hamming-Fenster überlagert ist. Der Fuß ist so sehr schmal geworden. Die Abtastfunktion, die analysiert wurde, war dabei genau 30 Perioden lang.



**Abbildung 3-11** DFT einer Sinusfunktion mit ganzzahliger Periodenlänge und Hamming-Fenster

## 4 Verwendete Programme

In diesem Kapitel werden die in dieser Arbeit verwendeten Programme vorgestellt, die die in Kapitel 3 eingeführten Analyseverfahren mit gemessenen Daten durchführen können.

Das wohl wichtigste Programm, um die Analyseverfahren überhaupt anwenden zu können, ist das **Aufnahmeprogramm**. Dieses Programm ermöglicht es, Schallformen aufzunehmen. Ein weiteres Programm ist **WAVcalc**. Es enthält die Lorenzabbildung, die Autokorrelationsfunktion und Konvertierungsprogramme. Anschließend wird das Programm **SimpleFFT** vorgestellt, welches die Fouriertransformierten berechnet. Zum Schluß wird noch auf das **Sonographie**-Programm eingegangen.

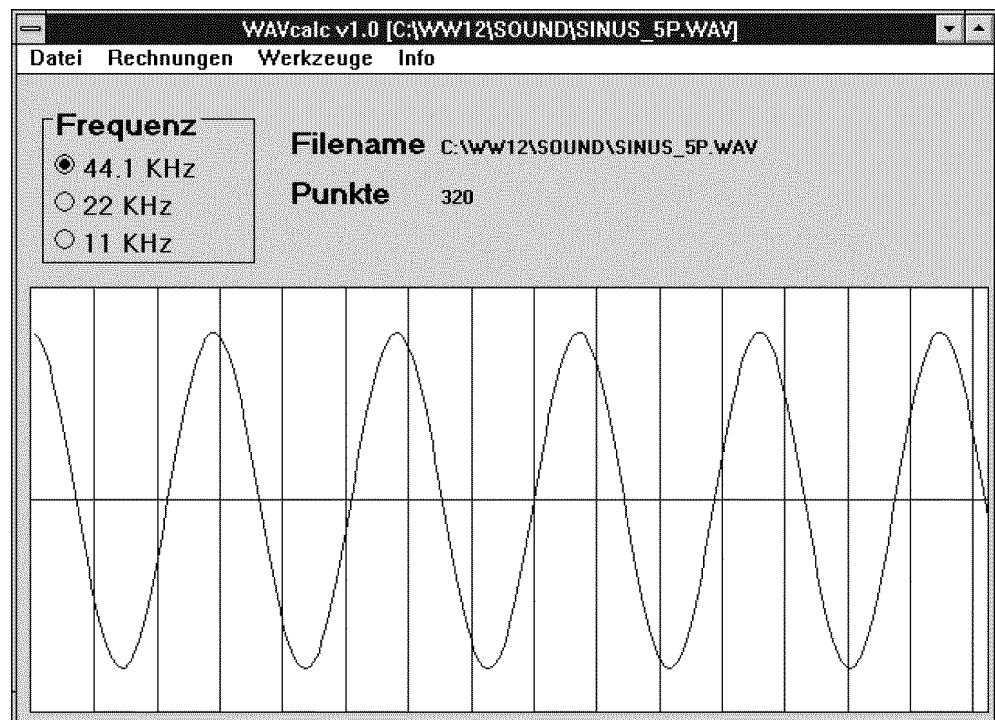
### 4.1 Aufnahmeprogramm

Das Aufnahmeprogramm dient zur Steuerung einer „Soundkarte“, zur Aufnahme von Audiosignalen. In diesem Programm können Audiodateien (Sounddateien) in der Art verändert werden, daß man die Audiodatei schneidet oder Stücke einfügt. Weiterhin gibt es viele Möglichkeiten der Veränderung einer Sounddatei, die hier nicht weiter erörtert werden. Infolge der Masse der Aufnahmeprogramme, die alle unterschiedlich aussehen, wird auf die Funktionsweise nicht weiter eingehen. Bei fast jeder Soundkarte werden derartige Programme mitgeliefert, deswegen wird auf jedes individuell mitgelieferte Aufnahmeprogramm verwiesen. Wichtig ist nur, daß das Programm Dateien im „WAV“-Format speichern kann.

## 4.2 WAVcalc

Am Anfang dieser Arbeit stellte sich das Problem, daß es kein Programm unter Windows gab, das WAV-Dateien einlesen konnte, um daraus die Lorenzabbildung oder die Autokorrelationsfunktion zu errechnen. Zu diesem Zweck mußte also ein Computerprogramm geschrieben werden, das diese beiden Analyseverfahren einschließt. Das von mir unter Visual Basic 4.0 Professional geschriebene Computerprogramm heißt WAVcalc Version1.0.

Zum Verständnis des Programms werden in den nächsten Abschnitten die wichtigsten Bedienschritte erklärt. Wenn eine WAV-Datei geöffnet wird (Menüpunkt: Datei öffnen), erscheint auf der weißen Fläche ein Bild der ausgewählten Audiodatei.



**Abbildung 4-1** Graphische Darstellung der ersten Form (WAVcalc)

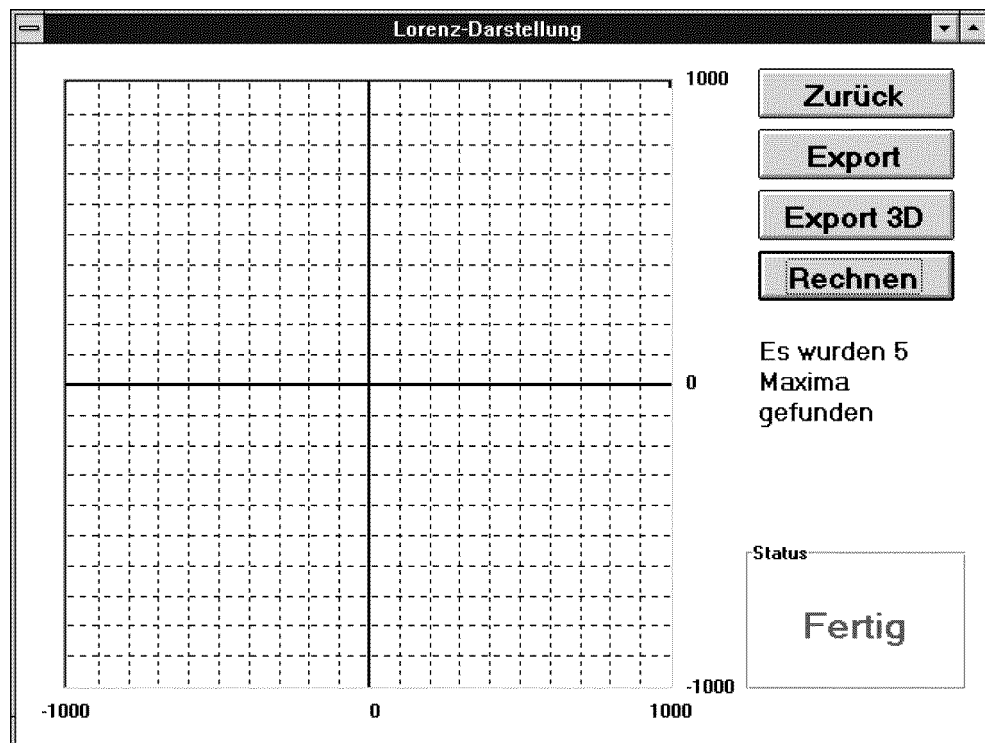
Um sicher zu sein, daß die richtige Datei geladen ist, erscheint noch einmal der Name der Datei unter „Filename“ und die Anzahl der Stützpunkte. Jetzt ist es

möglich, die Datei weiter zu bearbeiten. Abbildung 4-1 zeigt die erste Form<sup>1</sup> mit geladener Datei:

In der Menü-Leiste findet man unter der Option „Rechnungen“ die Lorenzabbildung und die Autokorrelationsfunktion. Außerdem beinhaltet diese Option die Möglichkeit die geladenen Dateien zu invertieren und zu normieren, so daß das Integral oberhalb und unterhalb der Nulllinie gleich ist.

#### 4.2.1 Die Lorenzabbildung als Computerprogramm

Wählt man die Lorenzabbildung aus, so öffnet sich eine neue Form. In Abbildung 4-2 kann man die geladene Datei als Lorenzabbildung darstellen, indem man den Knopf „Rechnen“ drückt.



**Abbildung 4-2** Lorenzabbildung der Sinusfunktion aus Abbildung 4-1

Das Statusfeld zeigt an, wie weit der Computer mit seiner Rechnung ist. Wenn die Rechnung abgeschlossen ist, wird oberhalb des Statusfeldes die Anzahl der gefundenen Maxima dargestellt. Abbildung 4-2 ist die berechnete Lorenzabbildung zu der in Abbildung 4-1 dargestellten Abtastfunktion. Die Maxima befinden sich alle oben rechts in der Ecke, neben der Zahl 1000.

<sup>1</sup> Mit Form bezeichnet man unter Visual Basic ein Fenster, wie es z.B. Abbildung 4-1 zeigt.

Möchte man das gewonnene Ergebnis exportieren, so hat man zwei Möglichkeiten. Einmal können die gewonnenen Werte „normal“ exportiert werden („Export“), d.h. die Werte werden als Text-Datei (.txt) abgespeichert und können als zweidimensionale Darstellung weiterverarbeitet werden. Zusätzlich bietet sich noch die Möglichkeit die Werte dreidimensional zu exportieren („Export 3D“). Die dreidimensionale Darstellung wird dadurch erreicht, daß über das zweidimensionale Bild ein Raster gelegt wird und die Maxima, die sich in jedem Feld befinden, aufsummiert werden. Abgespeichert wird diese Datei auch als Text-Datei (.txt). Zum Verlassen der Lorenzabbildung drückt man den „Zurück“-Knopf.

Die folgenden Zeilen beschreiben die Umsetzung der mathematischen Formulierungen in dieses Programm. Die aufgeführten Zeilen stellen den wichtigsten Ausschnitt des Programms dar. Zur genauen Programmierung von Visual Basic verweise ich auf die gängigen Handbücher [23,24]. Aus Gründen der Überschaubarkeit werden die kommenden Programm-Listings durchnummeriert.

```

1  For i=3 To anz-3
2      If lanz<10000 And sound(i-3)<sound(i)...
        ...And sound(i+3)<sound(i)...
        ...And sound(i-2)<=sound(i-1)...
        ...And sound(i+2)<=sound(i+1)...
        ...And sound(i+1)<sound(i)...
        ...And sound(i-1) < sound(i) Then
3          lanz = lanz + 1
4          lorenz(lanz) = sound(i)
5      End If
6  Next i

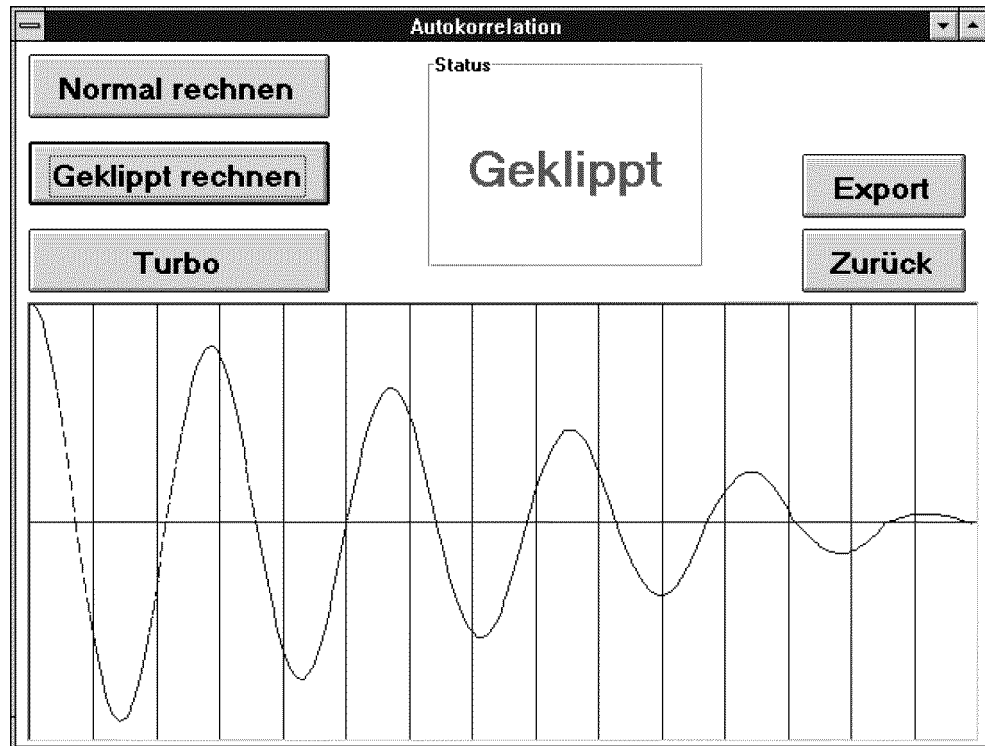
```

Mit der ersten For-Next-Schleife (Zeile 1/6) werden alle Werte, außer den Randwerten, bearbeitet. In Zeile 2 werden die lokalen Maxima gesucht. Das lokale Maximum ist dann ermittelt, wenn es größer ist als drei Werte links und rechts von ihm. Diese Definition des lokalen Maximums erwies sich als sehr sinnvoll, da die bearbeiteten Funktionen auch Rauschen beinhalten und so die Maxima herausgefiltert werden können, die nicht durch das Rauschen verursacht werden. Wenn ein lokales Maximum gefunden ist, wird die Anzahl der gefundenen Extrema `lanz` um eins erhöht (Zeile 3) und die Höhe des Maximums in `lorenz(lanz)` gespeichert.

#### 4.2.2 Die Autokorrelationsfunktion als Computerprogramm

Sucht man sich unter der Option „Rechnungen“ die Autokorrelationsfunktion aus, erscheint auf dem Bildschirm eine neue Form (Abbildung 4-3).





**Abbildung 4-3** Autokorrelationsfunktion der Sinusfunktion aus Abbildung 4-1

Diese Form ermöglicht es die Autokorrelationsfunktion auf drei verschiedene Arten zu lösen:

1. Sie kann normal gerechnet werden.  $\Rightarrow$  (Normal rechnen)
2. Sie kann geklippt gerechnet werden.  $\Rightarrow$  (Geklippt rechnen)
3. Sie kann im Turbo-Modus berechnet werden.  $\Rightarrow$  (Turbo)

Durch Drücken des jeweiligen Knopfs wird die gewünschte Rechnung durchgeführt und in der Form als Bild dargestellt. Die Rechnung kann dann als Text-Datei abgespeichert werden, indem der „Export“-Knopf gedrückt wird. Verlassen kann man das Programm durch den „Zurück“-Knopf und gelangt so wieder zur Ausgangsform (Abbildung 4-1).

Im weiteren werden die einzelnen Programmoperationen aufgelistet und knapp erläutert, denn gerade die Rechnung im Turbo-Modus bedarf einiger Erklärungen. Die **normale** Berechnung der Autokorrelationsfunktion wird in den nachfolgenden Zeilen dargestellt.

```

1  For i=1 To anz
2    hilfe(i)=sound(i)/f
3  Next i
4  For l=1 To anz-1
5    autokor(l)=0

```

```

6   For i=1 To anz-l-1
7       autokor(l)=autokor(l)+CLng(hilfe(i))*CLng(hilfe(i+1))
8   Next i
9   Next l

```

In der ersten For-Next-Schleife (Zeile 1-3) werden alle Werte `sound(i)` durch eine Konstante `f` geteilt und in `hilfe(i)` gespeichert. Die zweite For-Next-Schleife berechnet alle Werte  $R(l)$ . In Zeile 5 werden alle Werte `autokor`, die vorher gerechnet wurden, gelöscht. Die Berechnung der Autokorrelationsfunktion wird in Zeile 7 ausgeführt. Die Schleife von Zeile 4 bis 9 zählt alle Werte für `l`, während die Schleife in Zeile 6 und 8 alle `i` für ein bestimmtes `l` zählt.

Die **geklippte** Autokorrelationsfunktion unterscheidet sich von der normalen Autokorrelationsfunktion durch die in Kapitel 3.2 beschriebenen Merkmale. Hierzu das folgende Visual Basic Programm:

```

1   For i=1 To anz
2       hilfe(i)=Sgn(sound(i))+(Sgn(sound(i))=0)
3   Next i
4   For l=1 To anz-1
5       autokor(l) = 0
6       For i = 1 To anz - l - 1
7           autokor(l) = autokor(l) + sound(i) * hilfe(i + l)
8       Next i
9   Next l

```

Die erste For-Next-Schleife berechnet die Vorzeichenfunktion `sgn(sound(i))`, die in `hilfe(i)` gespeichert wird. Sonst unterscheidet sich die Programmierung der geklippten Autokorrelationsfunktion nicht von der vorhergehenden Version. Nur in Zeile 7 wird das Produkt der Funktion mit der Vorzeichenfunktion gebildet.

Schließlich muß noch die Berechnung der „**Turbo**“ Autokorrelationsfunktion beschrieben werden. Diese Art der Rechnung bringt einen beträchtlichen zeitlichen Vorteil mit sich, wenn die Funktion relativ wenige Vorzeichenwechsel hat.

```

1   For i=2 To anz
2       f = Sgn(sound(i))+(Sgn(sound(i))=0)
3       If hilfe(2,hanz)<>f Then
4           hanz=hanz+1
5       hilfe(1,hanz)=i
6       hilfe(2,hanz)=f
7       End If
8   Next i
9   autokor(1) = 0
10  For i = 1 To anz - 1 - 1
11      autokor(1)=autokor(1)+sound(i)*(Sgn(sound(i+1))
12          +(Sgn(sound(i+1))=0))
13  Next i

```

```

    start = 1
15 wert=autokor(1)
    autokor(1)=1000*Sgn(autokor(1))
    max=Abs(autokor(1))
    For l=2 To anz-1
        For i=start To hanz
20     If hilfe(1,i)-1>1 Then
            f=hilfe(2,i)
            wert=wert+2*CLng(sound(hilfe(1,i)-1))*f
        Else
            If i>2 Then
25         start=i
            End If
        End If
    Next i
    wert=wert-f*sound(anz-1)
30 autokor(1)=CLng(wert)/CLng(max)*1000
Next l

```

Die Berechnung erfolgt in zwei Schritten. Im ersten Schritt werden die Vorzeichenwechsel ermittelt und gespeichert. Ob ein Vorzeichenwechsel von  $i-1$  nach  $i$  stattgefunden hat oder nicht, wird in Zeile 3 entschieden. `hilfe(2,hanz)` ist das alte und `f` das neue Vorzeichen. Falls ein Vorzeichenwechsel stattgefunden hat, wird die Zahl der Vorzeichenwechsel `hanz` um eins erhöht (Zeile 4). Als nächstes merkt sich das Programm die Position des Vorzeichenwechsels (Zeile 5) und das neue Vorzeichen (Zeile 6).

Der zweite Programmschritt dient zur eigentlichen Rechnung. Der erste Wert der Autokorrelationsfunktion `autokor(1)` wird vorab gelöscht und im Anschluß daran neu berechnet (Zeile 10-13), genau wie bei der geklippten Autokorrelationsfunktion. Bei `start = 1` beginnt das Programm beim ersten gefundenen Vorzeichenwechsel. Es merkt sich den ersten Wert (Zeile 15) und normiert ihn (nächste Zeile). Der erste Wert ist stets das Maximum, weshalb dieser als normierender Faktor dient. Der Betrag des Maximalwerts wird in Zeile 17 in der Variable `max` gespeichert. Die nächste For-Next-Schleife (Zeile 18/31) berechnet alle Werte der geklippten Autokorrelationsfunktion  $r(l)$ . Zeile 19 ordnet an, alle gefundenen Vorzeichenwechsel abzuarbeiten. Alle Werte werden im folgenden um die Änderung am  $i$ -ten Vorzeichenwechsel korrigiert (Zeile 20-22). Darauf wird in den Zeilen 24-26 der erste zu beachtende Vorzeichenwechsel verschoben. Wenn diese Prozeduren abgelaufen sind, werden die Werte um die neue Summationsgrenze berichtigt. Dieses geschieht in Zeile 29. Die neuen Werte werden als `autokor(1)` gespeichert, zuvor werden sie jedoch normiert (Zeile 30).

Diese Kurzanleitungen sollen einen Überblick über die einzelnen Programme geben. Die Listings beinhalten nur die mathematischen Formulierungen und

nicht die übrigen Codes. Die lauffähigen Programme befinden sich auf der beigelegten Diskette.

Die Autokorrelationsfunktion wurde dafür verwendet, um unter der Vielzahl der Aufnahmen die Besten herauszusuchen. Durch Lautstärkeschwankungen in der Stimme und leichte Unterschiede in der Frequenz des Schalls weicht die Autokorrelationsfunktion vom Idealfall des linearen Abfalls ab. Eine derartige Klangprobe eignet sich nicht unbedingt zur Analyse, da man davon ausgehen muß, daß sich der Stimmapparat beim Sprechen verändert hat. Folglich kann die Güte der Aufnahmen durch den Verlauf der Autokorrelationsfunktion beurteilt und schlechte Aufnahmen ausgesondert werden.

### 4.3 Simple FFT

Für die Analyseverfahren fehlte noch ein Programm, das die Fourier-Transformierten berechnen kann. Durch das Internet fand sich das folgende Programm an der Universität Paderborn. Unter der folgenden Internet-Adresse findet man das Programm Simple FFT (sFFT):

**<ftp://ftp.uni-paderborn.de>**

**Verzeichnis: pub/simtel**

Dieses Programm ist für den freien Gebrauch zugänglich (Freeware). Man muß sich das Programm also nicht lizenzieren lassen.

Da in der nächsten Zeit viele Schulen in NRW vernetzt werden sollen, ist der Weg über das Internet ein geeigneter Weg, um freie Computerprogramme zu beziehen. Eine weitere interessante Internet-Adresse ist die nachstehende:

**<http://www.winsite.com>**

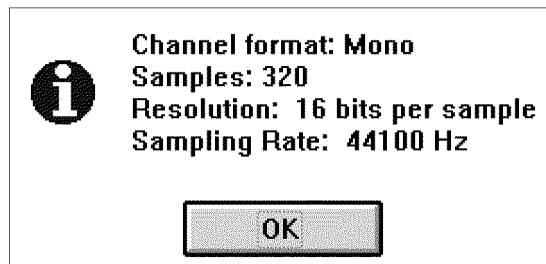
SimpleFFT v2.1 ist ein Fourier-Analyseprogramm, das verschiedene Möglichkeiten bietet, die geladene Datei zu bearbeiten. Das in englischer Sprache verfaßte Programm kann drei verschiedene Formatarten laden. ASCII-, Binär- und WAV-Formate können gelesen werden. Wie die einzelnen Formate aussehen sollten, damit sie mit sFFT bearbeitet werden können, soll durch die folgenden Kurzbeschreibungen erklärt werden:

- Das ASCII-Format sollte wie folgt aussehen:

**Realanteil** (Zahl) [**Leerzeichen**] **Imaginäranteil** (Zahl), wobei das Leerzeichen den Real- vom Imaginärteil trennt.

- Binär-Dateien können bearbeitet werden, wenn jeder Datenpunkt vier Byte des Real- und vier Byte des Imaginäranteils besitzt.
- WAV-Dateien können sofort gelesen werden. Sie müssen aber Mono aufgenommen worden sein. Überschreitet die Datei mehr als 32768 Punkte, kann sie nicht mehr bearbeitet werden.

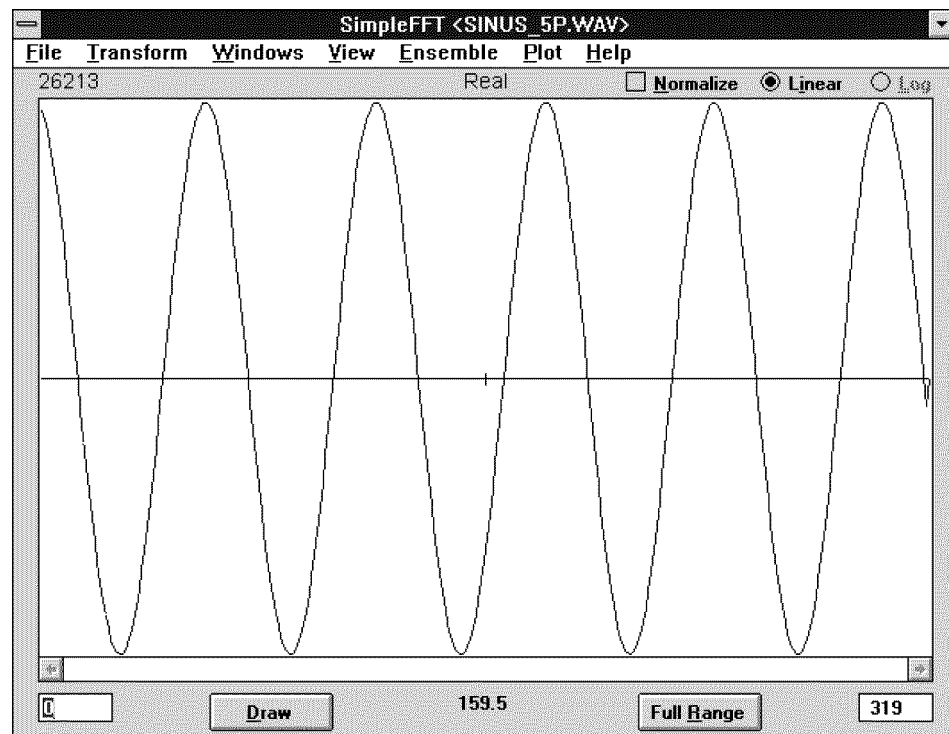
Ist eine Datei geladen worden, zeigt eine kleine Form an, woraus die Datei besteht. Die Form beinhaltet die Anzahl der Datenpunkte der geladenen Datei, das Aufnahmeformat und zusätzliche Daten, die in Abbildung 4-4 dargestellt sind.



**Abbildung 4-4** Darstellung des Aufnahmemodus

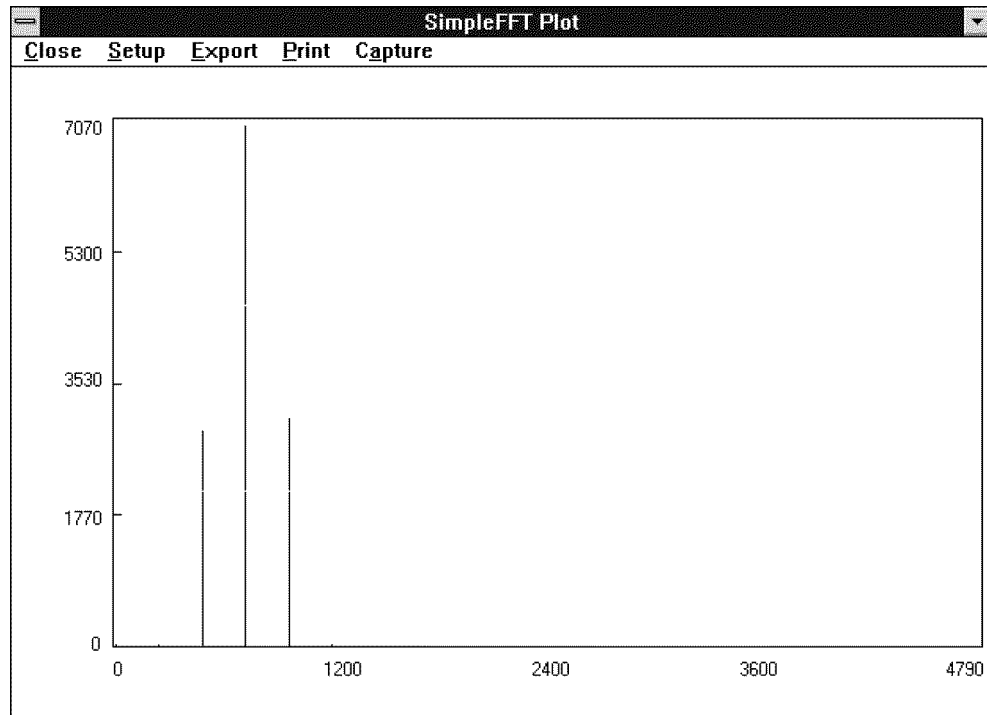
Nach Bestätigen dieser Form öffnet sich ein weiteres kleines Fenster, in dem man darauf hingewiesen wird, daß die FFT nur eine Zahl von Werten bearbeiten kann, die einer Potenz von zwei entspricht (siehe Kapitel 3.3). Aus diesem Grund sollte die Zahl der Datenpunkte auf die nächst kleinere Zweierpotenz gekürzt werden oder so viele Nullen hinzugefügt werden, daß die nächst höhere Zweierpotenz erreicht wird. Ein zusätzlicher Knopf bietet die Möglichkeit, die Datei nicht zu verändern und daher nur die DFT zu verwenden. Nachdem man sich für ein Verfahren entschieden hat, erscheint Abbildung 4-5.

Unter der Option „Windows“ kann man sich eine Fensterfunktion aussuchen mit der man die Funktion überlagern möchte. Es ist möglich, zwischen dem Hamming-Fenster und dem Blackman-Fenster auszuwählen. Das Blackman-Fenster setzt im Gegensatz zum Hamming-Fenster (Kapitel 3.3) die Randwerte genau gleich Null. Danach kann die Funktion transformiert werden („Transform“). Während die FFT in wenigen Sekunden gerechnet ist, benötigt die DFT erfahrungsgemäß eine bis zwei Stunden Rechenzeit, je nach Anzahl der Stützpunkte. Ist die Transformation beendet, so erhält man durch Betätigen des „Draw“-Knopfs, den Realanteil der Funktion.



**Abbildung 4-5** sFFT-Form mit geladener Sinusfunktion

Möchte man das Spektrum darstellen, so findet man in der Menü-Leiste unter „View“ die verschiedenen Darstellungsarten. Neben dem Spektrum, der Magnitude (Einhüllende), der Real- und Imaginäranzeige der Daten ist es ebenfalls möglich, die Daten als relative Phase von Real- und Imaginäranteil darzustellen. Die ausgesuchte Darstellungsform wird auf dem weißen Feld als Bild dargestellt. Zur Änderung des gezeigten Bereichs modifiziert man die Zahlen, die sich rechts und links unten in den beiden Kästchen befinden. Unter „Plot-Display“ öffnet sich eine neue Form (Abbildung 4-6).



**Abbildung 4-6** Fourierspektrum der in Abbildung 4-5 gezeigten Sinusfunktion

Um die Achsen zu beschriften, sucht man sich das Feld „Setup“ und erhält Abbildung 4-7.

The figure shows a "Setup" dialog box with three main sections. The first section, "Sample Frequency (1: no conversion)", has a text box containing "44100". The second section, "Scale x axis as", has two radio buttons: "Time" (unselected) and "Frequency" (selected). The third section, "Labels", has a "Plot Title:" label followed by a text box, and "X axis:" and "Y axis:" labels followed by text boxes. The fourth section, "3 Sig. Fig.", has two checkboxes: "x axis" (checked) and "y axis" (checked). An "Ok" button is at the bottom right.

**Abbildung 4-7** Form für die Beschriftung der Bilder

In dieser Form kann man

- die x-Achse entweder als Zeit- oder Frequenzachse darstellen.
- dem Bild einen Titel geben und die Achsen beschriften.
- mit „3 Sig. Fig.“ die Zahlen der jeweiligen Achsen auf drei signifikante Stellen runden.

Für die Weiterbearbeitung können die Werte aus Abbildung 4-6 auf zwei Arten exportiert werden. Zum einen unter „Export“ als Text-Datei (.txt), zum anderen als Bitmap-Datei („Capture“). Beenden läßt sich sFFT in gleicher Weise wie andere Windows-Programme.

Die Bedienung dieses Programms erscheint auf den ersten Blick nicht unbedingt leichtverständlich. Der Vorteil dieses Programms ist aber, daß es sich genau wie andere Programme unter Windows bedienen läßt. Falls man doch noch auf Probleme stoßen sollte, schließt dieses Computerprogramm noch eine Hilfe mit ein.

#### **4.4 Spectrogram**

Das Programm Spectrogram v2.3 kann man unter der gleichen Adresse finden, die schon bei sFFT angegeben wurde.

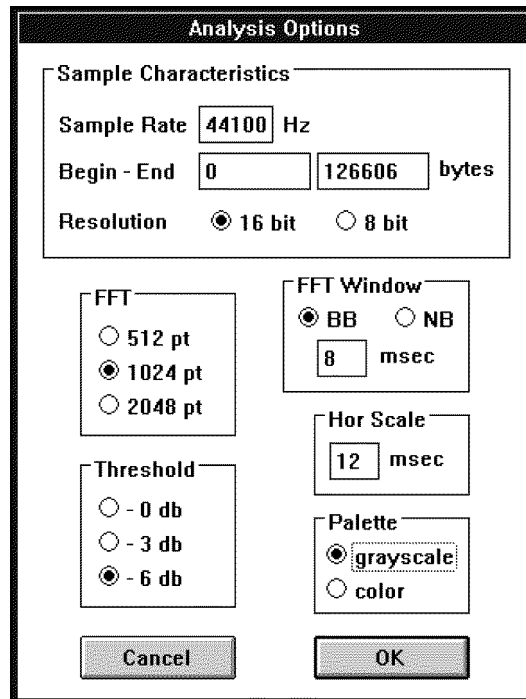
Das in Abschnitt 4.3 beschriebene Programm, das Amplituden-Frequenz-Diagramme berechnet, eignet sich nur für Klänge, die sich zeitlich nicht verändern, also stationär sind. Möchte man jedoch gerade die Änderung des Spektrums (Energieschwerpunkte, Änderung der Formanten) im zeitlichen Verlauf aufzeichnen, so ist ein dreidimensionales Diagramm notwendig, welches neben der Amplitude und der Frequenz auch noch die Zeit berücksichtigt. Faßt man die Frequenz und die Amplitude zusammen, so spricht man von einem Spektrum-Zeit-Diagramm.[9] Bevor es Computer gab, wurden diese Aufnahmen mit einem Sonagraphen<sup>2</sup> gemacht, deshalb heißen die gewonnenen Bilder auch Sonagramme. Heutzutage arbeitet man häufig mit Computerprogrammen, die wesentlich erschwinglicher sind als das Gerät. Eines dieser Programme ist Spectrogram v2.3.

Spectrogram bedient sich der Fast-Fourier-Transformation (Kapitel 3.3), um ein Spektrum-Zeit-Diagramm darzustellen. Das Programm analysiert kleine, aufeinanderfolgende Zeitfenster und stellt die zugehörigen Frequenzspektren nebeneinander graphisch dar. Möchte man eine Audiodatei öffnen, wählt man aus dem „File“- Menü die englische Bezeichnung für Öffnen („Open“). Ist die Audiodatei ausgewählt, erscheint auf dem Bildschirm eine Dialog-Box („Analysis Options“), in der man die Analyse der aufgenommenen Datei noch differenzieren kann. Abbildung 4-8 zeigt die unterschiedlichen Möglichkeiten der Analyse der geladenen Audiodatei.

---

<sup>2</sup> Mit elektrischen Filtern arbeitender Apparat. Er wird zur Analyse von Lauten in Hinblick auf Frequenzen, Intensität und Quantität eingesetzt. Das Ergebnis ist in Form unterschiedlicher Schwärzungen auf einem speziell belichteten Papier ablesbar. [25]





**Abbildung 4-8** Dialog-Box zum Analysieren einer Audiodatei in Spectrogram

Wenn man eine WAV-Datei geöffnet hat, werden die Aufnahmecharakteristika des Originals in „**Sample Characteristics**“ dargestellt. Zu ihnen gehören die Sample-Rate (in Hz), die Lage des Anfangs- und Endpunkts (in bytes) und die Auflösung (8 bit oder 16 bit) der Datei.

Unter „**FFT**“ kann man sich die Anzahl der Datenpunkte auswählen, die man Fourier-Transformieren möchte. Soll eine höhere Auflösung der Analyse erreicht werden, sollte man nicht mit 512, sondern mit 1024 oder 2048 Datenpunkten rechnen. Der Nachteil einer hohen Auflösung ist jedoch, daß die Zeit der Darstellung mit der Menge der Datenpunkte stärker anwächst.

Das FFT-Fenster („**FFT Window**“) beinhaltet zwei Arten von Analysefiltern: ein Schmalband- (NB) und ein Breitbandfilter (BB). Der Schmalbandfilter hat eine gute Frequenzauflösung, hingegen hat der Breitbandfilter eine bessere zeitliche Auflösung. Die zeitliche Auflösung (in [ms]) des Breitbandfilters kann noch variiert werden, indem man die Zahl unterhalb des BB-Knopfs ändert.

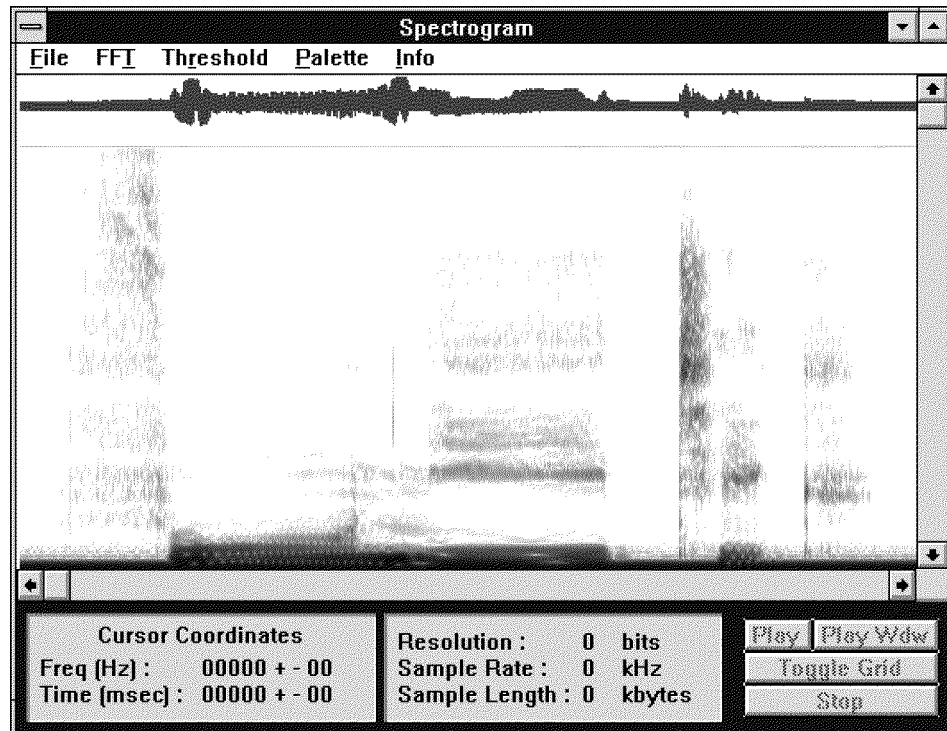
Man kann außerdem die horizontale Achse („**Hor Scale**“) modifizieren. Jede einzelne vertikale Linie im Sonagramm ist das Ergebnis aus einer FFT-Berechnung. Die horizontale Achse kann zwischen 1 und 500 ms variiert werden.

Ferner kann man Hintergrundgeräusche in einer Aufnahme bis zu einer gewissen Schallsschwelle („**Threshold**“) herausfiltern. Diese Schwellen liegen bei -3 dB oder -6 dB. Falls man eine Aufnahme ohne viele Störeffekte hat, sollte man eine Schwelle von 0 dB wählen.

In „**Palette**“ hat man zusätzlich die Wahl zwischen einer farblichen und einer grauen Darstellung des Bildes. Hohe Amplitudenwerte werden mit Rot und niedrige mit Dunkelblau gekennzeichnet. Entscheidet man sich für die Grauskalierung, so nehmen fiktive Amplitudenkoordinaten überall dort ihren Maximalwert an, wo starke Schwärzungen auftreten. Je weiter die Schwärzung abnimmt, desto kleiner wird die Amplitude.

Wenn man alle Optionen ausgewählt und auf „OK“ gedrückt hat, erscheint auf dem Bildschirm die bearbeitete Audiodatei als Bild. Als Ordinate wird die Frequenz und als Abszisse die Zeit aufgetragen. Die Amplitudenachse muß man sich als nach oben gerichtete Koordinate denken. Über dem Sonagramm ist die bearbeitete Abtastfunktion nochmals dargestellt. Abbildung 4-9 stellt ein Breitbandsonagramm des Ausdrucks „Phonetik“ dar, welches mit den gewählten Optionen aus Abbildung 4-8 bearbeitet wurde.

Ein genaues Auslesen der Cursor-Position in Frequenz [Hz] und Zeit [ms] wird durch die Anzeige links unten in Abbildung 4-9 ermöglicht. Durch Drücken des „Toggle Grid“ Knopfs kann man sich zusätzlich ein Koordinaten-Gitter anzeigen lassen.



**Abbildung 4-9** Breitband-Sonogramm des Ausdrucks „Phonetik“

Möchte man nun die Parameter des Sonagramms abändern, kann man mit Hilfe der Menü-Leiste ein Element (FFT, Threshold oder Palette) aussuchen. Mit dem neuen Wert baut sich das Sonogramm wieder auf. Wenn man mehr als eine Komponente korrigieren will, sucht man sich in der Menü-Leiste „File“ die Option „Modify“ aus. Es erscheint eine Dialog-Box, mit deren Hilfe mehrere Bestandteile verändert werden können.

Besitzt man eine Soundkarte, so kann sogar direkt im Programm aufgenommen werden. Sucht man sich im „File“-Menü die Option „Record New“ aus, wird die Aufnahme gestartet. Bevor man allerdings aufnehmen kann, erscheint wieder die Dialog-Box, in der die Parameter für die Analyse ausgesucht werden müssen.

Das Programm besitzt zusätzlich die Möglichkeiten, die Sounddatei abzuspielen und sich die Stelle im Sonogramm anzeigen zu lassen. Voraussetzung dafür ist selbstverständlich eine unter Windows installierte Soundkarte. Wenn der „Play“-Knopf gedrückt wird, wird die gesamte Länge der geladenen WAV-Datei abgespielt, während der „Play Wdw“ nur das auf dem Bildschirm sichtbare Stück der WAV-Datei abspielt.

Zur Weiterverarbeitung ist es möglich, die Audiodatei des Sonagramms als WAV-Datei zu speichern („Save Wave“). Ansonsten kann das Sonogramm als

Bitmap-Datei („Save Bitmap“) gespeichert werden. Falls noch Fragen zur Bedienung des Programms auftreten sollten, liegt als Anlage auf der Diskette noch eine Hilfe-Datei („readmegr.txt“) bei.

## 4.5 Zusammenfassung

Das vorangegangene Kapitel befaßte sich mit den in dieser Arbeit benutzten Programmen. Sie sollen die Möglichkeit bieten, die menschliche Sprache präzise zu analysieren und Aussagen über die Physik der menschlichen Sprache zu treffen.

In der Sprachforschung verwendet man ähnliche Programme, wie sie hier vorgestellt wurden. sFFT stellt den menschlichen Sprachschall in einem Amplituden-Frequenz-Diagramm dar. Diese Art der Illustration ist besonders von Vorteil, wenn man stationäre Klänge betrachten möchte. Charakteristische Änderungen von Klängen sind im Rahmen des sFFT-Programms nicht realisierbar. Zu diesem Zweck sollte das Sonographie-Programm eingesetzt werden. WAVcalc schließt zwei Analyseverfahren ein, die schon in Kapitel 3.1 und 3.2 beschrieben wurden. Zusätzlich findet sich dort ein Konvertierungsprogramm. Dieses Programm war erforderlich, weil sFFT die gerechneten Werte mit Punkten als Dezimaltrennzeichen liefert. Da man in deutschsprachigen Programmen die Werte nur mit Kommata als Trennzeichen bearbeiten kann, war es notwendig ein Konvertierungsprogramm zu schreiben, das Punkte durch Kommata ersetzt und umgekehrt.

Bei der Anwendung der Programme sollte darauf geachtet werden, daß die eingeladenen Dateien die bereits beschriebenen Formate haben, da sie sonst nicht analysiert werden können. Auch sollten nicht zu viele Datenpunkte in einer Datei gespeichert werden, da hierdurch die Rechenzeit erheblich anwächst. Hält man sich allerdings an die Hinweise in den Kurzanleitungen, sollte man keine Probleme haben, selbst Sprachproben zu analysieren. Im Anhang befinden sich daher zwei Disketten mit allen dargestellten Programmen. Der Inhalt und die Nutzung der Disketten wird in Kapitel 11.3 beschrieben.

## 5 Durchführung

Um das Zustandekommen der in den folgenden Kapiteln dargestellten Diagramme aufzuzeigen, wird hier der Weg der aufgenommenen akustischen Signale vom Sprecher bis zum fertigen Diagramm erläutert. Im speziellen wird auf den Meßaufbau eingegangen und das Zusammenwirken der in Kapitel 4 dargestellten Programme zum Ziel der Sprachanalyse darstellt. Damit systematische Fehler durch den speziellen Meßaufbau auszuschließen sind, wird das System anhand einer einfachen Prüfmessung getestet.

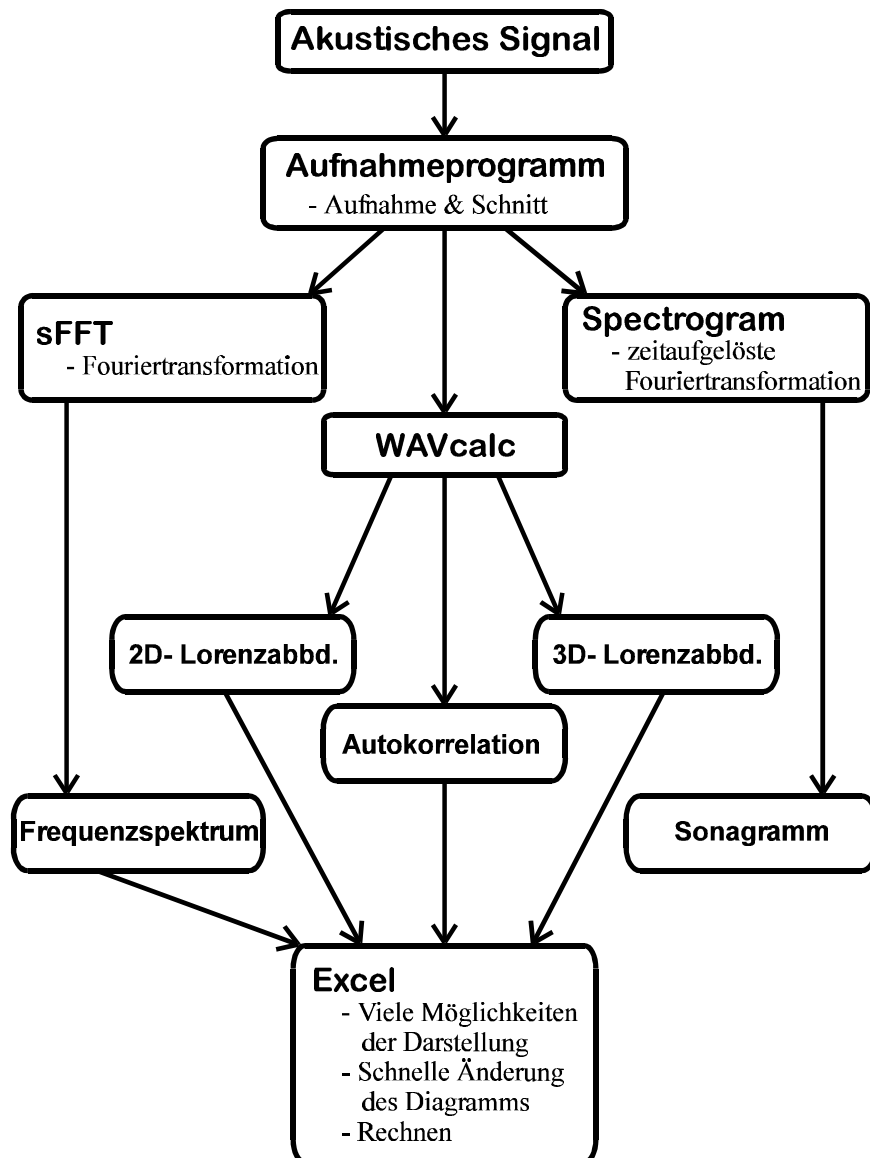
### 5.1 Meßwertaufnahme

Bei der menschlichen Sprache handelt es sich um akustische Signale, die mit Hilfe einer „Soundkarte“ und einem geeigneten Mikrofon digital abgetastet werden können. Die Abtastwerte können dann im Aufnahmeprogramm, wie in Kapitel 4.1 beschrieben, zur Analyse weiterverarbeitet werden.

Die „Soundkarte“ besteht im wesentlichen aus einem Analog-Digital-Wandler, der akustische Signale in digitale diskrete Abtastwerte umsetzt. Für die Aufnahmen wurde eine Soundkarte (Soundblaster 16 ASP) verwendet, die akustische Signale bis zu einer Auflösung von 16 Bit in digitale Werte umwandeln kann. Die Abtastrate der Soundkarte kann von 5 kHz bis 44,1 kHz variiert werden. Alle akustischen Signale wurden mit einer Auflösung von 16 Bit und einer Abtastrate von 44,1 kHz aufgenommen. Da die beschriebenen Programme unter Windows laufen, müssen die Aufnahmen im WAV-Format abgespeichert werden.

## 5.2 Meßabläufe

Das Flußdiagramm in Abbildung 5-1 zeigt die Zusammenhänge zwischen den einzelnen Programmen. Der Meßablauf beginnt beim akustischen Signal des Sprechers, über die in Kapitel 4 vorgestellten Programme bis hin zum endgültigen Diagramm.



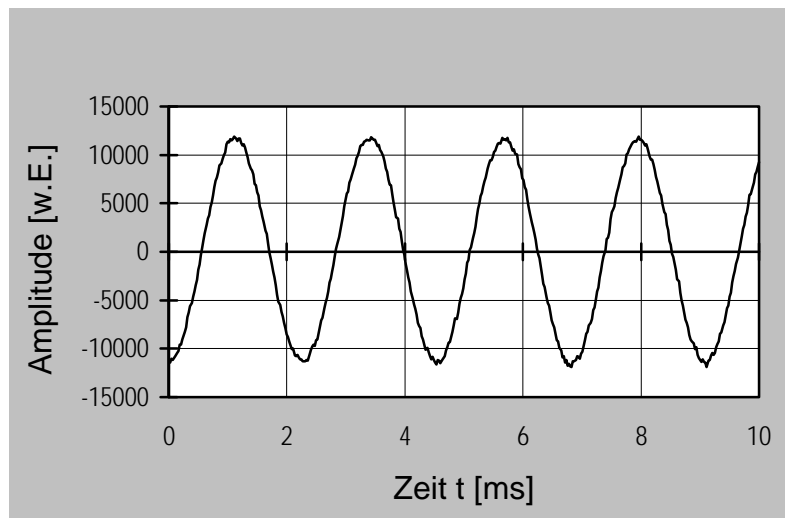
**Abbildung 5-1** Flußdiagramm der Meßabläufe

Man erhält letztendlich fünf verschiedene Diagramme mit fünf unterschiedlichen Aussagen.

### 5.3 Prüfmessung

Mittels einer einfachen Messung soll gezeigt werden, daß der Meßaufbau keinen oder nur einen geringen systematischen Fehler aufweist. Zur Prüfung des Meßaufbaus diene eine Stimmgabel, die eine Frequenz von 440 Hz besitzt. Das Signal wurde mit einem Mikrofon aufgenommen und dann mit den unterschiedlichen Verfahren analysiert. Durch die in Kapitel 3.1 beschriebene Theorie weiß man, was diese Prüfmessung ergeben sollte.

Die Abtastfunktion der Stimmgabel (Abbildung 5-2) zeigt einen fast periodischen und regulären Verlauf. Zur Illustration zeigt Abbildung 5-2 nur einen kleinen Ausschnitt der für die Analyse verwendeten Abtastfunktion.

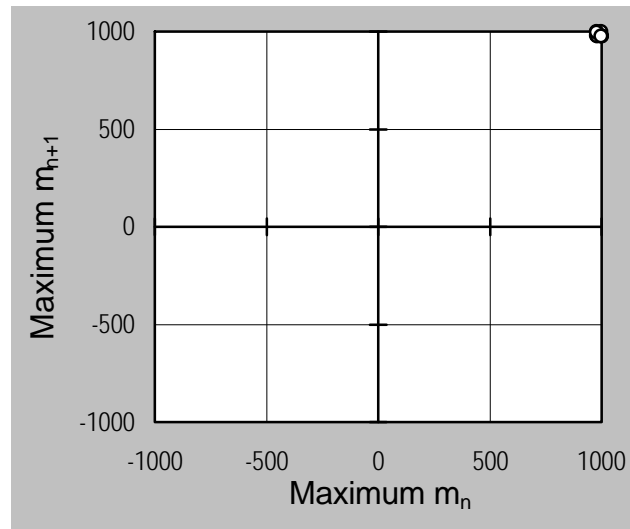


**Abbildung 5-2** Abtastfunktion der Stimmgabel

**Anmerkung:** Die Werte für die Amplitude der Abtastfunktion sind die von der Soundkarte ermittelten digitalen Werte.

Die leichten Störungen an den Maxima/ Minima lassen darauf schließen, daß das Signal durch ein leichtes Hintergrundrauschen verfälscht wird. Weil die Aufnahmen aber nicht in einem schalldichten Raum aufgenommen wurden, ist diese Störung nicht verwunderlich.

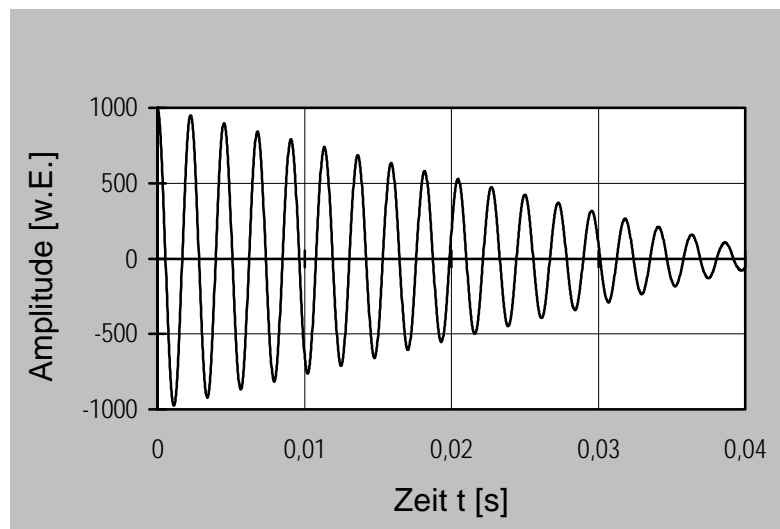
In der Lorenzabbildung (Abbildung 5-3) sieht man einen fast scharfen Punkt, der durch die konstante Amplitude der Abtastfunktion zustande kommt. Bedingt durch das leichte Rauschen in den Maxima ergibt sich kein einzelner Punkt.



**Abbildung 5-3** Lorenzabbildung der Abtastfunktion (Stimmgabel)

**Anmerkung:** Die durch die Lorenzabbildung ermittelten Werte sind so normiert, daß das höchste Maximum der Abtastfunktion den Wert 1000 erhält.

Die Autokorrelationsfunktion in Abbildung 5-4 weist im zeitlichen Verlauf einen linearen Abfall auf. Nur bei optimaler Korrelation, also wenn keine Frequenzschwankung auftritt, ist ein linearer Abfall zu erwarten.



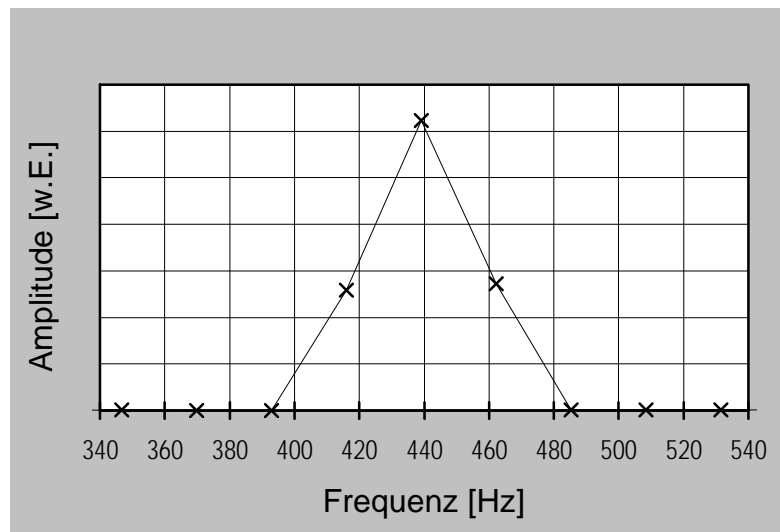
**Abbildung 5-4** Autokorrelationsfunktion der Abtastfunktion (Stimmgabel)

**Anmerkung:** Der erste Wert der Autokorrelationsfunktion ist das Integral über das Quadrat der Abtastfunktion und gleichzeitig stets der höchste Wert der Autokorrelationsfunktion. Die Autokorrelationsfunktion ist so normiert, daß der erste Punkt den Wert 1000 annimmt.

Abbildung 5-5 zeigt einen Ausschnitt aus dem Frequenzspektrum der Stimmgabel, um die Ablesegenauigkeit für das Maximum zu erhöhen. Man



sieht ein scharfes Maximum etwa bei 440 Hz. Auf der Stimmgabel ist eine Frequenz von 440 Hz angegeben.



**Abbildung 5-5** Ausschnitt des Frequenzspektrums (Stimmgabel)

**Anmerkung:** Für die Betrachtung der Frequenzspektren ist nur die Lage der Maxima auf der Frequenzachse und ihre relative Höhe zueinander wichtig. Aus Gründen der Übersichtlichkeit wird auf die Angabe der Zahlenwerte der y-Achse verzichtet.

Die Prüfmessungen zeigen das Verhalten, das man aufgrund der Theorie erwartet. Der systematische Fehler bei der Bestimmung von Frequenzen beläuft sich demnach auf wenige Prozent. Gleichlaufschwankungen der Aufnahme und Schwankungen in der Amplitude können durch die verwendeten Aufnahme-geräte ausgeschlossen werden. Demzufolge kann der Meßaufbau für die nachfolgenden Aufnahmen mit hinreichender Genauigkeit verwendet werden.

## 6 Experimentelle Analyse der Grundschallformen

Der menschliche Stimmapparat ist eines der interessantesten Instrumente zur Tonerzeugung, jedoch auch das am schwierigsten zu analysierende. Um das Verständnis zur allgemeinen Schwingungs- und Wellenlehre zu fördern und auch Schüler zu motivieren sich mit diesem Thema auseinanderzusetzen, wäre

das menschliche Sprechorgan sicherlich ein reizvolles Beispiel. Die Schüler würden sich mit einem Modell auseinandersetzen, das sie jederzeit benutzen, aber keineswegs begreifen.

In diesem Kapitel wird der Begriff der Grundschaallform erläutert. Es werden Kategorisierungen von Schallformen vorgenommen, die durch Beispiele konkretisiert werden.

## **6.1 Was ist eine Grundschaallform?**

Nach Neppert [9] werden in der Sprachlautbeschreibung die Schallformen in unterschiedliche Kategorien aufgeteilt. Man unterscheidet vier verschiedene Grundschaallformen, die „stumme“ Schallform, den Explosionsschall, das Friktionsrauschen und den quasiperiodischen Schall. Weil die Schallformen vom menschlichen Sprechorgan hervorgebracht werden, ist es nicht möglich das Signal eindeutig zu interpretieren, ohne die Beziehung zwischen der Physiologie der Sprechmotorik und dem Schallsignal darzustellen. Die Definition der vier Grundschaallformen stellt eine extreme Simplifizierung dar, doch macht diese Vereinfachung die Schallsignale beschreibbar. Die Schallformen werden mit Hilfe der in Kapitel 3 beschriebenen Analyseverfahren bearbeitet.

## 6.2 Die „stumme“ Schallform

Der Begriff der „stummen“ Schallform ist von der Wortwahl zugegeben etwas widersprüchlich, doch findet sich in der Literatur noch kein besserer Ausdruck für diese Schallerscheinung. „Stumme“ Schallformen sind solche Schallerscheinungen, deren Amplitudenwerte Null oder fast Null sind, d.h. man kann keinen nutzbaren Schall wahrnehmen. Diese Schallform kommt zum Beispiel in der Verschlußphase stimmloser Verschlußkonsonanten [p], [t], [k]<sup>3</sup> vor. Ihr folgt ein Explosionsschall. Der Unterschied zwischen einer „stummen“ Schallform und einer Pause zwischen Wörtern kann akustisch nicht festgestellt werden. Nur die Sprechweise macht einen Unterschied. Ohne sprechphysiologische Daten könnte man ansonsten das Signal kaum erklären. Da diese Schallform nicht auswertbar ist, wird auf diese Schallform nicht weiter eingegangen.

## 6.3 Der Explosionsschall

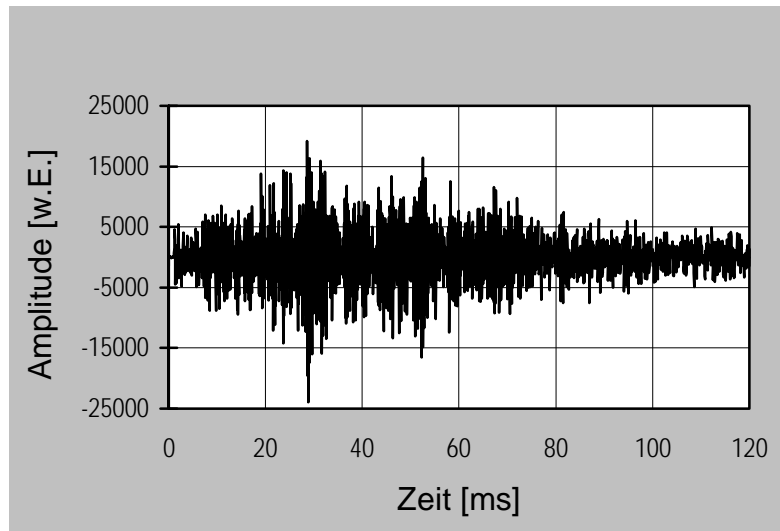
Wenn im Ansatzrohr oder an der Glottis ein Verschluß durch Muskeltätigkeit rasch gelöst wird, entsteht ein Explosionsschall (Explosivlaut)<sup>4</sup>. [25] Er ist durch seine kurze Dauer charakterisiert. Die Grundlage für die Entstehung eines Explosionsschalls ist der Druckunterschied zwischen den beiden Seiten des Verschlusses. Wenn sich der Verschluß gelöst hat, strömt die aufgestaute Luft nach außen.

Abbildung 6-1 zeigt anhand des Explosionsschalls [t] einer weiblichen Testperson die Form der Abtastfunktion.

---

<sup>3</sup> In den Klammern stehen die jeweils gesprochenen Buchstaben. Im folgenden Ablauf wird diese Bezeichnung benutzt.

<sup>4</sup> Auch Sprenglaut genannt. Ein bei oralem Verschluß gebildeter nicht-nasaler Sprachlaut.

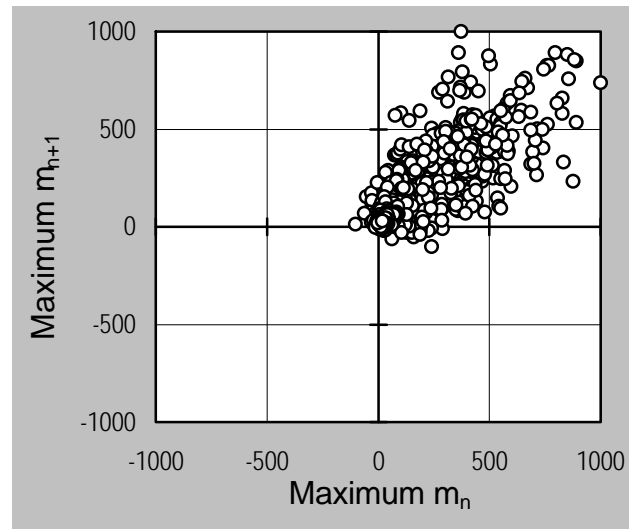


**Abbildung 6-1** Abtastfunktion des Explosionsschalls [t] (weibl.)

Nachdem sich der Verschluß im Ansatzrohr gelöst hat, sieht man in der Abtastfunktion einen raschen Amplitudenanstieg, dessen Maximum nach etwa 30 ms erreicht wird. Zu diesem Zeitpunkt verläßt die meiste angestaute Luft den Mundbereich. Nach etwa 120 ms ist der Explosionsschall ausgeklungen und es strömt keine weitere Luft mehr nach.

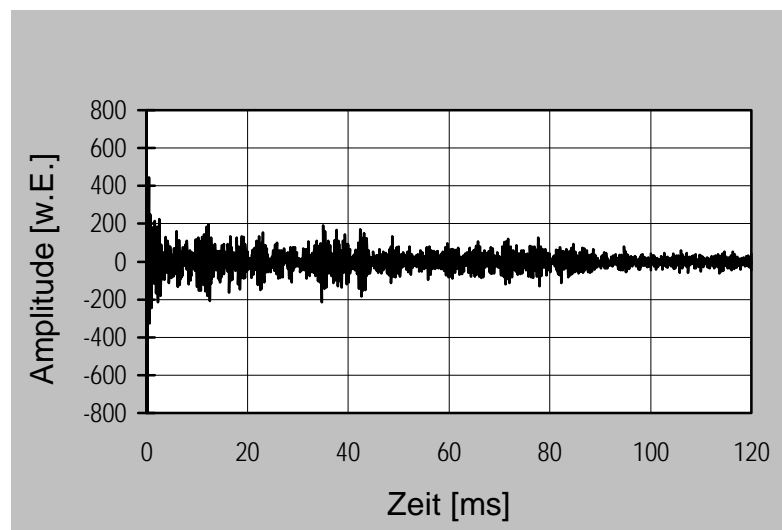
Da bei allen Testpersonen und bei fast allen Explosionsschallarten das Amplitudenmaximum nach etwa 30 ms erreicht wird, kann man davon ausgehen, daß es sich bei dieser Zeit um eine durch den Stimmapparat gegebene Konstante handelt. Die Konstante von 30 ms ist die Zeit, die die angestaute Luft braucht, um einen Verschluß zu lösen. Diese Zeitkonstante ist unabhängig von der Verschlußstelle. In der Sekundärliteratur findet sich für dieses Phänomen keine Erklärung, deshalb muß von diesem Erklärungsansatz ausgegangen werden.

In der Lorenzabbildung (Abbildung 6-2) verteilen sich die Punkte innerhalb des ersten Quadranten mit einem Schwerpunkt, der um den Nullpunkt herum bewegt. Das läßt darauf schließen, daß die meisten Maxima der Abtastfunktion über dem Nullpunkt liegen. Ausgehend vom Nullpunkt zerstreut sich die Lorenzabbildung in den ersten Quadranten hinein. Die Häufung der Punkte um den Nullpunkt herum entsteht durch das Abklingen der Abtastfunktion. Dies ist ein wichtiges Merkmal des Explosionsschalls.



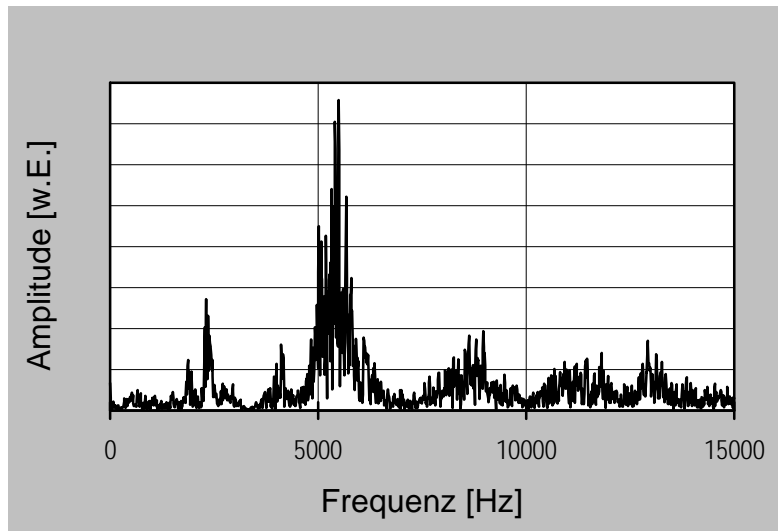
**Abbildung 6-2** Lorenzabbildung der Abtastfunktion [t] (weibl.)

Wendet man die Autokorrelationsfunktion auf die Abtastfunktion des Explosionsschalls [t] an, so sieht man, daß die Abtastfunktion keine Korrelation in der zeitlichen Abfolge aufweist. In Abbildung 6-3 ist die Autokorrelationsfunktion des Lauts [t] dargestellt.



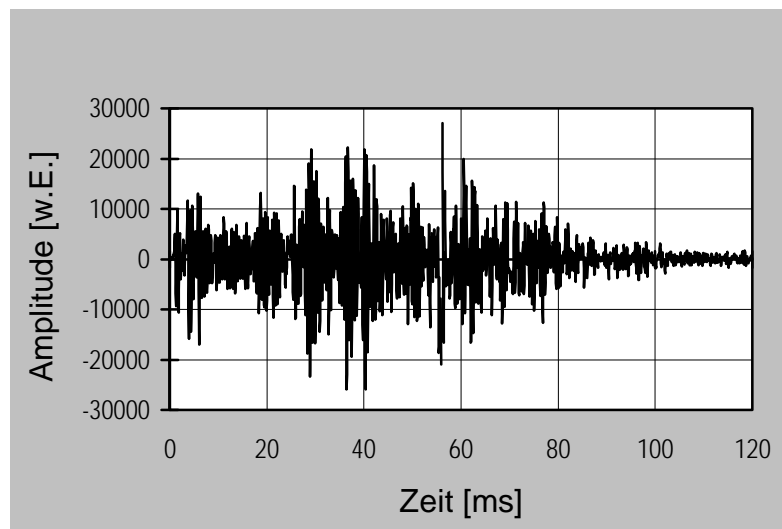
**Abbildung 6-3** Autokorrelationsfunktion des Explosionsschalls [t] (weibl.)

Im Frequenzspektrum (Abbildung 6-4) des Explosionsschalls [t] existiert ein ausgeprägtes Maximum bei 5400 Hz. Neben diesem Maximum liegen noch kleine Nebenmaxima.



**Abbildung 6-4** Frequenzspektrum des Explosionsschalls [t] (weibl.)

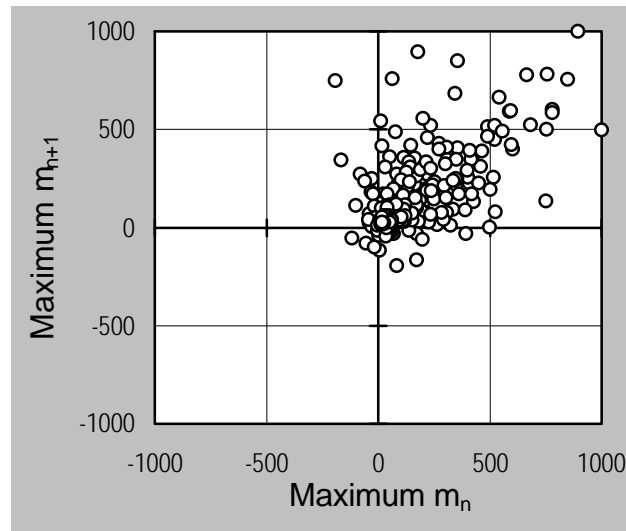
Vergleicht man den Explosionslaut [t] mit dem Explosionsschall [k], so stellt sich der Verlauf der Autokorrelationsfunktion ähnlich dar. Aus diesem Grund wird im weiteren nur die Abtastfunktion, die Lorenzabbildung und das Frequenzspektrum zur Erläuterung dieser Schallart verwendet.



**Abbildung 6-5** Abtastfunktion des Explosionsschalls [k] (weibl.)

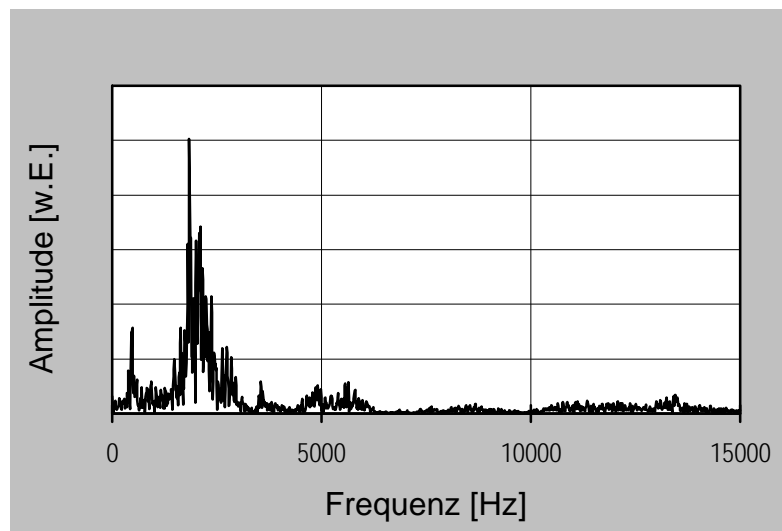
Wie in Abbildung 6-5 zu sehen ist, liegt das Maximum der Amplitude wieder bei ungefähr 30 ms. Wiederum ist der Explosionsschall [k] der weiblichen Testperson nur von kurzer Dauer, denn nach 120 ms läuft die Abtastfunktion fast gegen Null.

Die Lorenzabbildung weist eine ähnliche Form auf, die schon der Explosionsschall [t] gezeigt hat. Ein weiteres Mal liegt der Schwerpunkt der Punkteverteilung um den Nullpunkt herum. In Abbildung 6-6 wird die Lorenzabbildung des Explosionsschalls [k] dargestellt.



**Abbildung 6-6** Lorenzabbildung der Abtastfunktion aus Abbildung 6-5 (weibl.)

Das Frequenzspektrum des Explosionsschalls [k] der weiblichen Testperson unterscheidet sich vom Explosionsschall [t] in der Form, daß die Lage des absoluten Maximums von 5400 Hz auf 2000 Hz gefallen ist. Wenn man sich die beiden Explosionsschallarten anhört, kann man auch den Unterschied der beiden wahrnehmen, weil sich das [k] tiefer anhört als das [t]. Außerdem ist die Amplitude der Nebenmaxima kleiner geworden.



**Abbildung 6-7** Frequenzspektrum des Explosionsschalls [k] (weibl.)

Die nächsten Diagramme zeigen den Expolsivlaut [k] einer männlichen Testperson.

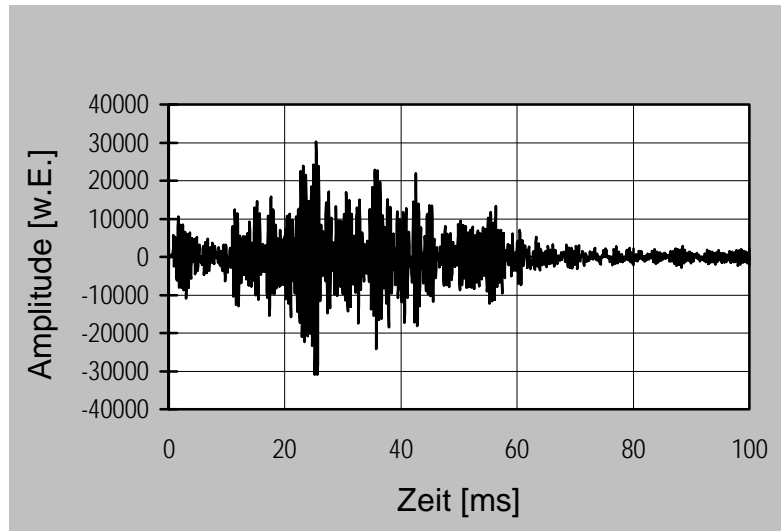
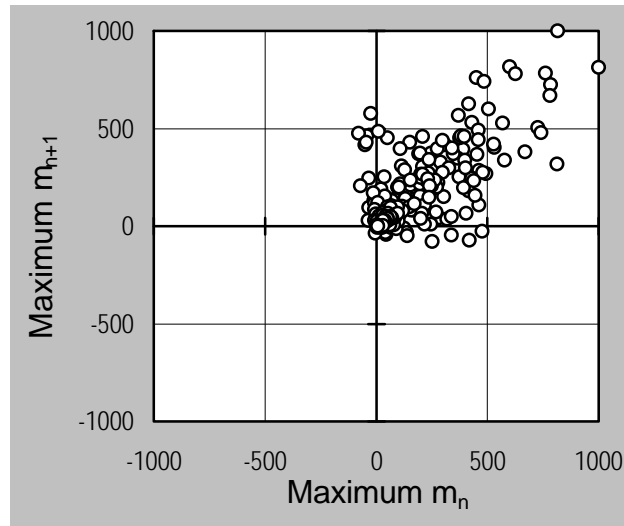
**Abbildung 6-8** Abtastfunktion des Explosionsschalls [k] (männl.)

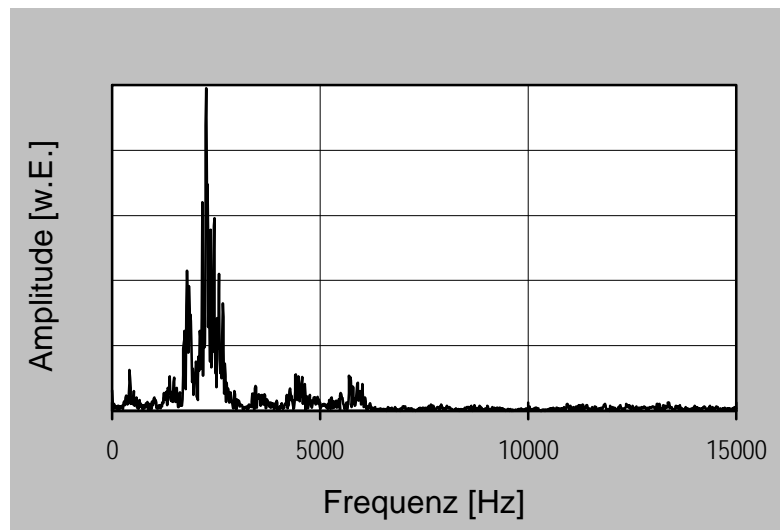


Abbildung 6-8 stellt die Abtastfunktion des Explosionsschall [k] der männlichen Testperson dar und Abbildung 6-9 die dazugehörige Lorenzabbildung.



**Abbildung 6-9** Lorenzabbildung der Abtastfunktion aus Abbildung 6-8 (männl.)

Im Frequenzspektrum in Abbildung 6-10 findet sich das absolute Maximum auch wieder bei ungefähr 2000 Hz und zeigt einen ähnlichen Verlauf wie Abbildung 6-7.



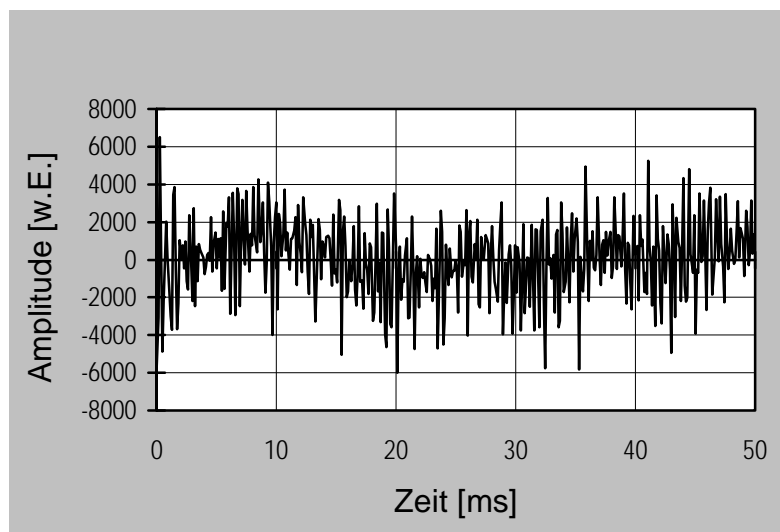
**Abbildung 6-10** Frequenzspektrum des Explosionsschalls [k] (männl.)

An Hand von drei Abtastfunktionen wurde der Explosionsschall beschrieben. Der Explosionsschall ist im allgemeinen charakteristisch für alle Typen von

Explosionslauten. Typisch für alle Explosionslaute war, daß das [k] sein Hauptmaximum im Frequenzspektrum immer bei etwa 2000 Hz hatte und das [t] bei 5500 Hz. Dieser Unterschied ist auch hörbar, denn das [t] klingt höher als das [k]. Da sich zwischen den einzelnen Explosionslauten keine großen Unterschiede erkennen ließen, wird auf weitere Beispiele verzichtet.

## 6.4 Das Friktionsrauschen

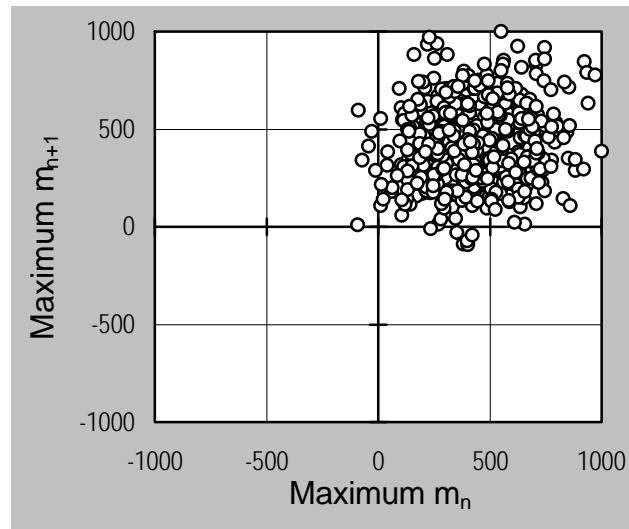
Ein Friktionsrauschen wird durch eine kritische Verengung im Ansatzrohr hervorgerufen. Die Luftströmung geht durch diese Verengung von ihrer laminaren Form vor der Verengung in eine turbulente Strömung, hinter der Verengung, über. Das heißt, daß eine parallele Teilchenbewegung in eine verwirbelte überwechselt.[2,9] Abbildung 6-11 stellt eine Abtastfunktion des Friktionsrauschens [f] dar. Die Abtastfunktion wurde aus einem Stück des Friktionsrauschens, das als Einzellaute gehalten wurde, herausgeschnitten.



**Abbildung 6-11** Abtastfunktion des Friktionsrauschens [f] (weibl.)

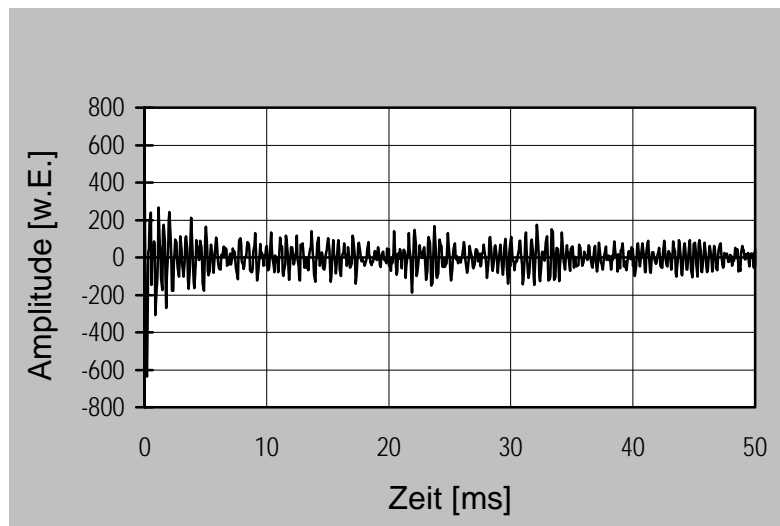
Die Abtastfunktion zeigt ein völlig aperiodisches Verhalten. Unterscheidbar werden die Friktionsrauschen im wesentlichen durch die Art und den Grad der Engbildung im Ansatzrohr.

Die Lorenzabbildung des Friktionsrauschens aus Abbildung 6-11 weist ein ähnliches Bild auf wie die Untersuchung des Rauschsignals aus Kapitel 3.1. Die meisten Maxima der Abtastfunktion liegen oberhalb des Nullpunkts. Im Gegensatz zum Explosionsschall klingt das Friktionsrauschen nicht aus, sondern kann beliebig lang gesprochen werden. Deshalb liegt eine Häufung der Punkte in der Mitte des ersten Quadranten vor. Im Gegensatz zum Explosionsschall, bei dem die Amplitude der Abtastfunktion gegen Null strebt und ein Häufungspunkt in der Nähe des Koordinatenursprungs besteht.



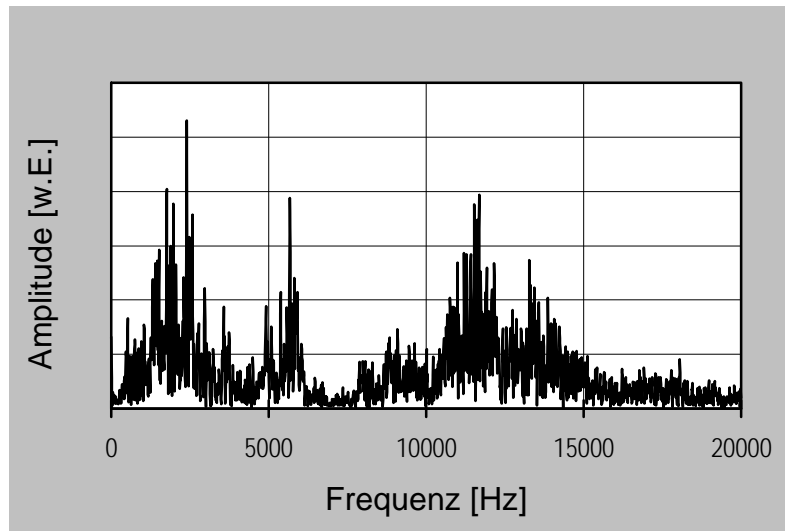
**Abbildung 6-12** Lorenzabbildung des Friktionsrauschens [f] (weibl.)

Die Autokorrelationsfunktion des Friktionsrauschens [f] zeigt keine Korrelation in der zeitlichen Abfolge, wie in Abbildung 6-13 dargestellt.



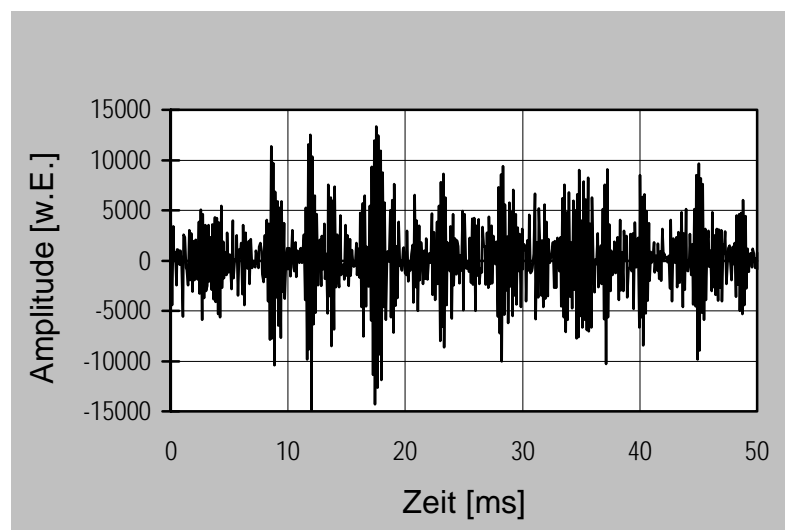
**Abbildung 6-13** Autokorrelationsfunktion des Friktionsrauschens [f] (weibl.)

Im Frequenzspektrum erreicht das Friktionsrauschen [f] das erste Maximum bei annähernd 2500 Hz. Es folgen Maxima etwa bei 5500 Hz, 8000 Hz, 12000 Hz und bei 13500 Hz.



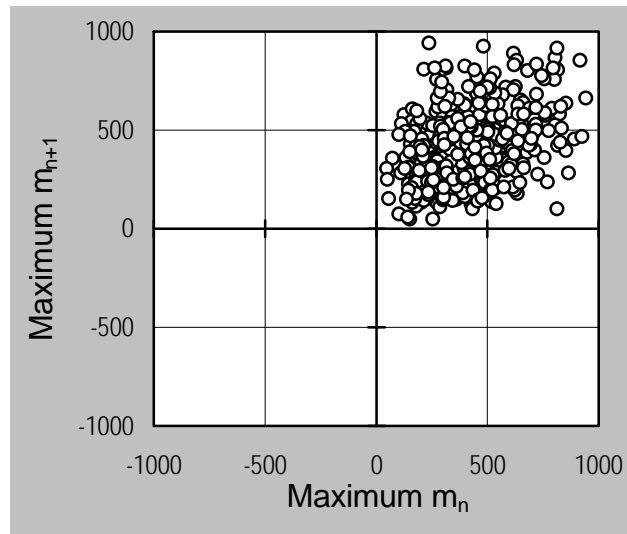
**Abbildung 6-14** Frequenzspektrum des Friktionsrauschens [f] (weibl.)

Bei der Gegenüberstellung aller Friktionsrauschen der Testpersonen ergeben sich abermals keine gravierenden Unterschiede in der Abtastfunktion, sowie in der Lorenzabbildung und auch in der Autokorrelationsfunktion. In den Frequenzspektren liegen die Maxima an verschiedenen Stellen. Das sollte auch so sein, weil man sonst die verschiedenen Friktionsrauschen nicht auseinanderhalten könnte. Die in Abbildung 6-15 dargestellte Abtastfunktion des Friktionsrauschens [s] ist eine Aufnahme einer männlichen Testperson.



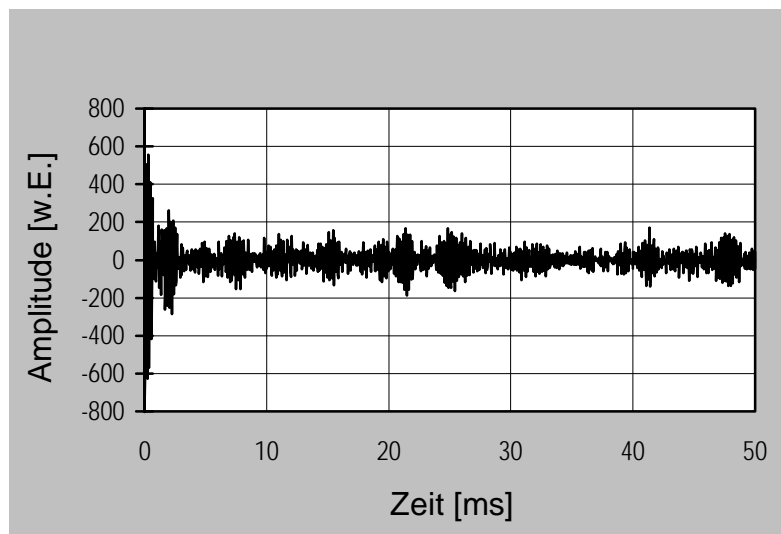
**Abbildung 6-15** Abtastfunktion des Friktionsrauschens [s] (männl.)

Die Lorenzabbildung in Abbildung 6-16 läßt eine ähnliche Form erkennen wie die des Friktionsrauschens [f] aus Abbildung 6-12.



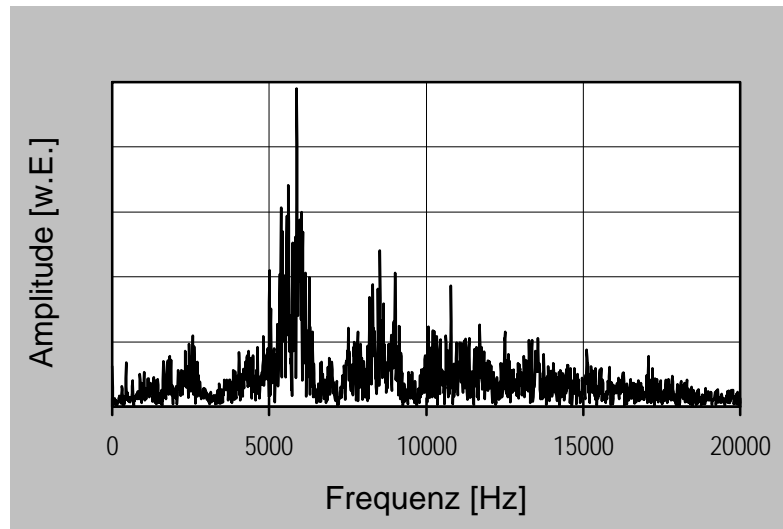
**Abbildung 6-16** Lorenzabbildung des Friktionsrauschens [s] (männl.)

Abbildung 6-17 stellt die Autokorrelationsfunktion des als Einzellaut gesprochenen Friktionsrauschens [s] dar. Im Vergleich zur Autokorrelationsfunktion aus Abbildung 6-13 kann man keine großen Unterschiede nachweisen.



**Abbildung 6-17** Autokorrelationsfunktion des Friktionsrauschens [s] (männl.)

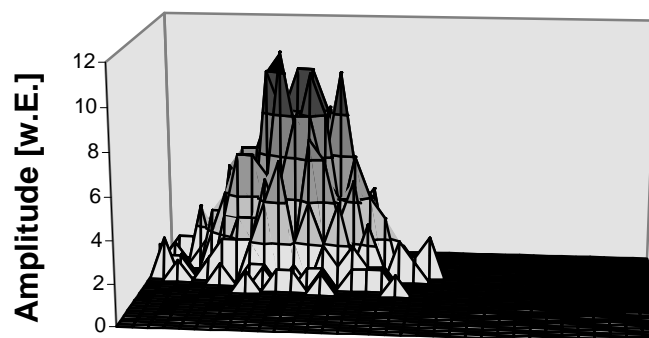
Das Hauptmaximum im Frequenzspektrum der Abbildung 6-18 liegt ungefähr bei 6000 Hz. Die folgenden Maxima liegen bei 8000 Hz und bei 11500 Hz. Im Vergleich zum Frequenzspektrum des Friktionsrauschens [f] haben sich die Amplitudenschwerpunkte der Maxima verschoben. Das Friktionsrauschen [f] hat das Schwerpunktmaximum bei 2500 Hz und erzeugt einen tieferen Ton als beim Friktionsrauschen [s], dessen Maximum bei ungefähr 6000 Hz liegt. Wenn man diese Laute spricht, hört man den Unterschied in der Tonhöhe.

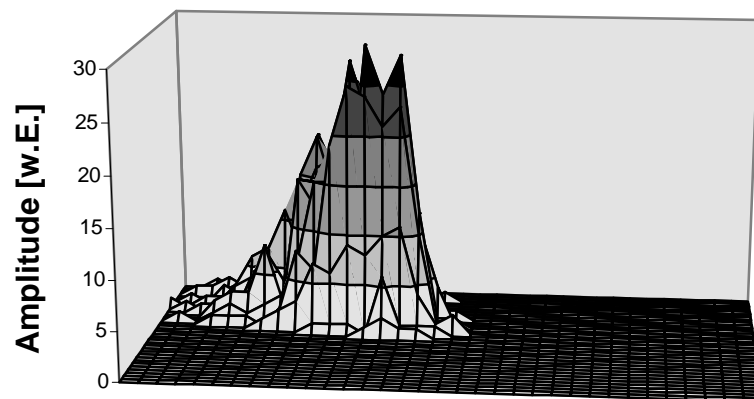


**Abbildung 6-18** Frequenzspektrum des Friktionsrauschen [s] (männl.)

Während der Analyse zeigte sich, daß sich die Lorenzabbildungen der Friktionsrauschen [f], [s] doch unterscheiden. In der 3D-Lorenzabbildung erkennt man den Unterschied.

Das Friktionsrauschen [f] hat um das Hauptmaximum eine Ansammlung von kleinen Nebenmaxima, die man in Abbildung 6-19 erkennen kann. Betrachtet man dagegen die 3D-Lorenzabbildung des Friktionsrauschens [s], so finden sich keine Anhäufungen von Nebenmaxima mehr. Auf der rechten Seite ist der „Berg“ steil und es finden sich keine Nebenmaxima. Auf der linken Seite hingegen geht das Hauptmaximum in kleine Nebenmaxima über. Die Erhebung der Maxima läuft langsam aus. Dieses Verhalten der Friktionsrauschen ist sprecherunabhängig.



**Abbildung 6-19** 3D-Lorenzabbildung des Friktionsrauschens [f]**Abbildung 6-20** 3D-Lorenzabbildung des Friktionsrauschens [s]

Der große Vorteil der 3D-Lorenzabbildung ist, daß mehr Details sichtbar werden. Gerade beim Friktionsrauschen können Feinheiten hervorgehoben werden, die in der 2D-Lorenzabbildungen nicht zu sehen sind.

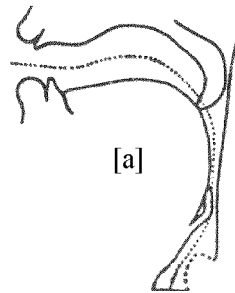


## 6.5 Der periodische Schall

Durch die Vibration der Stimmlippen wird dieser Stimmklang erzeugt. Er ist die Grundlage für die Kardinalvokale<sup>5</sup>, Nasale<sup>6</sup> und Semivokale<sup>7</sup>. [9,25] Dieses Signal kann man aber nie in der ursprünglichen Form hören, da der eigentliche Klang nur kurz hinter der Glottis wahrnehmbar ist. Man hört also einen durch den Rachenraum gefilterten vokalartigen Stimmklang. Der biophysikalische Entstehungsmechanismus der Vokale wurde in Kapitel 2.2 beschrieben. Die Tönung der Kardinalvokale wird durch die Position der Zungenstellung und die damit verbundenen Verengungen bzw. Erweiterungen des Mundraums bestimmt. Im folgenden wird der periodischen Schall mittels der Kardinalvokale erörtert.

### 6.5.1 Der Kardinalvokal [a]

Die typische Mundstellung um den Vokal [a] zu bilden, zeigt Abbildung 6-21. [21]



**Abbildung 6-21** Mundstellung bei der Artikulation des Vokals [a] nach Fant [3]

In Abbildung 6-22 ist der quasiperiodische Schwingungsverlauf des Kardinalvokals [a] zu erkennen. Im Vergleich zu den anderen Abtastfunktionen der Vokale zeigt die Abtastfunktion des Vokals [a] eine sehr regelmäßige

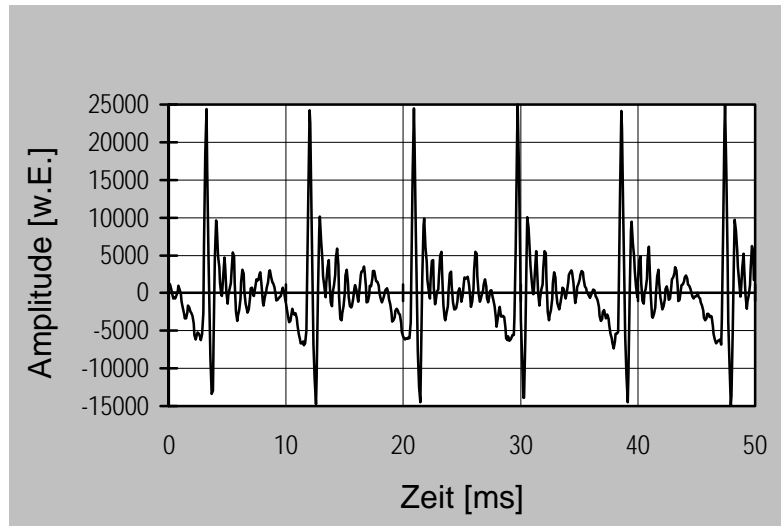
---

<sup>5</sup> Bezeichnung für die Vokale [a], [e], [i], [o], [u].

<sup>6</sup> Der Weg zur Nasenhöhle ist nicht, wie bei den Explosivlauten, verschlossen, sondern frei. Der weiche Gaumen senkt sich und ermöglicht so, daß die strömende Luft auch durch die Nasenhöhle entweichen kann. Beispiele für einen Nasallaut sind [m] und [n].

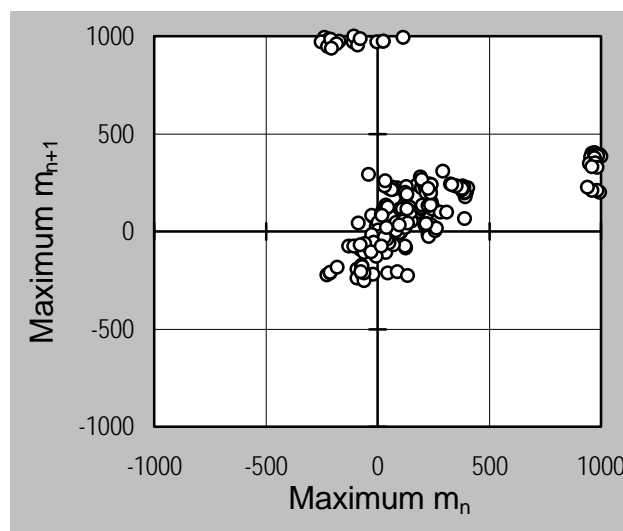
<sup>7</sup> Semivokale sind Halbvokale. Sie liegen in ihrer Artikulation zwischen Konsonanten und Vokalen. Ein Beispiel ist das [j], wie zum Beispiel in „Kojé“.

Struktur. Dieser Vokal stammt von der ersten männlichen Testperson (männl. 1). Eine Periode des Vokals beträgt annähernd 9 ms. Das entspricht einer Grundfrequenz von 111 Hz.



**Abbildung 6-22** Abtastfunktion des Vokals [a] (männl.1)

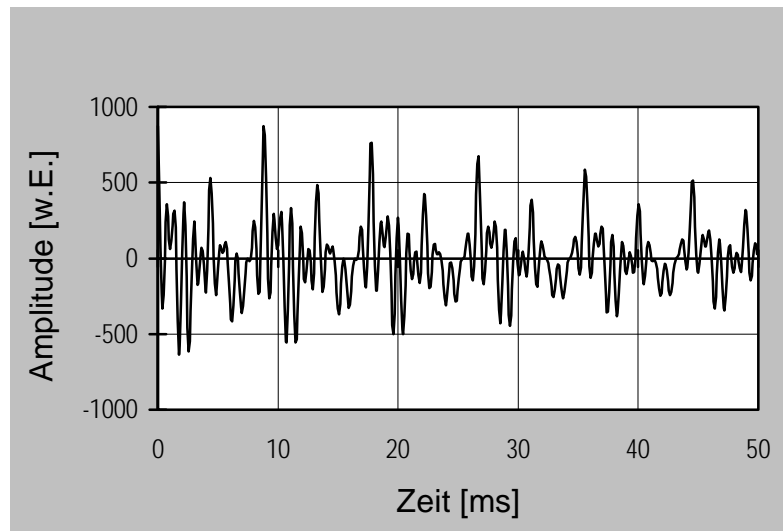
Die Lorenzabbildung weist eine Häufung der Punkte im Koordinatenursprung auf. Weitere verteilen sich oberhalb und rechts von dieser Anhäufung. Die Ansammlung von Punkten weist auf eine annähernd periodische Funktion hin, deren Maxima über dem Nullpunkt liegen. Durch die Lorenzabbildung kann man diese Schallart von den anderen unterscheiden. In Abbildung 6-23 ist die Lorenzabbildung des Vokals [a] dargestellt.



**Abbildung 6-23** Lorenzabbildung des Vokals [a] (männl.1)

Während der Arbeit stellte sich heraus, daß sich die Lorenzabbildung nicht nur zur Bestimmung einer Schallart eignet, sondern durch sie auch ein „sprecher-spezifischer Fingerabdruck“ erstellt werden kann. Bei den verschiedenen Testpersonen fielen einige Besonderheiten auf, die sich in den folgenden Lorenzabbildungen zeigen.

In der Autokorrelationsfunktion (Abbildung 6-24) zeigt sich ein fast linearer Abfall. Der ideale Fall wäre ein linearer Abfall der Autokorrelationsfunktion, der jedoch bei Aufnahmen der menschlichen Stimme kaum erreicht werden kann.



**Abbildung 6-24** Autokorrelationsfunktion der Abtastfunktion [a] (männl.1)

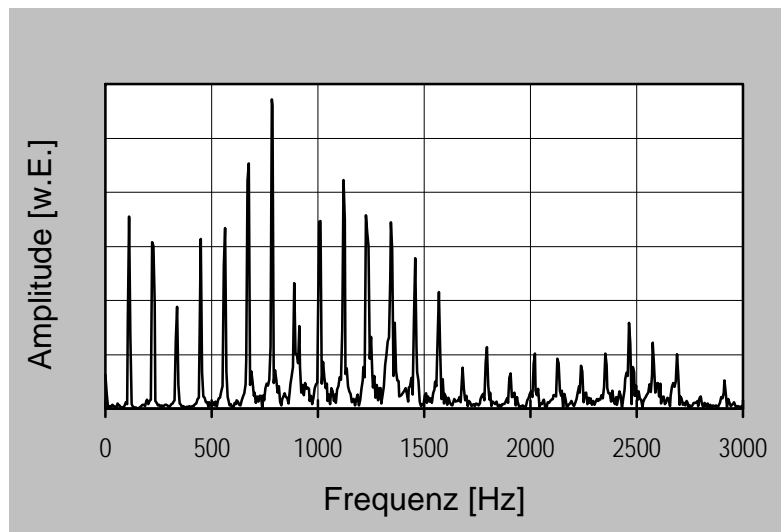
Betrachtet man das zugehörige Frequenzspektrum aus Abbildung 6-25, so werden die Auswirkungen der Mundstellung auf das Frequenzspektrum sichtbar. Durch die ersten zwei Formanten kann man einen Kardinalvokal hinreichend genau charakterisieren. Die Formanten befinden sich an den Stellen, wo sich ein scharfes Maximum ausprägt. Zusätzlich kann man aus dem Frequenzspektrum die Grundfrequenz der jeweiligen Testpersonen bestimmen, indem man die Frequenz des n-ten Maximums durch n teilt.

Die Grundfrequenz dieser Testperson beim Vokal [a] liegt bei 111 Hz. Die ersten beiden Formanten<sup>8</sup> finden sich bei 750 Hz und 1100 Hz. Der dritte

---

<sup>8</sup> Zur Erinnerung: Formanten sind vergleichbar mit den Eigenresonanzen im einseitig geschlossenen Rohr (siehe Kapitel 2.2.3).

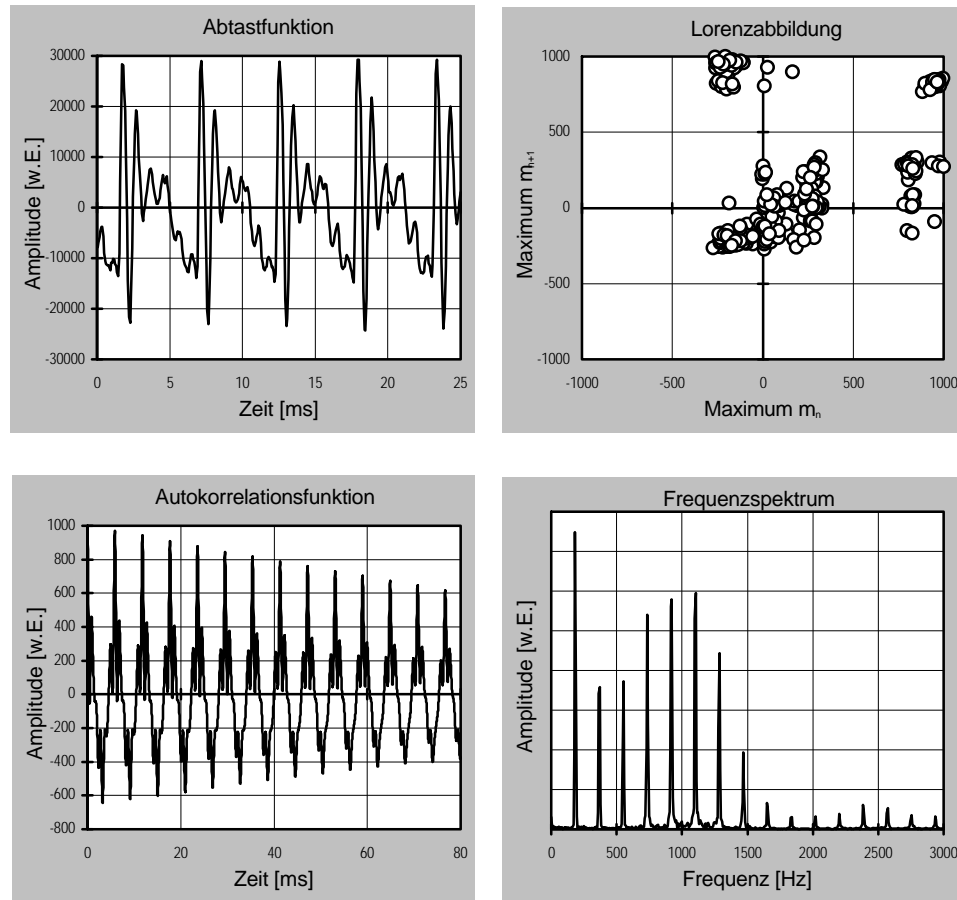
Formant, der für die Beschreibung der Vokale nicht von großer Bedeutung ist, liegt bei 2500 Hz.



**Abbildung 6-25** Frequenzspektrum des Vokals [a] (männl.1)

Bei den anderen Testpersonen wurden die gleichen Analyseverfahren angewandt. Um einen Vergleich zwischen den einzelnen Testpersonen ziehen zu können, werden die Ergebnisse für den Vokal [a] zusammenfassend auf den nächsten Seiten dargestellt.

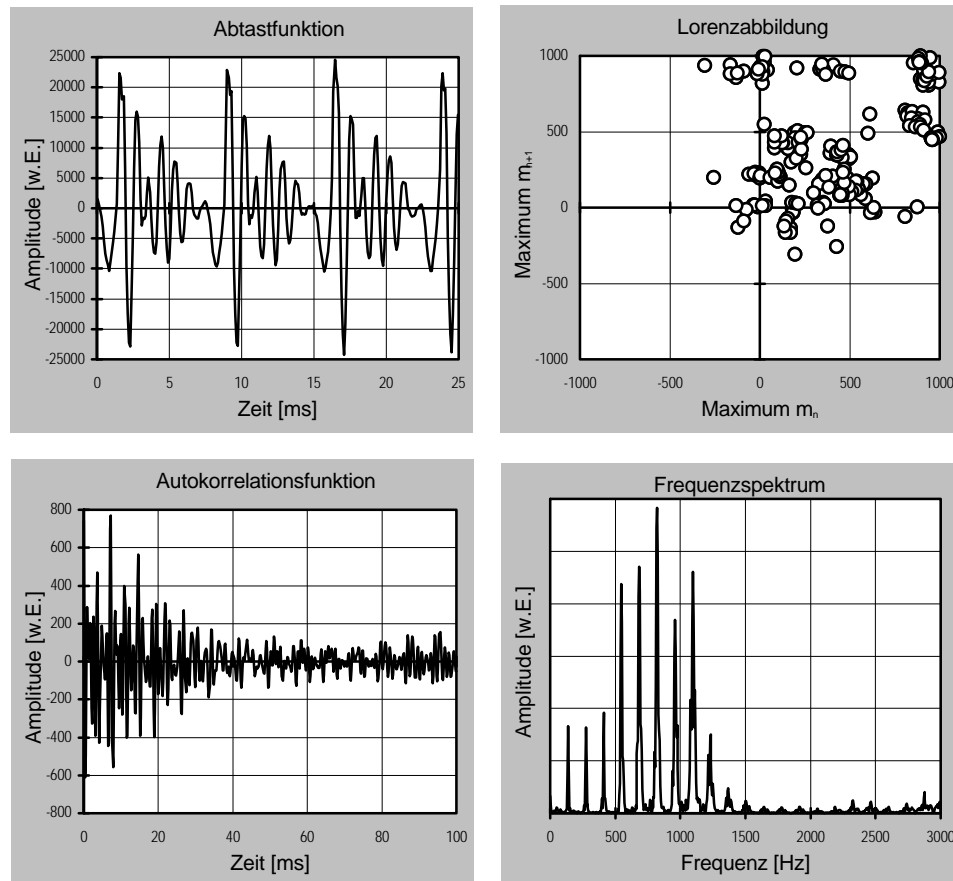
Der Vokal [a] in Abbildung 6-26 wurde von der weiblichen Testperson gesprochen. Die Abtastfunktion des gesprochenen Vokals ist annähernd periodisch und hat eine Periodenlänge von 5 ms. In der Lorenzabbildung finden sich wieder Anhäufungen von Punkten um den Nullpunkt herum. Die Lorenzabbildung zeigt ein ähnliches Bild wie die zuvor besprochene Abbildung 6-23. Die Besonderheit dieses gesprochenen Vokals befindet sich rechts oben im ersten Quadranten. Für diese Testperson ist die Punktanhäufung in der Ecke charakteristisch. Die Autokorrelationsfunktion ist linear abfallend.



**Abbildung 6-26** Ergebnisse der Analyseverfahren des Vokals [a] (weibl.)

Im Frequenzspektrum liegen die Formanten sehr nahe beieinander. Die Grundfrequenz des gesprochenen Kardinalvokals [a] beträgt 182 Hz. Der erste Formant liegt ungefähr bei 850 Hz und der zweite bei 1150 Hz. Vergleicht man die ersten beiden Formanten der männlichen und weiblichen Testperson, so sind die Formantenfrequenzen der weiblichen Testperson höher als die der männlichen. Das zeigt sich auch schon an der viel höheren Grundfrequenz der weiblichen Stimme. Da die weibliche Stimme höher ist als die männliche, hat man oft Probleme die Frequenzspektren auszuwerten, denn die ersten beiden Formanten lassen sich häufig kaum unterscheiden. Weil die Dichte der Stützpunkte (Maxima des Frequenzspektrums) für die Einhüllende kleiner ist als bei den männlichen Stimmen, kann man nicht unbedingt unterscheiden, wo die Formanten liegen. Aus diesem Grund wurde die Anzahl an männlichen Testpersonen höher gewählt, als die der weiblichen.

In Abbildung 6-27 sind die Ergebnisse einer zweiten männlichen Testperson (männl. 2) dargestellt.

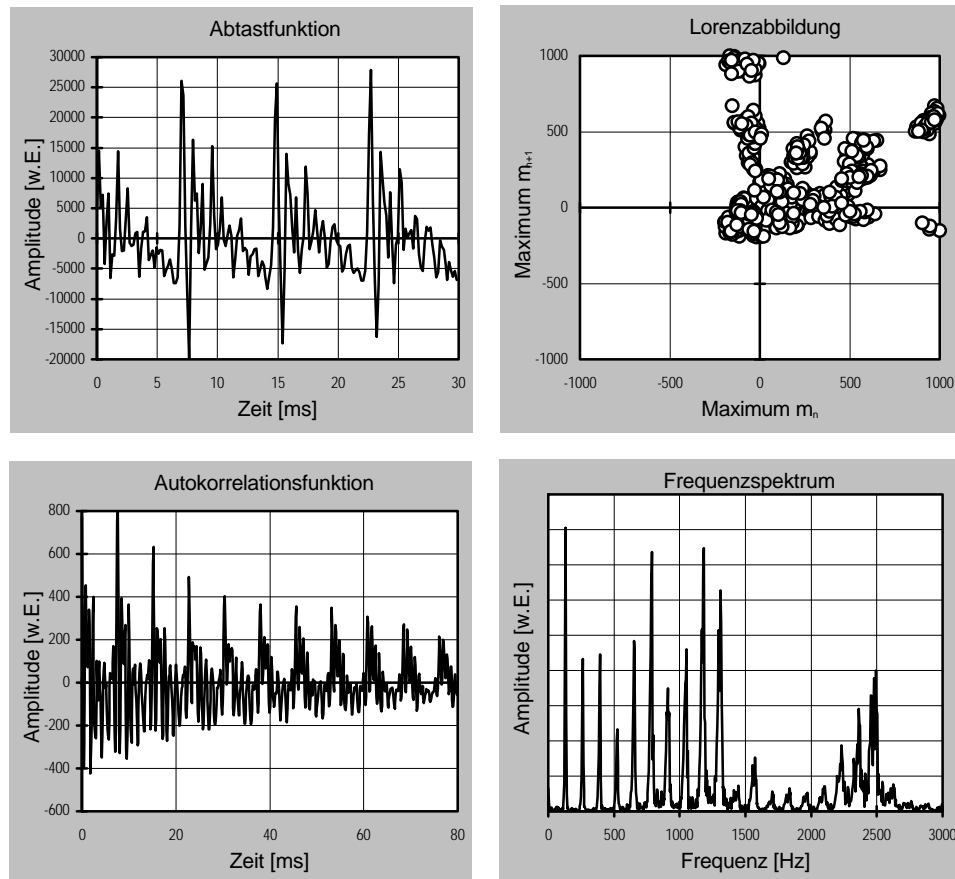


**Abbildung 6-27** Ergebnisse der Analyseverfahren des Vokals [a] (männl. 2)

Die Lorenzabbildung zeigt ein Punktmuster, welches sich im ersten Quadranten bewegt. Die Punktanhäufungen finden sich, wie schon bei den anderen Testpersonen beschrieben, in den Ecken des ersten Quadranten. Die Besonderheit dieses Sprechers zeigt sich in der Sprengelung der Punkte.

Die Autokorrelationsfunktion fällt nicht linear ab, sondern erfährt nach 80 ms noch einmal einen Amplitudenanstieg. Der Amplitudenanstieg wird durch leichte Amplitudenschwankungen in der Stimme und durch Unterschiede in der Periodizität der Abtastfunktion verursacht. Die Formanten liegen bei 750 Hz und 1050 Hz. Diese Testperson spricht das [a] mit einer etwas höheren Grundfrequenz (135 Hz), als die erste männliche Testperson.

Für die dritte männliche Person (männl. 3) zeigen sich folgende Ergebnisse:



**Abbildung 6-28** Ergebnisse der Analyseverfahren des Vokals [a] (männl. 3)

Die Abtastfunktion zeigt wieder ein periodisches Verhalten, allerdings mit einem Überhang im positiven Bereich. Die Autokorrelation spiegelt das Verhalten der Abtastfunktion wieder. Die Grundfrequenz des Vokals [a] liegt bei diesem Sprecher bei 130 Hz. Die Formanten liegen bei 750 Hz und 1100 Hz.

In Tabelle 6-1 sind die Frequenzen der Formanten der einzelnen Sprecher zusammengefaßt. Die weibliche Testperson wird im Durchschnittswert der Formanten nicht berücksichtigt.

**Tabelle 6-1** Frequenzen der Formanten des Kardinalvokals [a]

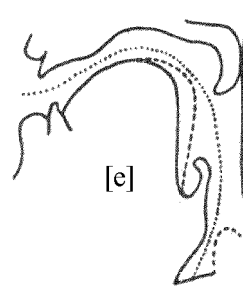
	weibl.	männl. 1	männl. 2	männl. 3	Durchschnitt (männl.)
<b>Grundfrequenz [Hz]</b>	182	111	135	130	—
<b>1. Formant [Hz]</b>	850	750	750	750	750
<b>2. Formant [Hz]</b>	1150	1100	1050	1100	$1080 \pm 20$

Da aus den Autokorrelationsfunktionen für den Kardinalvokal [a] keine Schlüsse gezogen werden konnten, wird in den folgenden Betrachtungen auf die Autokorrelationsfunktion verzichtet.



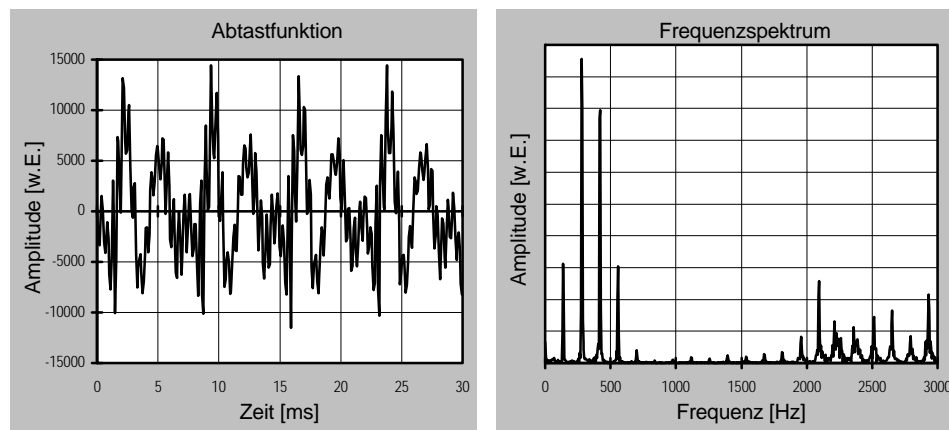
### 6.5.2 Der Kardinalvokal [e]

Abbildung 6-29 stellt die charakteristische Mund- und Zungenstellung für den Kardinalvokal [e] dar.



**Abbildung 6-29** Mundstellung für den Kardinalvokal [e] nach Fant [3]

Da die Abtastfunktionen und die Frequenzspektren der einzelnen Testpersonen nur geringe Unterschiede erkennen lassen, wird die Beschreibung dieser Analyseverfahren bei den nächsten Kardinalvokalen auf eine Testperson beschränkt. Die Abtastfunktion und das zugehörige typische Frequenzspektrum einer Testperson, die den Vokal [e] gesprochen hat, wird in Abbildung 6-30 dargestellt.



**Abbildung 6-30** Abtastfunktion und Frequenzspektrum des Vokals [e] (männl. 2)

Die Abtastfunktion des Kardinalvokals [e] weist eine etwas diffusere Struktur als die des Vokals [a] auf, denn in seinen zwei großen Maxima sind noch kleine Nebenmaxima, die die Abtastfunktion des Vokals [a] nicht zeigt. In dieser etwas unscharfen Form findet sich eine Periodenlänge von 7 ms. Im Frequenzspektrum des Kardinalvokals [e] sieht man die Auswirkungen der Mund- und Zungenstellungen. Die ersten beiden Formanten haben sich weit

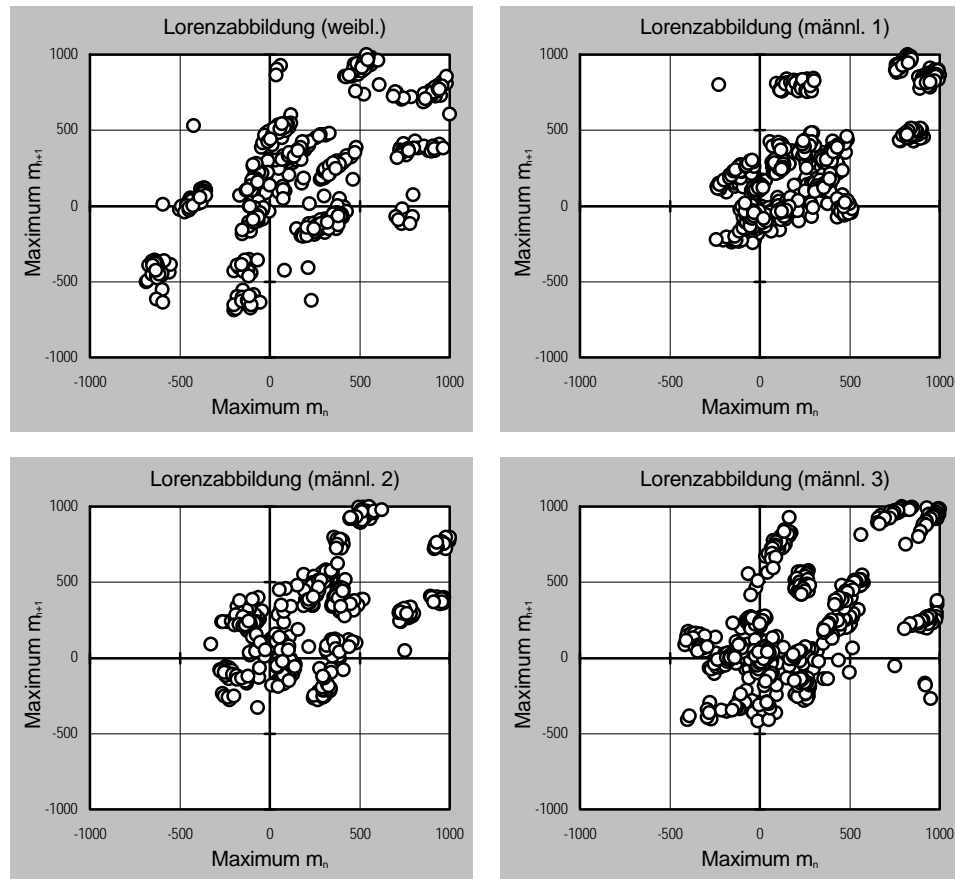
voneinander entfernt. Sie liegen nicht mehr so nah beieinander, wie beim Vokal [a]. Der erste Formant liegt bei ungefähr 400 Hz und der zweite bei 2000 Hz. Die Grundfrequenz des gesprochenen Vokals liegt bei 140 Hz. Die restlichen Formantenfrequenzen der einzelnen Testpersonen sind in Tabelle 6-2 zusammengefaßt.

**Tabelle 6-2** Frequenzen der Formanten des Kardinalvokals [e]

	weibl.	männl. 1	männl. 2	männl. 3	Durchschnitt (männl.)
<b>Grundfrequenz [Hz]</b>	210	130	140	130	—
<b>1. Formant [Hz]</b>	550	370	400	350	$370 \pm 20$
<b>2. Formant [Hz]</b>	2400	2100	2000	2050	$2050 \pm 40$

Die Formantenfrequenzen der weiblichen Testperson liegen wieder über denen der männlichen Personen.

Stellt man die Lorenzabbildungen der Testpersonen gegenüber, sieht man wiederum Unterschiede in der Darstellung. Die kleinen Unterschiede in der Abtastfunktion, die man normalerweise nicht wahrnehmen würde, sieht man in der Lorenzabbildung deutlich. Sie stehen als Repräsentanten für die Unterschiede in der Abtastfunktion. Durch die unterschiedlichen Frequenzgemische in der Stimme ergeben sich bei jedem Probanden zwangsweise leicht unterschiedliche Abtastfunktionen, was man an der Lorenzabbildung gut erkennt. In den Lorenzabbildungen werden die kleinen Nebenmaxima der Abtastfunktion des Vokals [e] sichtbar, denn es treten mehrere kleine Punktanhäufungen auf. In Abbildung 6-31 sieht man die Unterschiede zwischen den einzelnen Sprechern.

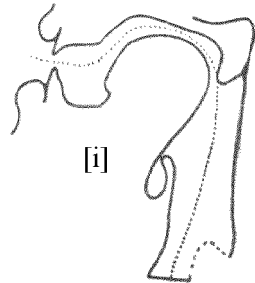


**Abbildung 6-31** Lorenzabbildungen des Vokals [e]

Bei der weiblichen Testperson zieht sich die Punkthäufung vom ersten in den dritten Quadranten hinein. Das weist darauf hin, daß sich viele Maxima der Abtastfunktion immer noch oberhalb des Nullpunkts befinden. Einige aufeinanderfolgende Maxima liegen jedoch auch unter dem Nullpunkt. Bei den männlichen Probanden liegen die meisten Maxima oberhalb des Nullpunkts. Allen Lorenzabbildungen aus Abbildung 6-31 ist aber gemeinsam, daß sie in ihrer Form länglicher sind als die des Vokals [a].

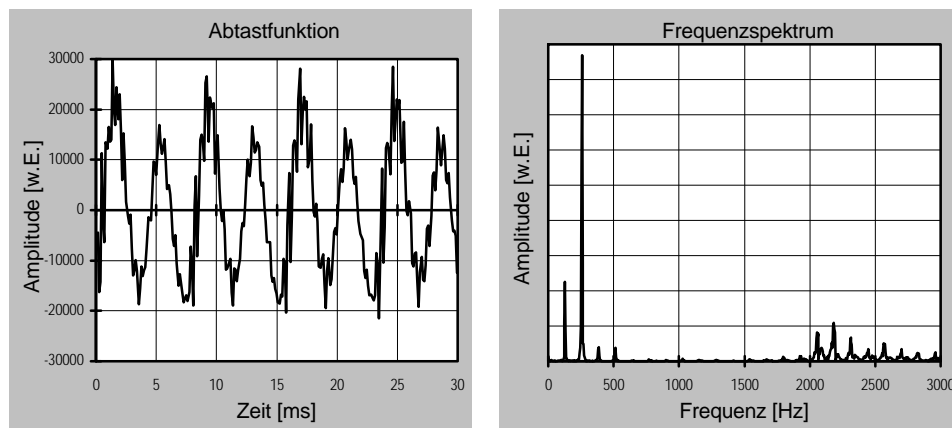
### 6.5.3 Der Kardinalvokal [i]

Bei der Artikulation des Vokals [i] nimmt der Mundbereich die in Abbildung 6-32 skizzierte Stellung ein. [21]



**Abbildung 6-32** Mund- und Zungenstellung bei der Artikulation des Vokals [i] nach Fant [3]

In Abbildung 6-33 werden die zur Mundstellung des Vokals [i] entsprechende Abtastfunktion und das dazugehörige Frequenzspektrum dargestellt.



**Abbildung 6-33** Abtastfunktion und Frequenzspektrum des Vokals [i] (männl. 3)

Der Proband (männl. 3) hat diesen Vokal mit einer Grundfrequenz von 125 Hz gesprochen. Die Formanten des Vokals [i] liegen noch weiter auseinander als die des zuvor besprochenen [e]. Der erste Formant liegt unweit der Grundfrequenz bei 250 Hz. Im Gegensatz dazu liegt der zweite Formant bei 2250 Hz. Die Ergebnisse der anderen Testpersonen bestätigen die Formantenfrequenzen des Frequenzspektrums. In Tabelle 6-3 liegen die Frequenzen der Formanten und die entsprechenden Grundfrequenzen der Probanden für den Kardinalvokal [i] vor. Besonders offensichtlich wird der Unterschied zwischen einer weiblichen und einer männlichen Stimme in der Lage des zweiten Formanten.

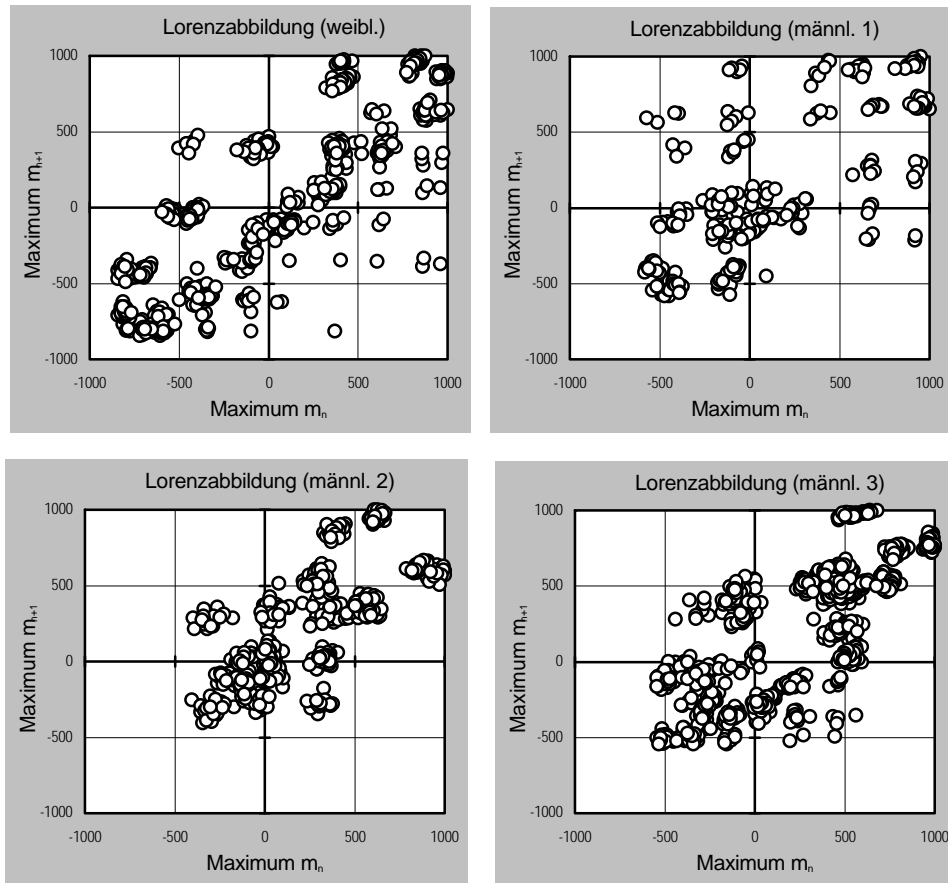
Der zweite Formant der weiblichen Testperson liegt bei 2700 Hz. Es besteht demzufolge eine Differenz von mehr als 400 Hz beim zweiten Formanten.

**Tabelle 6-3** Frequenzen der Formanten des Kardinalvokals [i]

	weibl.	männl. 1	männl. 2	männl. 3	Durchschnitt (männl.)
<b>Grundfrequenz [Hz]</b>	200	155	140	125	—
<b>1. Formant [Hz]</b>	300	250	280	250	$260 \pm 10$
<b>2. Formant [Hz]</b>	2700	2300	2300	2250	$2280 \pm 20$

Da die Abtastfunktionen der Vokale [e] und [i] sehr ähnlich sind, zeigen sich nur geringfügige Unterschiede in den Lorenzabbildungen. In Abbildung 6-34 ist der Vokal [i] der Probanden als Lorenzabbildung dargestellt.

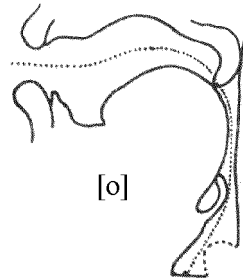
Auf den ersten Blick könnte man nicht unbedingt sofort sagen welcher Vokal gesprochen wird. Vergleicht man jedoch die Lorenzabbildungen der Vokale sprecherspezifisch, so ergeben sich Differenzen. Die Lorenzabbildung des Vokals [i] bei der weiblichen Testperson ist gestreckter als beim [e]. Die aufeinanderfolgenden Maxima beim [i] weisen im negativen Bereich einen kleineren Amplitudenunterschied auf. Bei den männlichen Probanden kann man auch eine Streckung der Punkthäufungen erkennen. Das Verhalten der Vokale [e] und [i] ließen sich durch mehrere Sprechproben reproduzieren.



**Abbildung 6-34** Lorenzabbildungen des Vokals [i]

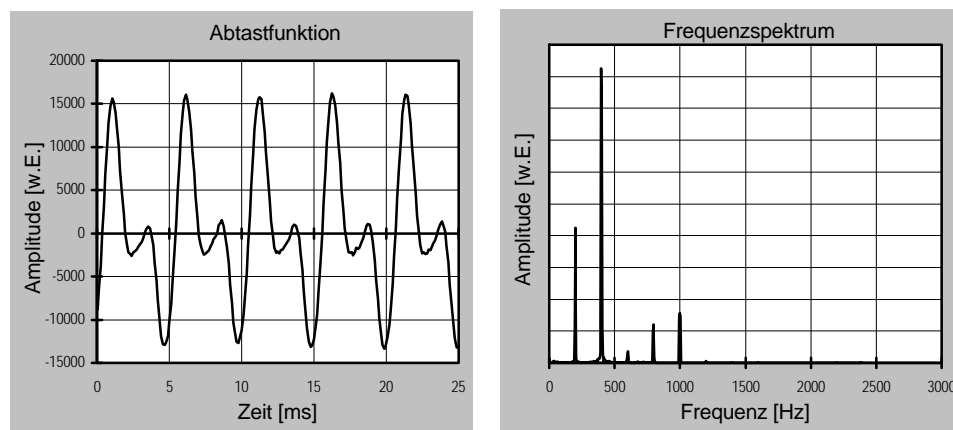
#### 6.5.4 Der Kardinalvokal [o]

Wie ausschlaggebend die Mundstellung für die Bildung der Vokale ist, hat sich schon in den letzten Kapiteln gezeigt. Ein weiterer Kardinalvokal ist das [o], dessen unverkennbare Mundstellung zur Artikulation in Abbildung 6-35 gezeigt wird.



**Abbildung 6-35** Mundstellung für die Artikulation des Kardinalvokals [o] nach Fant [3]

Die Auswirkung dieser Mundstellung auf den Schall soll durch Abbildung 6-36 verdeutlicht werden. Die Abtastfunktion zeigt eine Periodenabfolge mit zwei verschiedenen Amplituden. Nach einer Schwingung mit einer großen Amplitude folgt ein Ausschlag mit einer kleineren Amplitude. Die Abtastfunktion zeigt einen fast sinusförmigen Verlauf. Die Formanten des Vokals [o] bei der weiblichen Testperson befinden sich bei 550 Hz und 900 Hz.



**Abbildung 6-36** Abtastfunktion und Frequenzspektrum des Vokals [o] (weibl.)

Die einfache Form der Abtastfunktion spiegelt sich im Frequenzspektrum wieder, denn im Spektrum wird ein Gemisch von vier Sinusschwingungen

sichtbar. Die Frequenzspektren der anderen Testpersonen ergaben ähnliche Frequenzen für die beiden Formanten. Tabelle 6-4 beinhaltet die ersten beiden Formanten und die Tonhöhe (Grundfrequenz) des gesprochenen Kardinalvokals.

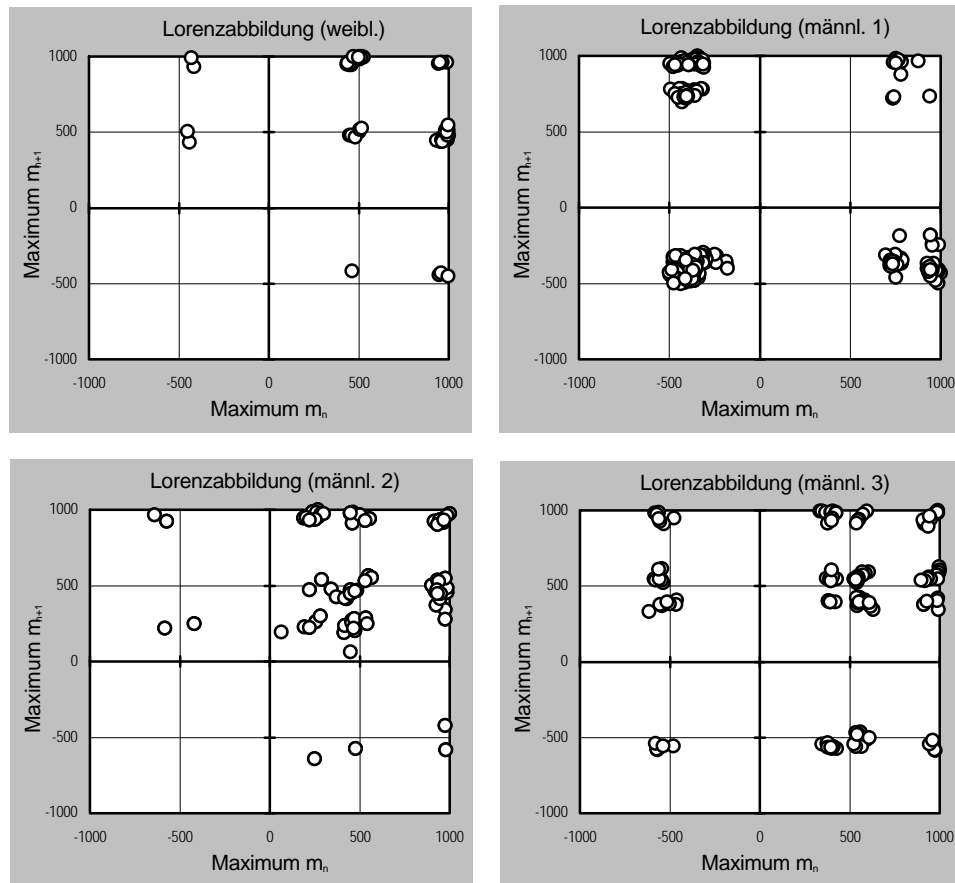
**Tabelle 6-4** Frequenzen der Formanten des Kardinalvokals [o]

	weibl.	männl. 1	männl. 2	männl. 3	Durchschnitt (männl.)
<b>Grundfrequenz [Hz]</b>	200	140	130	125	—
<b>1. Formant [Hz]</b>	550	500	500	520	$510 \pm 10$
<b>2. Formant [Hz]</b>	900	850	830	750	$810 \pm 40$

In den Lorenzabbildungen (Abbildung 6-37) zeigt sich die nahezu sinusförmige Abtastfunktion des Tons.

Die Lorenzabbildungen der weiblichen sowie der zweiten und dritten männlichen Testperson weisen Ähnlichkeiten auf. Die meisten Punkthäufungen befinden sich im ersten Quadranten zu einem Viereck gesetzt. Die Anzahl der Punkte unterscheidet diese Sprecher. Im Gegensatz zu diesen Sprechern liegen die Punkte der ersten männlichen Testperson nur in den vier Quadranten. Im ersten Quadranten zeigt sich bei ihm keine Anhäufung der Punkte zu einem Viereck.

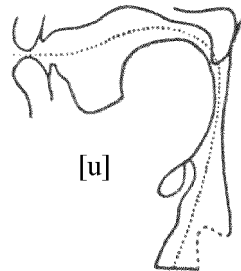




**Abbildung 6-37** Lorenzabbildungen des Vokals [o]

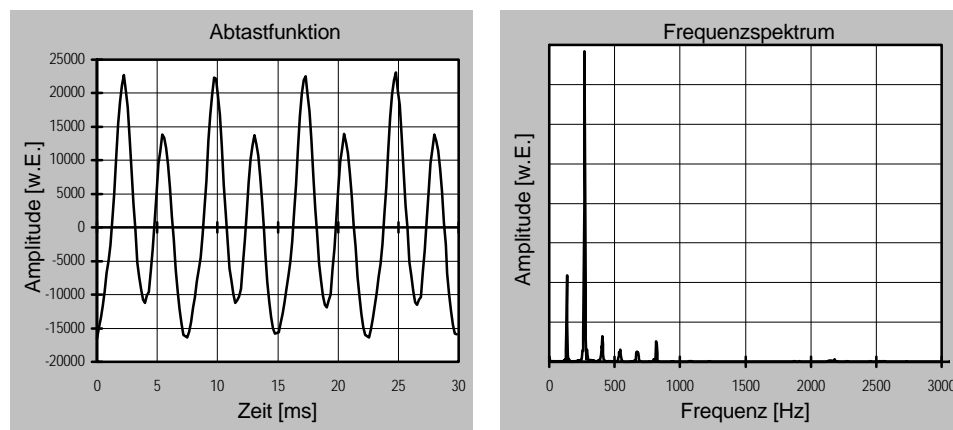
### 6.5.5 Der Kardinalvokal [u]

Der letzte Kardinalvokal [u] hat ebenfalls eine Mund- und Zungenstellung, die unverkennbar ist. Die Lippen sind nahezu geschlossen, während die Zunge weit hinten am weichen Gaumen sitzt und den Rachenraum somit verengt. Zur Veranschaulichung dient Abbildung 6-38.



**Abbildung 6-38** Mundstellung für die Artikulation des Kardinalvokals [u] nach Fant [3]

Abbildung 6-39 beinhaltet exemplarisch die Abtastfunktion und das Frequenzspektrum des Kardinalvokals [u]. In der Abtastfunktion zeigt sich ein vergleichbarer Kurvenverlauf wie beim Vokal [o]. Der Amplitudenausgang des zweiten Maximums ist jedoch erheblich größer als beim [o]. Das Frequenzspektrum stellt den charakteristischen Verlauf des Kardinalvokals [u] dar. Zwischen den Formanten liegen ungefähr 600 Hz, denn der erste Formant liegt bei 250 Hz und der zweite bei ca. 850 Hz.



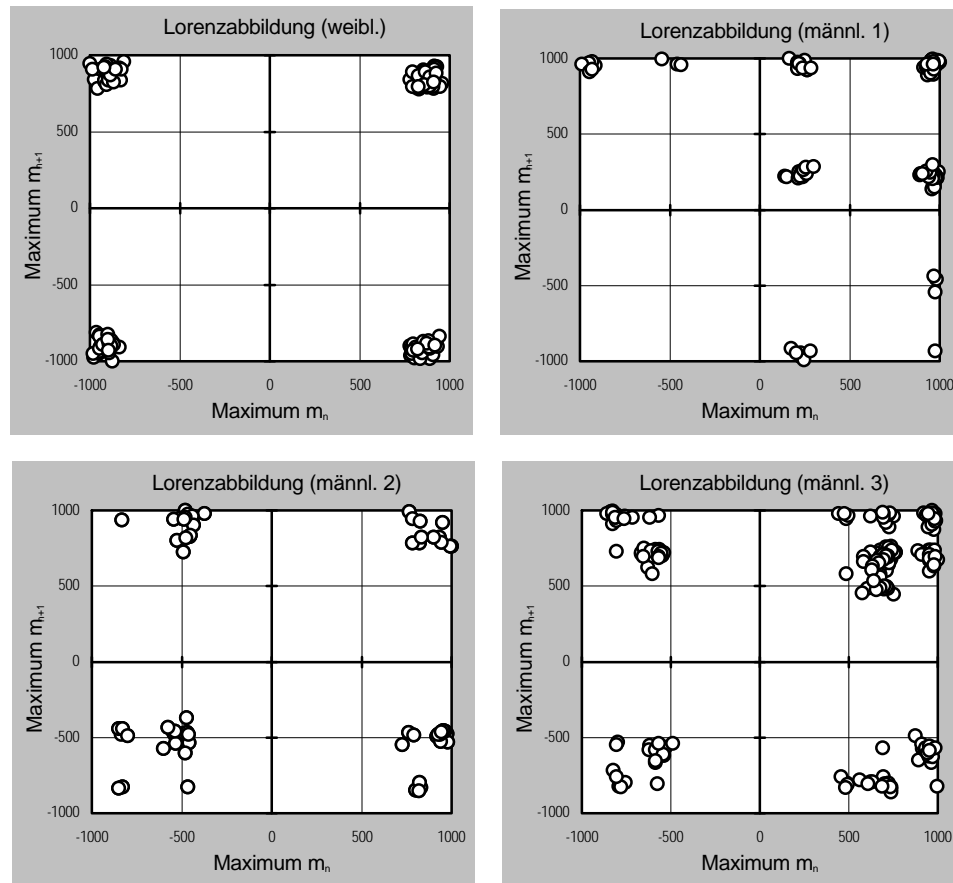
**Abbildung 6-39** Abtastfunktion und Frequenzspektrum des Vokals [u] (männl. 3)

Um das gefundene Ergebnis zu untermauern, werden die Frequenzen der Formanten der anderen Testpersonen in Tabelle 6-5 zusammengefasst.

**Tabelle 6-5** Frequenzen der Formanten des Kardinalvokals [u]

	weibl.	männl. 1	männl. 2	männl. 3	Durchschnitt (männl.)
<b>Grundfrequenz [Hz]</b>	200	140	140	130	—
<b>1. Formant [Hz]</b>	350	280	270	250	$270 \pm 10$
<b>2. Formant [Hz]</b>	940	790	800	850	$810 \pm 30$

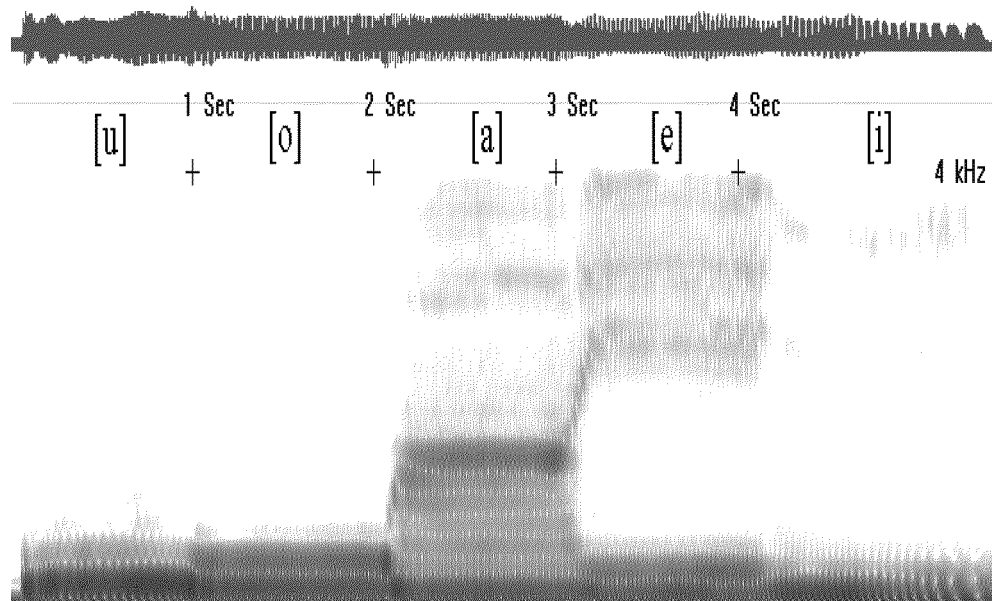
Bei den Lorenzabbildungen der weiblichen Testperson und der männlichen Probanden zwei und drei zeigen sich Ähnlichkeiten in der Abbildung. Charakteristisch für die weibliche Person sind die Punktanhäufungen in allen vier Ecken der Lorenzabbildung. Bei den beiden männlichen Testpersonen sind die Punktanhäufungen auch in den Ecken zu finden, doch sind die Anhäufungen nicht so zentriert wie bei der weiblichen Testperson, sondern fransiger. Bei dem zweiten männlichen Probanden sind immer vier markante Punktanhäufungen zu sehen. In der Lorenzabbildung des dritten männlichen Probanden haben sich diese Anhäufungen von Punkten zerteilt. Verwunderlich erscheint jedoch die Lorenzabbildung des ersten männlichen Probanden. Sie zeigt die Lorenzabbildung des Vokals [o] der anderen drei Probanden. Man könnte nun vermuten, daß die Lorenzabbildungen der Vokale [o] und [u] des ersten männlichen Probanden versehentlich vertauscht wurden, doch trifft das nicht zu. Wenn man sich die WAV-Datei des Vokals [u] im Aufnahmeprogramm anhört, nimmt man eindeutig diesen Vokal wahr. Diese Besonderheit macht den ersten männlichen Probanden aus. In Abbildung 6-40 werden die Merkmale der Lorenzabbildungen jeder Testperson bildlich dargestellt.



**Abbildung 6-40** Lorenzabbildungen des Vokals [u]

### 6.5.6 Sonagramm der Kardinalvokale

In den letzten Kapiteln wurden die Formantenfrequenzen der einzelnen Kardinalvokale ermittelt. Um die Formantstellungen der Vokale in zeitlicher Abfolge zu präzisieren, werden die Kardinalvokale schließlich in einem Breitband-Sonagramm in Abbildung 6-41 dargestellt.

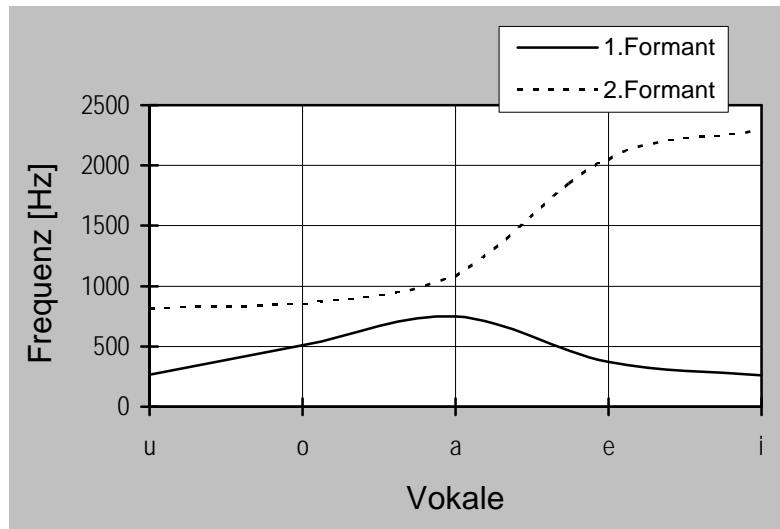


**Abbildung 6-41** Breitband-Sonagramm der Vokale [u], [o], [a], [e], [i]

Vergleicht man die Zungenstellungen der beiden Vokale [e] und [i], so kann man einen Zusammenhang zwischen den Zungenstellungen und dem ersten Formanten finden. Je höher die Zungenstellung im Mundraum ist, desto niedriger ist die Frequenz des ersten Formanten. Noch deutlicher zeigt sich dieser Zusammenhang bei den Vokalen [a] und [i]. Während der Artikulation hat das [a] eine sehr flache Zungenlage im Mundraum und deswegen einen ersten Formanten mit hoher Frequenz.

Der zweite Formant ist korreliert mit der horizontalen Zungenposition. Vom Vokal [u], der eher im hinteren Bereich des Mundes gebildet wird, bis zum Vokal [i] erhöhen sich die Frequenzen des zweiten Formanten ständig.

Um das systematische Verhalten der beiden ersten Formanten in einen Zusammenhang zu bringen, wird in Abbildung 6-42 diese Beziehung als eingängiges Bild dargestellt. Diese Abbildung beinhaltet die Werte aus den Frequenzspektren der männlichen Testpersonen.



**Abbildung 6-42** Schema der Formantenbewegung der Vokale

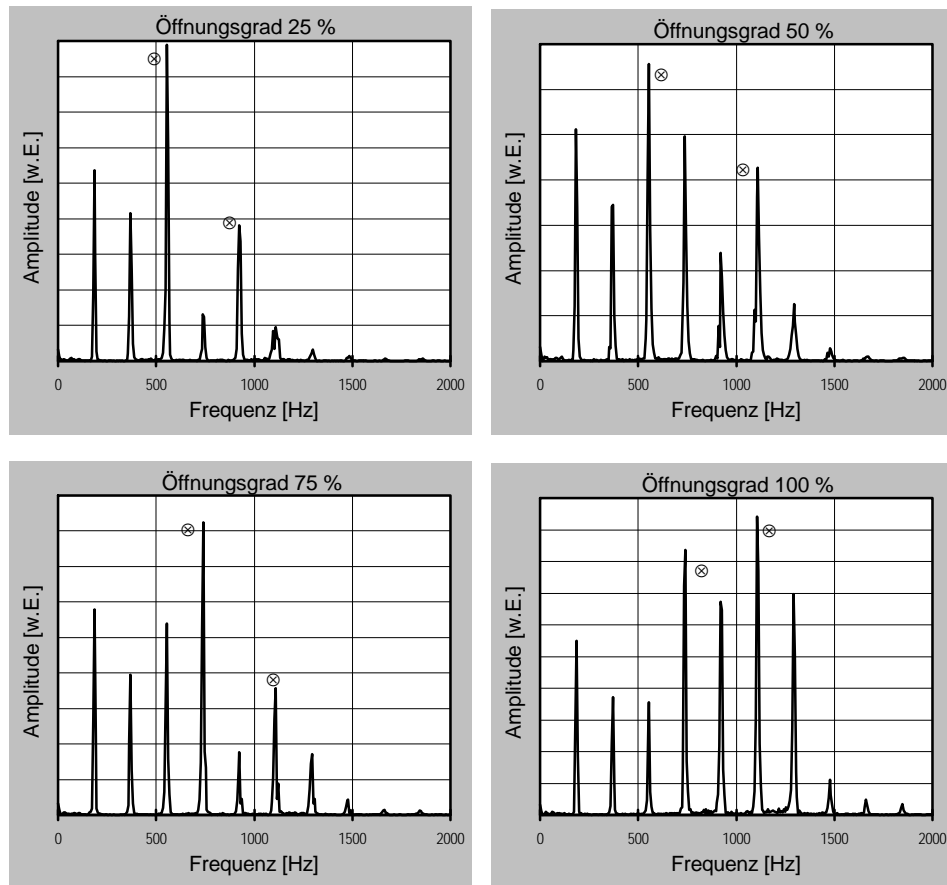
## 7 Untersuchungen am Stimmapparat

Wenn der Mensch spricht, sind die Veränderungen des Stimmapparats sehr komplex. Möchte man mit dem Stimmapparat experimentieren, so sind gezielte Änderungen im Rachenraum notwendig. Nur durch die Änderung eines einzelnen Parameters des Stimmapparats kann man die Auswirkungen auf den Ton untersuchen. In den nachfolgenden Experimenten wird einmal nur die Öffnung des Mundes geändert. Ein weiteres Experiment beschäftigt sich mit der Variation der Tonhöhen. Die letzten beiden Versuche behandeln die Stimme einer Opernsängerin und die Auswirkung des Einatmens von Helium.

### 7.1 Öffnen und Schließen des Mundes

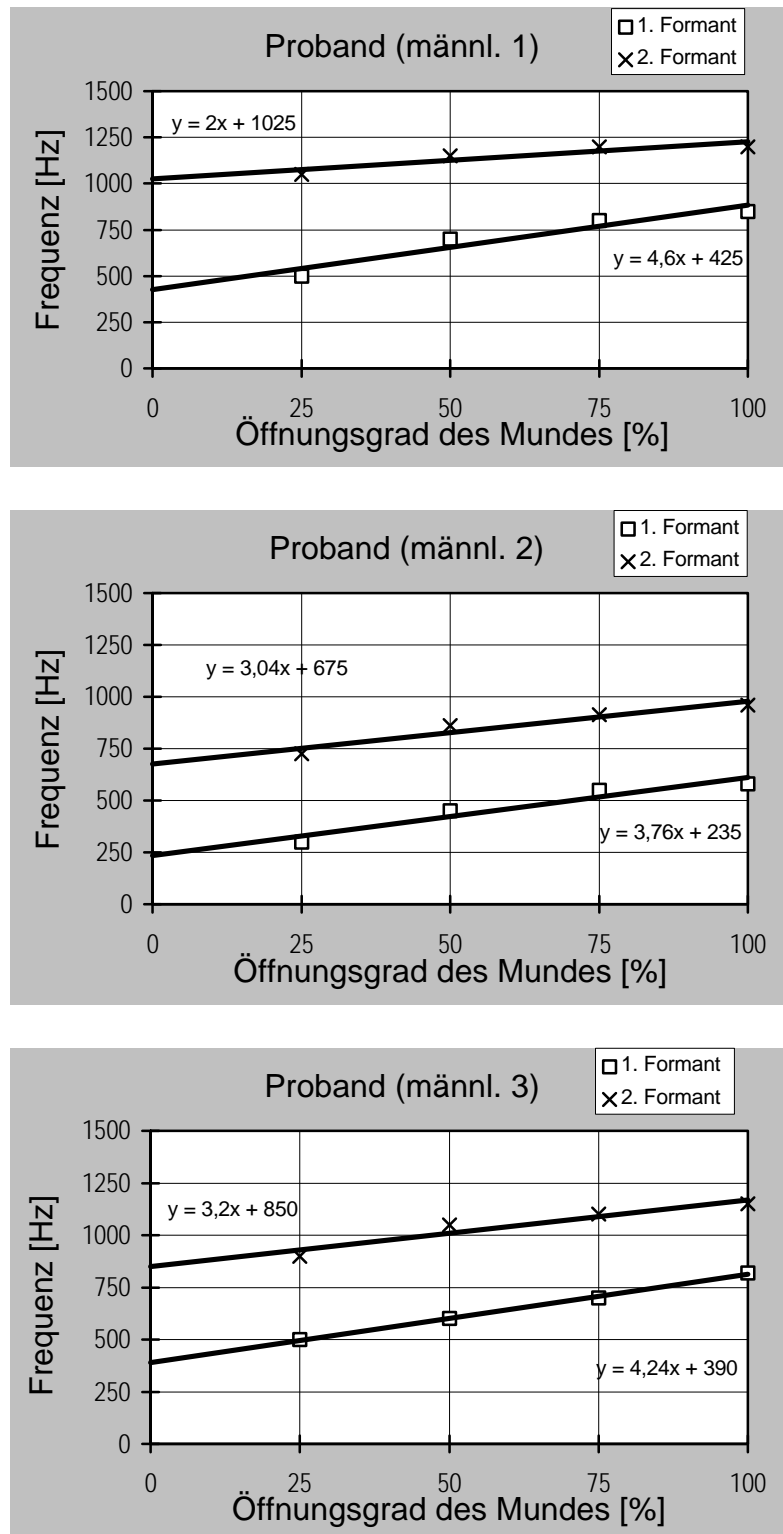
Die erste Veränderung, die am menschlichen Stimmapparat vorgenommen wurde, war das Öffnen und Schließen des Mundes. Die Probanden sollten den Kardinalvokal [a] mit unterschiedlich weit geöffnetem Mund sprechen. Die Auswirkung der Verformung des Rachenraums auf das Frequenzspektrum des Vokals wird in Abbildung 7-1 dargestellt. Durch die Prozentangaben wird der

ungefähre Öffnungsgrad des Mundes angezeigt. Bei 100 % ist der Mund nahezu vollständig geöffnet. Die Abnahme der Prozentangabe soll die Schließung des Mundes verdeutlichen. Zur Vereinfachung sind die Positionen der Formanten durch einen gekreuzten Kreis gekennzeichnet.



**Abbildung 7-1** Frequenzspektren des Vokals [a] beim Öffnen des Mundes (männl. 3)

Die Grundfrequenz, die durch die Schwingungen der Stimmlippen hervorgerufen wird, ändert sich während dieses Experiments nicht, weil die Stimmlippen an der Variation des Tons nicht beteiligt sind. Beobachtet man die Lage der Formanten, so erkennt man, daß sich ihre Frequenz beim Öffnen zu größeren Werten hin verschiebt. Diesen Zusammenhang kann man gut sehen, wenn man die Lage der Formanten in ein Diagramm aufträgt. In Abbildung 7-2 findet sich die Auswertung der drei männlichen Probanden. Die Formanten der weiblichen Testperson ließen sich nicht genau festlegen, weil die Dichte der Stützpunkte für die einhüllende Filterfunktion zu klein war.

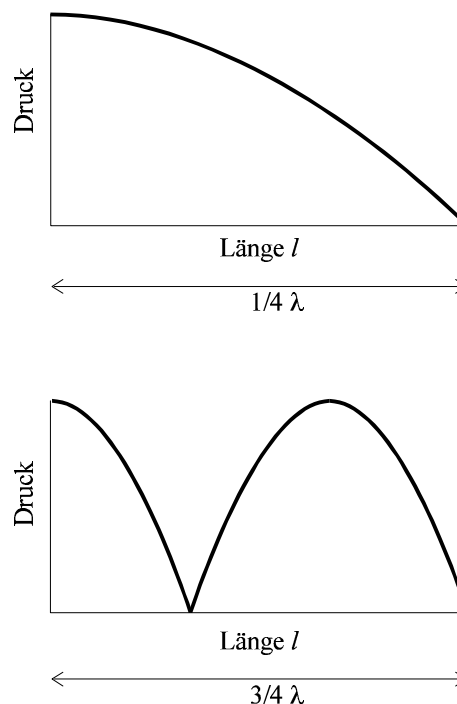


**Abbildung 7-2** Formantenfrequenz in Abhängigkeit vom Öffnungsgrad des Mundes



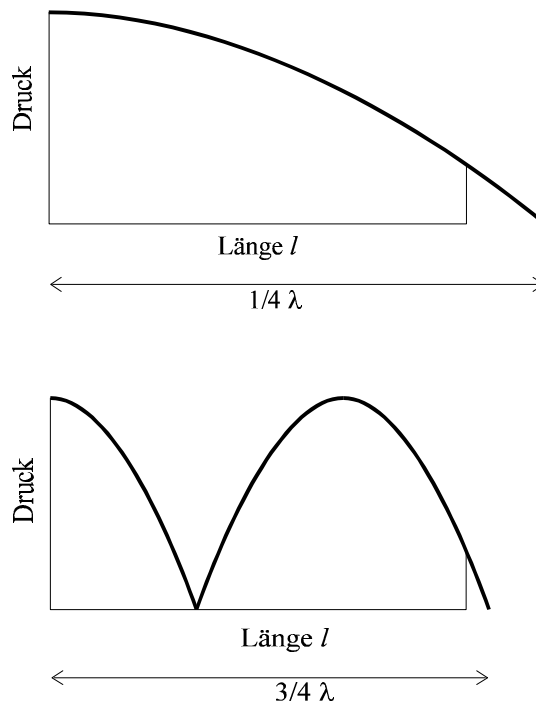
Man kann einen linearen Zusammenhang zwischen dem Öffnungsgrad des Mundes und den Formantenfrequenzen annehmen. Der erste Formant verschiebt sich stärker als der zweite, was man auch in den Gleichungen der Regressionsgeraden, die neben den Graphen stehen, erkennen kann. Beim ersten männlichen Probanden fällt besonders auf, daß der zweite Formant bei geöffnetem Mund einen sehr hohen Wert besitzt. Das läßt darauf schließen, daß der Mund dieses Probanden sehr weit geöffnet war. Den Öffnungsgrad von null Prozent kann man eigentlich gar nicht erreichen, weil das bedeuten würde, daß der Mund völlig geschlossen ist. Die Ausgleichsgerade ermöglicht die Interpolation dieses Punktes.

Aus welchem Grund aber wandern die Formanten? Zur Beantwortung dieser Frage dient das einfache Beispiel der einseitig geschlossenen Orgelpfeife. Abbildung 7-3 zeigt die Druckverteilung der Grundschiwingung und der ersten Oberschwingung in einer einseitig geschlossenen Orgelpfeife. Der Druck am offenen Ende des Rohrs ist null.



**Abbildung 7-3** Druckverteilung in einem einseitig geschlossenen Rohr

Wird das Rohr am offenen Ende nun ein wenig eingeschnürt, ist der Druck an dieser Stelle von Null verschieden. Der Druckknoten muß an die Stelle wandern, an der der Druck wieder null sein kann. Aus diesem Grund verlagert sich der Druckknoten nach außen hin (Abbildung 7-4).

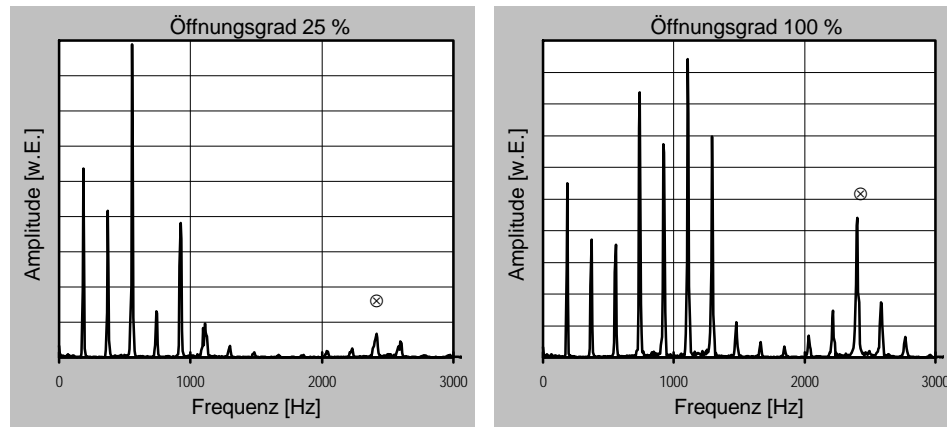


**Abbildung 7-4** Auswirkung der Druckänderung am offenen Ende eines einseitig geschlossenen Rohrs auf die Wellenlänge

Für die Grundfrequenz der Orgelpfeife, die dem ersten Formanten des Stimmapparats entspricht, wandert der Druckknoten weit von der ursprünglichen Stelle weg (Abbildung 7-4 oben). Da die Druckvariation der ersten Oberschwingung in der Nähe der Einengung stärker ist, muß sich der Druckknoten hier weniger weit nach außen hin verlagern (Abbildung 7-4 unten). So läßt sich erklären, warum die Frequenz des ersten Formanten stärker ansteigt als die des zweiten. Durch die scheinbare Verlängerung des Ansatzrohres wird die Wellenlänge der Grundschiwingung größer und somit die Frequenz der Formanten kleiner.

Schon bei diesem einfachen Modell sieht man, daß man die feste Wellenlängenbeziehung der Orgelpfeife (siehe Kapitel 2.2.3) nicht unbedingt auf den menschlichen Stimmapparat übertragen kann. Wie man sieht, ist die Wellenlänge der Grundschiwingung in Abbildung 7-4 nicht mehr das dreifache der Wellenlänge der Grundschiwingung. Dieses Verhalten entspricht nicht dem, das man von der Orgelpfeife erwarten würde.

Sollte sich die Theorie des Modells bestätigen, müßte sich der dritte Formant kaum aus seiner Lage bewegen. Abbildung 7-5 zeigt die Frequenzspektren des dritten männlichen Probanden mit fast geschlossenem und mit offenem Mund. Die Lage des dritten Formanten ist wieder gekennzeichnet.



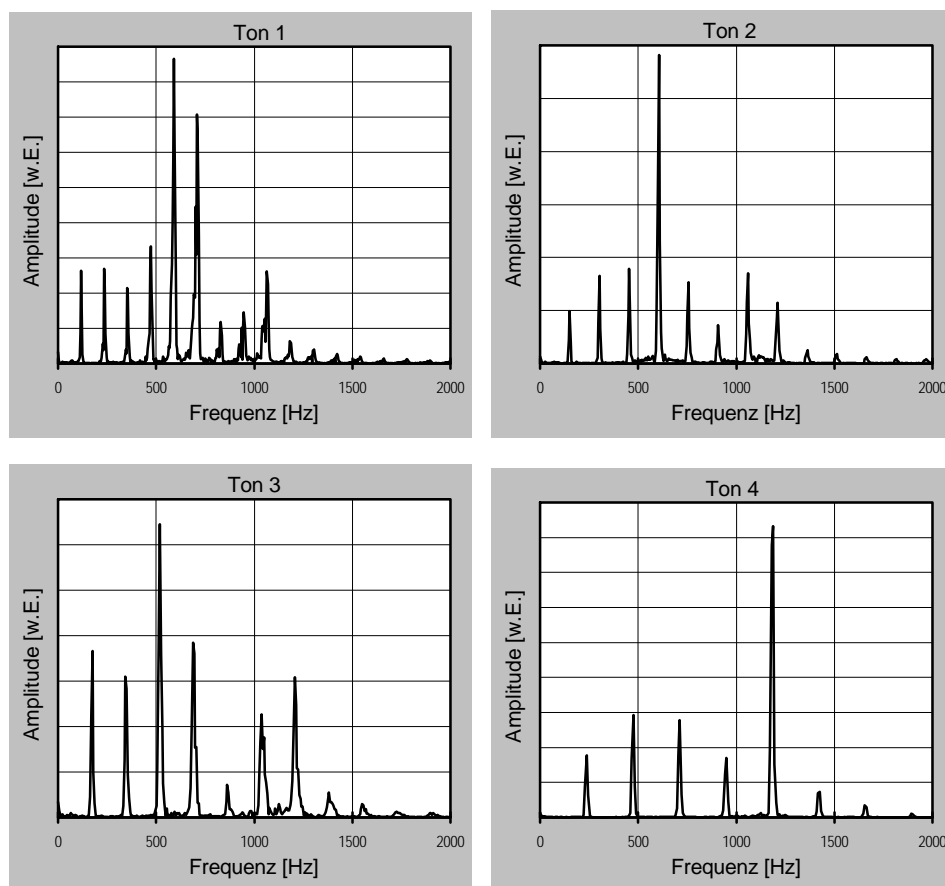
**Abbildung 7-5** Lage des dritten Formanten (männl. 3)

Wie man in den Bildern sieht, bewegt sich der dritte Formant kaum. Man kann also davon ausgehen, daß dieses einfache Modell zur Beschreibung dieses Effekts geeignet ist. Es läßt allerdings keine quantitativen Aussagen über die Verschiebung der Formanten zu, sondern dient zum qualitativen Verständnis.

Grundzüge dieser Theorie finden sich bei Lieberman und Blumenstein [12]. Zur Beschreibung der Längenänderung des zweiten und dritten Formanten mußte der Erklärungsansatz zur Druckverteilung in einem einseitig geschlossenen Rohr erweitert werden.

## 7.2 Variation der Tonhöhe

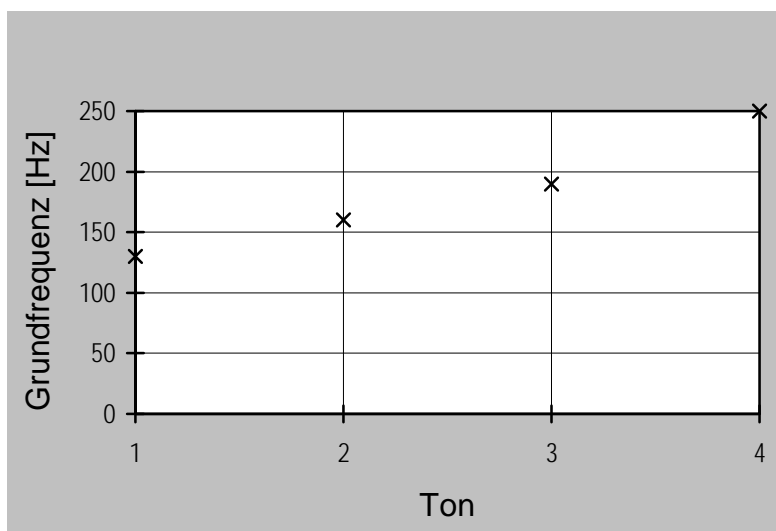
Dieses Experiment beschäftigt sich mit der Variation der Tonhöhe eines Kardinalvokals. Die Probanden sollten den Vokal [a] in vier verschiedenen Tonhöhen sprechen, die ich vorsprach. Die Abfolge der Tonhöhenänderung wies eine steigende Tendenz auf, wobei die Form des Ansatzrohrs nicht verändert werden sollte. Die gemessenen Frequenzspektren sind in Abbildung 7-6 dargestellt. Die Variation begann mit dem niedrigsten Ton, der in der folgenden Abbildung mit Ton (1) bezeichnet wird. Die steigende Nummer der Töne bezeichnet den Anstieg in der Tonhöhe.



**Abbildung 7-6** Variation der Tonhöhe des Vokals [a] (männl. 2)

Die Frequenzspektren zeigen, daß sich die Dichte der Stützpunkte zu größeren Tonhöhen hin immer weiter verringert. Die Lage der ersten beiden Formanten ändert sich hingegen kaum. Der erste Formant liegt bei dieser Testperson im Durchschnitt bei 600 Hz und der zweite bei 1070 Hz. Die Schwankungen der Formantenfrequenzen liegen bei höchstens 50 Hz. Da sich die Stützpunkte

verringern, kann man den Wert der Formanten bei den größeren Tonhöhen nicht mehr exakt bestimmen. Was sich aber in den Frequenzspektren ändert, ist die Grundfrequenz des gesprochenen Kardinalvokals [a]. Die Grundfrequenz erhöht sich mit steigender Tonhöhe. Dieser Zusammenhang wird in Abbildung 7-7 dargestellt.



**Abbildung 7-7** Grundfrequenzänderung mit steigender Tonhöhe (männl. 2)

Das Zustandekommen der verschiedenen Spektren läßt sich wie folgt erklären. Die Lage der Formanten ändert sich kaum, da sich der Resonanzraum bei der Tonhöhenvariation nicht ändert. Die leichten Schwankungen in den Frequenzen der Formanten entstehen dadurch, daß die Probanden das Ansatzrohr leicht geändert haben. Die Stimmbänder schwingen bei ansteigender Tonhöhe schneller, weil der Druck hinter den Stimmlippen steigt. Somit erhöht sich die Grundfrequenz der Stimmbandschwingungen und die Grundfrequenz des Frequenzspektrums.

Bei den anderen Probanden ergaben sich ähnliche Ergebnisse. Aus diesem Grund werden die Frequenzspektren nicht abgebildet. In Tabelle 7-1 sind nur die Änderung der Grundfrequenzen beschrieben.

**Tabelle 7-1** Änderungen der Grundfrequenzen durch Tonhöhenänderung

	weibl.	männl. 1	männl. 2	männl. 3
<b>Ton 1 [Hz]</b>	180	100	130	130
<b>Ton 2 [Hz]</b>	220	125	160	170

<b>Ton 3 [Hz]</b>	260	160	190	200
<b>Ton 4 [Hz]</b>	350	200	250	260

Wenn man die Abfolge der Grundfrequenzänderungen genauer betrachtet, lassen sich Zusammenhänge zwischen den Tonhöhen der verschiedenen Töne feststellen. Die Frequenz von Ton 2 und Ton 1 bewegen sich mit einer Genauigkeit von 10 % in einem Verhältnis von 5/4. Die Frequenzen des dritten Tons ist hingegen 3/2 zu der Frequenz des ersten Tons. Außerdem ist die Frequenz des vierten Tons das doppelte der Frequenz des ersten Tons. Dieser Zusammenhang kann mit dem Frequenzverhältnis einer diatonischen Tonleiter<sup>9</sup> verglichen werden. Ton 1 entspricht dem ersten Ton dieser Tonleiter, der Prime („c“). Der Ton 2 liegt in der Nähe der Terz („e“). Die Terz ist der dritte Ton der diatonischen Tonleiter. Ton 3 besitzt die gleichen Frequenzverhältnisse zur Prime wie die Quinte („g“), dem fünften Ton der Tonleiter und Ton 4 ist die Oktave zum Ton 1 und somit das „c“. Ihr Frequenzverhältnis zur Prime beträgt 2:1.

---

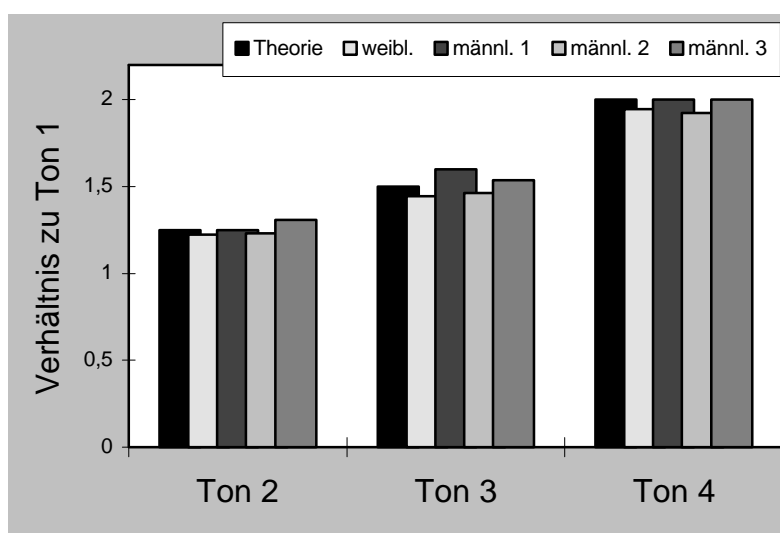
<sup>9</sup> Bezeichnung für ein Tonsystem, das aus einer Kombination von Ganz- und Halbtonschritten bestehen.

In Tabelle 7-2 sind die Frequenzverhältnisse der einzelnen Töne dargestellt.

**Tabelle 7-2** Aufbau der Frequenzverhältnisse der Töne

Ton	Name des Tons	Intervallbezeichnung	Frequenzverhältnis bezogen auf C
1	c	Prime	1
2	e	Terz	5/4
3	g	Quinte	3/2
4	c	Oktave	2/1

Die diatonische Tonleiter ist darauf ausgelegt relativ harmonisch klingende Tonsysteme zu schaffen. Ein wohlklingender Ton ist angenehmer anzuhören als ein dissonant klingender. Aus kulturellen Gründen wählen wir sehr oft die diatonische Tonleiter.



**Abbildung 7-8** Verhältnisse des theoretischen und gemessenen Wert der Töne zu Ton 1

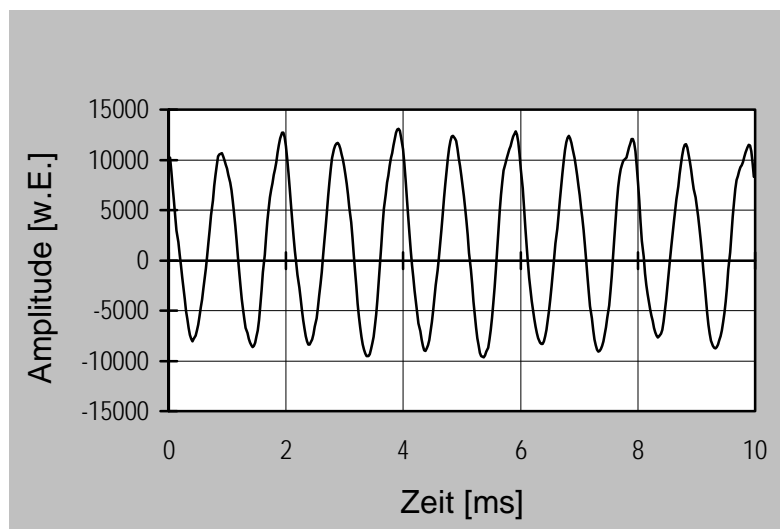
Abbildung 7-8 faßt noch einmal die Änderung der Grundfrequenzen durch die Tonhöhenänderung aus Tabelle 7-1 zusammen. Im Säulendiagramm werden die Verhältnisse des theoretischen Werts der Frequenzverhältnisse mit den gemessenen Werten verglichen. Der Vergleich zeigt, daß die gemessenen Werte sehr nah am theoretischen Wert liegen.

### 7.3 Die Singstimme

Die Singstimme beruht auf demselben Erzeugungsmechanismus wie die Sprechstimme. Doch worin unterscheiden sich dann diese zwei Stimmarten? Einige Besonderheiten der Gesangsstimme werden anhand eines Beispiels erläutert.

Die von mir ausgesuchte Gesangsstimme ist die von Maria Callas. Sie ist eine Opernsängerin, deren Stimme im italienischen Stil ausgebildet wurde. In der Regel singen Opernsänger mit Orchesterbegleitung und gerade das verhindert eine Analyse der Gesangsstimme. Es mußte also eine Oper gefunden werden, wo die Sängerin teilweise ohne Orchesterbegleitung singt. Solche Passagen finden sich zum Beispiel im Stück „La mamma morta“ aus der Oper „Andrea Chenier“<sup>10</sup> von Umberto Giordano.

Maria Callas singt in diesem Stück einen Vokal, der eine völlig andere Form aufweist, als wenn er gesprochen würde. In Abbildung 7-9 ist ein Ausschnitt der Abtastfunktion des gesungenen Kardinalvokals [a] dargestellt. Im Stück findet man diesen Ausschnitt nach ungefähr 75 s.



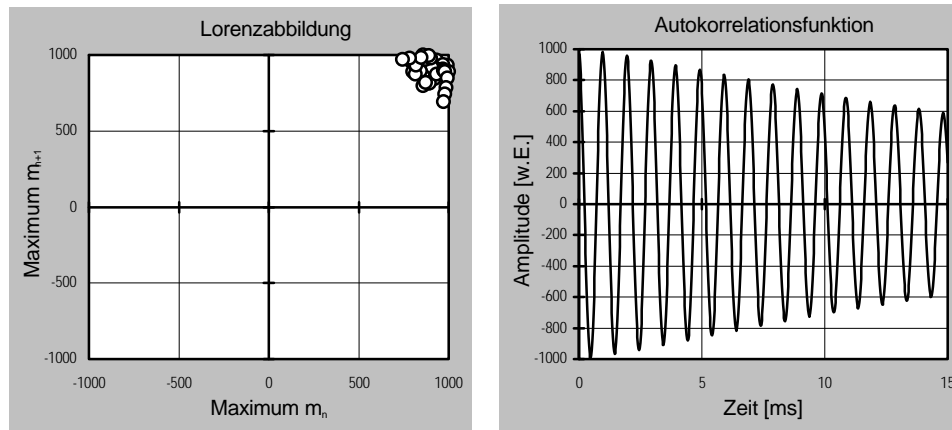
**Abbildung 7-9** Abtastfunktion eines gesungenen Vokals [a]

Die Abtastfunktion läßt die Struktur des Vokals [a] kaum erkennen, denn es zeigt sich eine fast sinusförmige Funktion. Trotz dieses Erscheinungsbildes

<sup>10</sup> © EMI-Records



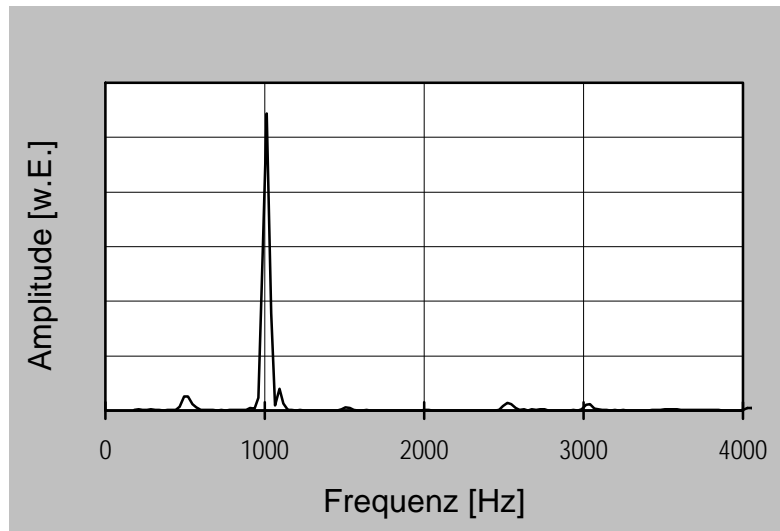
nimmt man den gesungenen Vokal als [a] wahr. In der Autokorrelationsfunktion und in der Lorenzabbildung lassen sich Ähnlichkeiten zum sinusförmigen Verlauf finden. Abbildung 7-10 stellt die Ergebnisse der beiden Analyseverfahren dar.



**Abbildung 7-10** Darstellungsarten des gesungenen Vokals [a]

Die Punkthäufung in der Lorenzabbildung tritt lediglich in der rechten oberen Ecke auf. Das heißt, daß alle Maxima der Abtastfunktion oberhalb der Nulllinie liegen und etwa gleich hoch sein müssen. In der Autokorrelationsfunktion zeigt sich ein hohes Maß an Selbstähnlichkeit der Abtastfunktion. Der Idealfall eines linearen Abfalls der Funktion ist annähernd erreicht. Es treten kaum Lautstärkeschwankungen im gesungenen [a] auf, und die Unterschiede in den Frequenzen des Schalls sind ebenfalls unwesentlich.

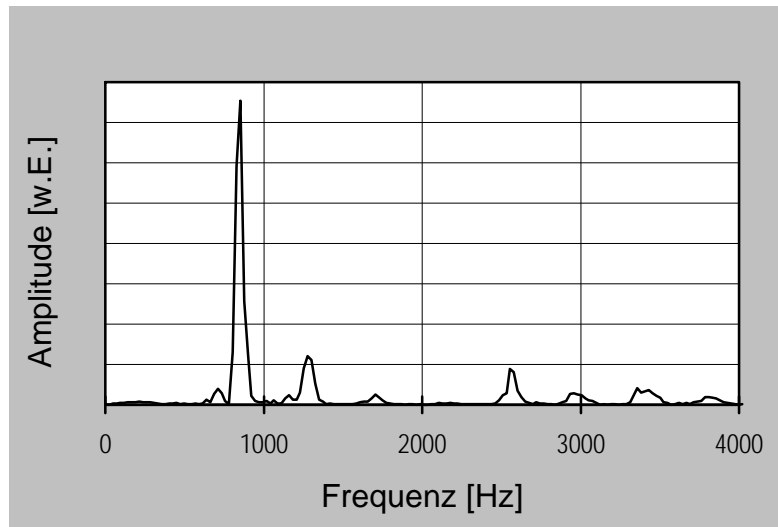
Im Frequenzspektrum in Abbildung 7-11 sieht man, daß sich ein starkes Maximum bei 1000 Hz ausgebildet hat. Neben diesem Hauptmaximum sind nur sehr kleine Nebenmaxima. Auch hier sieht man den sinusartigen Verlauf der Abtastfunktion. Die Entstehung für diesen Klang wird hauptsächlich durch die Stimmlippen beeinflusst. Je straffer die Stimmlippen gespannt sind, desto höher ist der erzeugte Ton und die Stimmbänder schwingen schneller.



**Abbildung 7-11** Frequenzspektrum des gesungenen Vokals [a]

In der Musik wird diese Art der Stimmführung als Kopfstimme bezeichnet.[1,8,14] Bei ihr sind die Stimmlippen so straff gespannt, daß sich nur noch die Ränder bewegen können. Die Masse der bewegten Teile nimmt ab, die Stimmlippen schwingen viel schneller, und die Frequenz des gesungenen Tons wird höher. Aufgrund der schnellen Stimmlippenschwingungen ist es schwierig in der Abtastfunktion genau eine Periode auszumachen. Da sich die Stimmlippen bei dieser Art der Stimmführung nicht mehr vollständig schließen, nehmen die Maxima im oberen Bereich des Frequenzspektrums ab. Die Kopfstimme erscheint deshalb als eine sehr dünne Stimme. Zusätzlich ist ein Formant fast oder genau auf der 2. Harmonischen (1000 Hz). Das hat zur Folge, daß die 2. Harmonische maximal verstärkt wird und gegenüber den anderen Maxima stark überhöht erscheint.

Eine weitere Art der Stimmführung wird als Bruststimme bezeichnet.[1,8,14] Weil bei der Bruststimme die Stimmlippen nicht so stark gespannt sind, können sie durchschwingen. Außerdem schließen sich die Stimmlippen bei der Bruststimme. Der dadurch erzeugte Ton ist tiefer und obertonreicher, denn je länger die Verschlußphase der Stimmlippen dauert, desto mehr Obertöne finden sich im Frequenzspektrum. In Abbildung 7-12 wird dieser Zusammenhang dargestellt. Die Abtastfunktion, die zu diesem Frequenzspektrum führt, findet man wieder im Stück „La mamma morta“ bei 70 s.

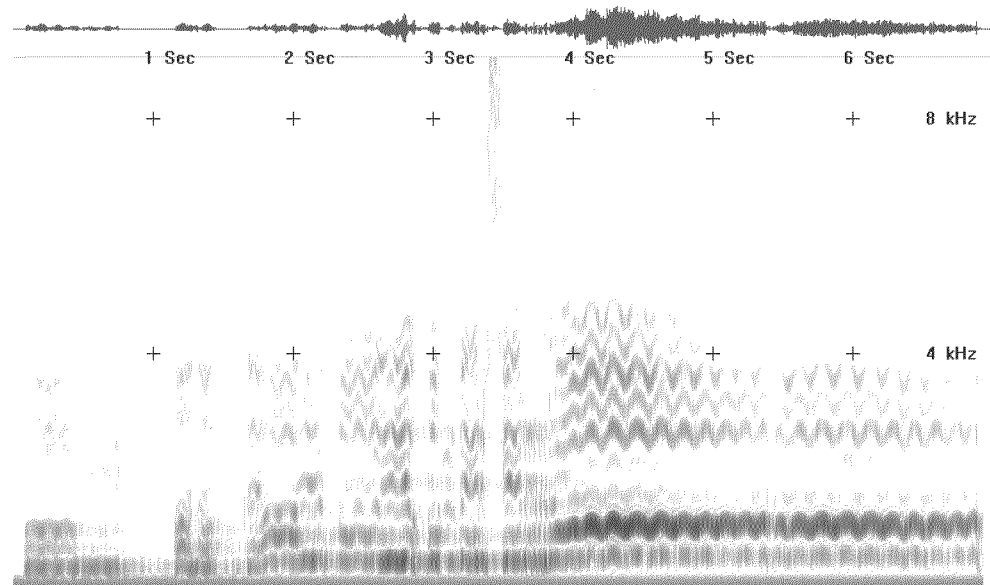


**Abbildung 7-12** Obertonreiches Frequenzspektrum einer Bruststimme [a]

Ferner sieht man im oberen Frequenzbereich ein ausgeprägteres Maximum, das bei 2500 Hz liegt. In der Literatur wird dieses Maximum als „Singformant“ bezeichnet. Der Singformant hängt mit der Lage des Kehlkopfes zusammen. Er liegt beim Singen viel tiefer als beim Sprechen. Das hat zur Folge, daß sich der Kehlkopfraum oberhalb der Stimmlippen und der Rachenraum erweitert. Dieser besitzt charakteristische Eigenresonanzen, die den Bereich zwischen 2500 Hz und 3500 Hz besonders verstärken (Abbildung 7-12). [1,8,15]

Eine letzte Besonderheit, die bei der Sängerstimme auftritt, ist das Vibrato.[4,9] Es fällt auf, wenn man einen Teil des Stücks „La mamma morta“ als Sonagramm darstellt. Das Vibrato wird durch eine periodische Modulation der Frequenz oder der Amplitude bewirkt. Einen endgültigen Erklärungsansatz für die Ursache des Vibratos gibt es nicht. Die Ursache für ein Tonhöhenvibrato kann in den Schwingungsschwankungen der Stimmlippen begründet sein. Liegt ein Klangfarbenvibrato vor, so kann das mit einer Schwankung des Vokaltrakts zusammenhängen. [1,8]

In Abbildung 7-13 sieht man ein Frequenzvibrato. Besonders deutlich wird das Vibrato von der vierten bis zur sechsten Sekunde. An dieser Stelle singt Maria Callas den Vokal [a], der schon zuvor untersucht wurde.

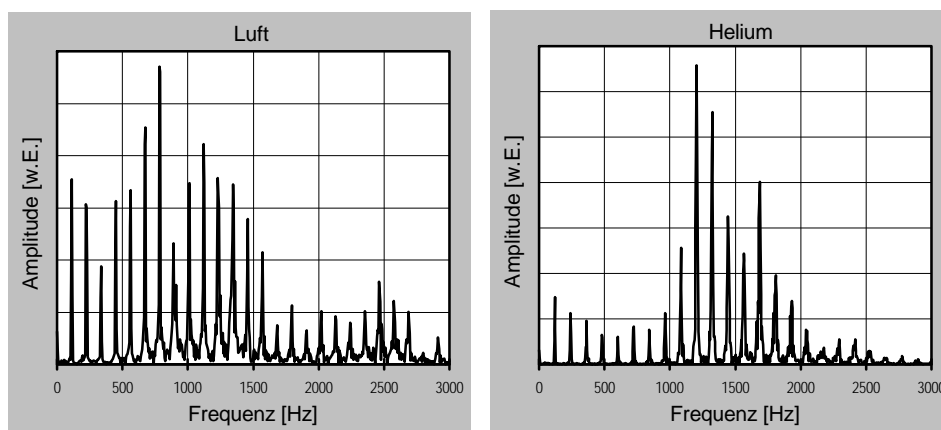


**Abbildung 7-13** Breitbandsonagramm eines Ausschnitts aus „La mamma morta“

## 7.4 Die Heliumstimme

Vielen Menschen ist die Heliumstimme bekannt, weil sie schon selbst Helium aus einem Luftballon eingeatmet haben. Taucher, die in großen Tiefen tauchen, atmen ein Helium-Sauerstoff-Gemisch. Dies dient dazu den Stickstoff, der in normaler Pressluft zu 70 % enthalten ist, durch ein Gas zu ersetzen, das nicht beim Auftauchen im Blut aufschäumen. In Reportagen über Tiefseetaucher kann man deren Helium-Stimme hören. Die Stimme wird unter Heliumeinfluß zum Teil bis zur Unkenntlichkeit verstellt. Wodurch ist dieses Phänomen zu erklären? Ein Proband (männl. 1) stellte sich für dieses Experiment zur Verfügung.

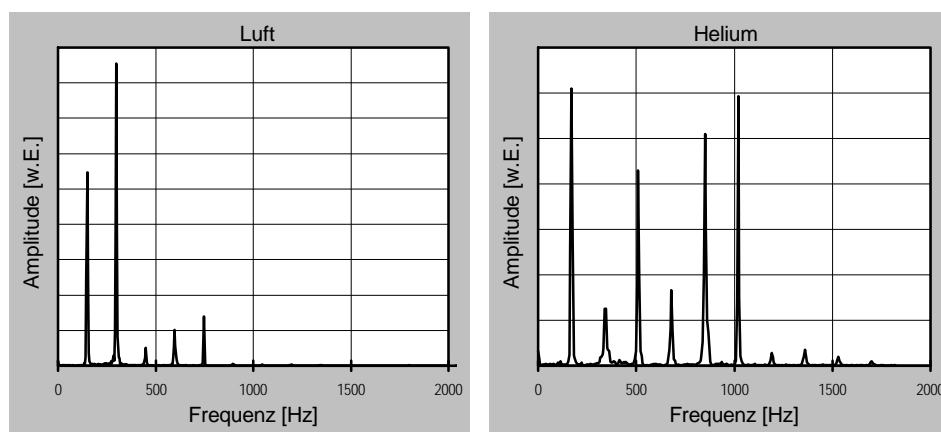
Anhand der Kardinalvokale werden die Auswirkungen des Einatmens von Helium auf die Stimme physikalisch erklärt. In Abbildung 7-14 ist rechts das Frequenzspektrum des unter Heliumeinfluß gesprochenen Kardinalvokals [a] zu sehen. Die ersten beiden Formanten liegen bei 1250 Hz und 1650 Hz. Die Grundfrequenz hat sich auf 120 Hz erhöht. Links findet sich zum Vergleich das Spektrum ohne Helium.



**Abbildung 7-14** Frequenzspektren des Vokals [a] mit Luft und Helium

Der Vergleich der beiden Frequenzspektren zeigt eine Formantenverschiebung bei Helium zu höheren Frequenzen hin, denn die Frequenzen der Formanten unter „normalen“ Bedingungen liegen bei 750 Hz und 1100 Hz.

Abbildung 7-15 zeigt die Frequenzspektren des gesprochenen Kardinalvokals [u]. Der erste Formant verschiebt sich bei Helium von 280 Hz nach 500 Hz, der zweite von 780 Hz nach 1000 Hz.



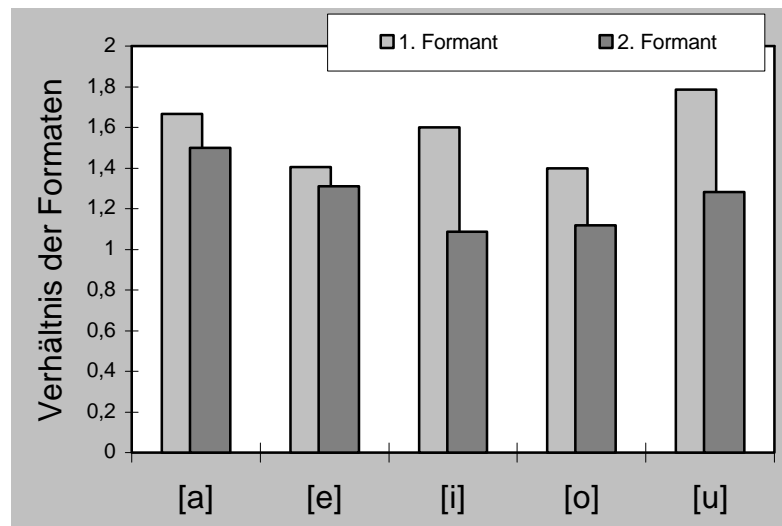
**Abbildung 7-15** Frequenzspektren des Vokals [u] mit Luft und Helium

In Helium beträgt die Schallgeschwindigkeit  $1007 \frac{m}{s}$ , im Gegensatz zu  $330 \frac{m}{s}$  in Atemluft. Da der Schall in Helium schneller ist, verkürzt sich die effektive Länge des Ansatzrohrs. Aus diesem Grund wird die Wellenlänge kleiner und die Frequenzen der Formanten größer. In Tabelle 7-3 sind Frequenzen der ersten beiden Formanten der Kardinalvokale verglichen. Einmal unter „normalen“ Bedingungen und unter Heliumeinfluß.

**Tabelle 7-3** Formantenänderung unter Einfluß verschiedener Gase

	[a]	[e]	[i]	[o]	[u]
<b>1. Formant (Luft) [Hz]</b>	750	370	250	500	280
<b>1. Formant (Helium) [Hz]</b>	1250	520	400	700	500
<b>2. Formant (Luft) [Hz]</b>	1100	2100	2300	850	780
<b>2. Formant (Helium) [Hz]</b>	1650	2750	2500	950	1000

Der Vergleich zeigt, daß die Frequenzen der beiden Formanten sich zu höheren Frequenzen hin verschieben. Bildet man das Verhältnis der Formanten unter Heliumeinfluß zu den Formanten in Luft, so ergibt sich ein über alle Vokale gemittelter Faktor von 1,4 für den ersten Formanten und 1,2 für den zweiten Formanten. Abbildung 7-16 zeigt das Verhältnis der Formanten jedes einzelnen Kardinalvokals.



**Abbildung 7-16** Verhältnis der Formanten von Helium- zu Lufteinfluß

Die Frequenzen der Formanten unter Heliumeinfluß sollten um den Faktor drei gegenüber denen unter Lufteinfluß verschoben sein, weil die Schallgeschwindigkeit in Luft  $1/3$  von der in Helium beträgt. Dieser Faktor kann aber niemals erreicht werden, denn der Schall breitet sich nicht in Helium, sondern in einem Gemisch aus Helium und Luft aus. In der Lunge und dem Ansatzrohr ist noch Restgas in Form von Luft vorhanden, außerdem desorbiert Kohlendioxyd ( $\text{CO}_2$ ) aus der Lunge. Man sieht, daß nur ein Teil der Luft in der Lunge durch einmaliges Ein- und Ausatmen durch Helium ersetzt werden kann.

Weiterhin ist auffällig, daß der zweite Formant weniger stark unter Heliumeinfluß wandert. In Abbildung 7-16 sieht man deutlich, daß die

Verhältnisse bei allen Vokalen für den zweiten Formanten kleiner sind als für den ersten. Da bei allen Messungen dies der Fall ist, muß man davon ausgehen, daß dieser Effekt nicht auf einen Meßfehler zurückzuführen ist. Leider konnte kein befriedigender Erklärungsansatz für dieses Phänomen gefunden werden.

Man sieht aber deutlich, daß, je höher die Schallgeschwindigkeit eines Gases ist, die Vokalformanten sich nach höheren Frequenzen hin verschieben.

## **8 Simulation des Stimmapparats**

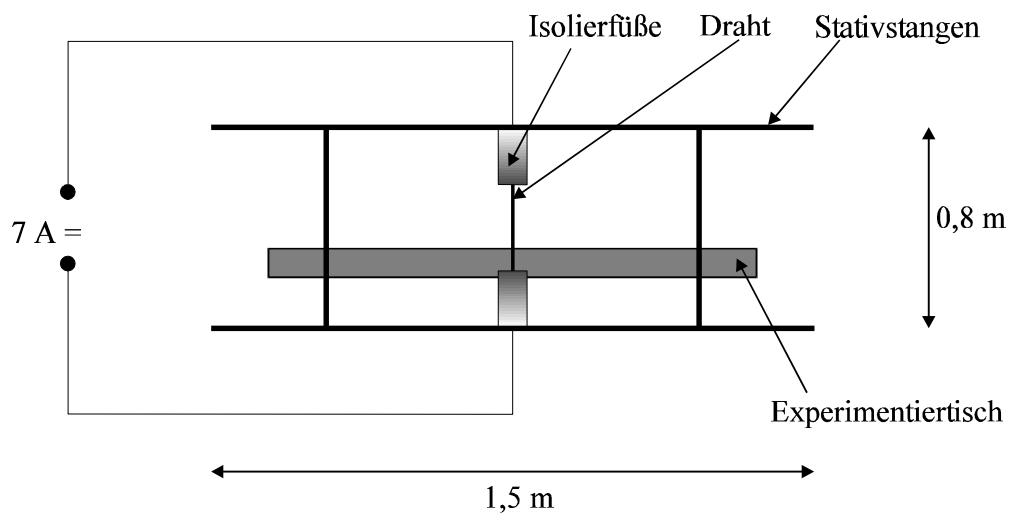
Irgendwann stößt man an Grenzen, wenn es um Experimente mit dem Stimmapparat geht. Es ist selbstverständlich nicht möglich den Stimmapparat einer lebenden Person zu öffnen, um dort gezielte Änderungen vorzunehmen. Infolgedessen ist es sinnvoll den Stimmapparat nachzubauen. An einem Experiment ist es möglich, reproduzierbare Experimente durchzuführen. Ein solches Experiment würde sicherlich auch das Interesse von Schülern wecken, denn dort könnten sie kreativ werden und handwerklich arbeiten. Dieses Projekt wäre eine Aufgabenstellung für eine Projektwoche oder eine Arbeitsgruppe.

### **8.1 Aufbau des Experiments**

Möchte man den menschlichen Stimmapparat nachbauen, so ist es nicht sinnvoll den Ton der Stimmlippen nachzuempfinden. Um die Filterfunktion des nachgebauten Rachenraums messen zu können, ist es besser, nur einzelne Sinusschwingungen durch den Versuchsaufbau filtern zu lassen und die Amplitude des durchgelassenen Tons aufzunehmen. Aus der relativen Amplitude der Töne kann man die Filterfunktion des Rachens bestimmen. Zu diesem Zweck wurde die folgende Tonsequenz synthetisiert:

Die Frequenz des Testtons wächst linear von 100 Hz bis 2000 Hz in Schritten von 10 Hz bei konstanter Amplitude. Pro Sekunde erfolgen fünf Schritte. Somit besitzt der Testton eine Länge von 38 s. Die Membran eines Autolautsprechers fungiert als Ersatz für die Stimmlippen.

Der erste Versuchsaufbau bestand aus Legosteinen, also aus Hartplastik. Mit diesem Aufbau wurden keine verwertbaren Ergebnisse erzielt, weil das Plastik den Schall fast vollständig reflektiert und kaum absorbiert. Die Absorption spielt eine große Rolle, denn der Mensch besteht zu 90 % aus Wasser und das ist ein guter Absorber für Schall. Der zweite Versuchsaufbau besteht aus diesem Grund aus Styropor, weil es besser absorbiert und preiswert ist. Ferner läßt sich dieses Material gut bearbeiten. Spannt man einen Draht senkrecht zwischen zwei Stativstangen und läßt durch ihn einen Strom von einigen Ampere fließen, so kann man mit ihm das Styropor schneiden. Durch den heißen Draht wird das Styropor nicht nur geschnitten, sondern man erhält auch glatte Schnittkanten. Der von mir benutzte Aufbau der Styroporsäge wird durch eine schematische Zeichnung in Abbildung 8-1 dargestellt.

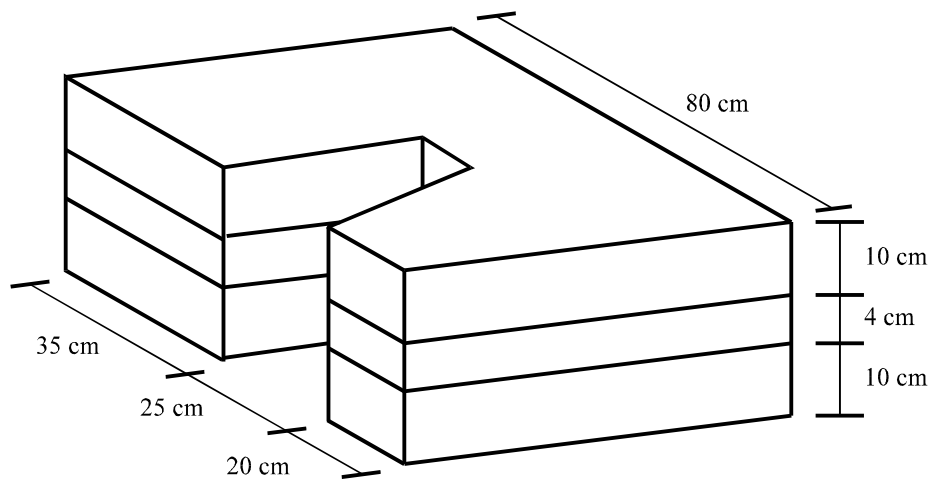


**Abbildung 8-1** Schematischer Aufbau der Styroporsäge

Da der Aufbau nur aus Schulmaterialien besteht, kann man ihn in jeder Schule nachbauen und damit arbeiten. Die Netzversorgung muß eine Strombegrenzung besitzen, weil der Draht keinen nennenswerten Widerstand hat. In meinem speziellen Aufbau hat sich ein Strom von 7 A als sinnvoll erwiesen.

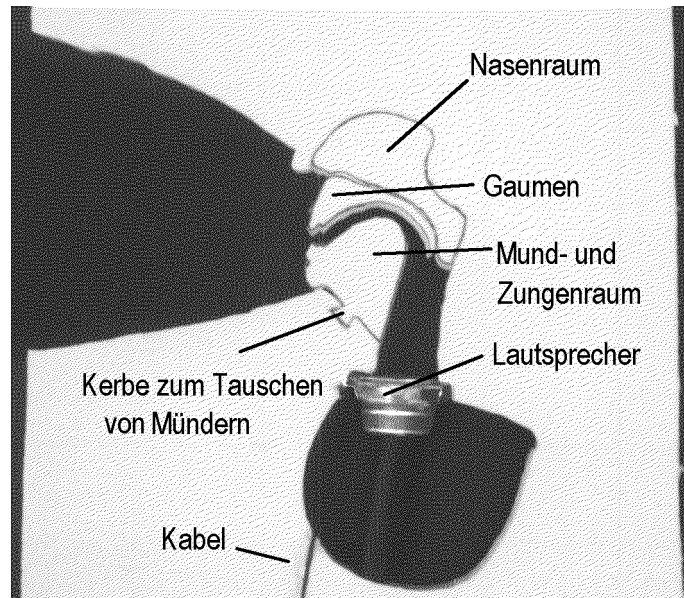


Das Experiment besteht im Prinzip aus drei Styroporplatten (Abbildung 8-2). Die obere und die untere dicke Platte dienen zur Dämpfung und Isolierung nach außen hin. Zur weiteren Dämpfung wurden kleine Filzlappen an die Innenseiten der dickeren Platten geklebt. Die dünnere, mittlere Platte beinhaltet das Ansatzrohr, das mittels einer Schablone aus der Platte herausgeschnitten wurde. Als Schablone diente eine schematische Schnittzeichnung aus [9], die auf die natürliche Größe vergrößert wurde. Um den Mund- und Nasenraum für das Experiment variabel zu gestalten, wurden verschiedenen Mund- und Gaumenstellungen der Kardinalvokale angefertigt. Diese lassen sich durch eine Auskerbung (Abbildung 8-3), ähnlich eines Puzzleteils, exakt an die richtige Position einpassen.



**Abbildung 8-2** Schematische Zeichnung des Aufbaus

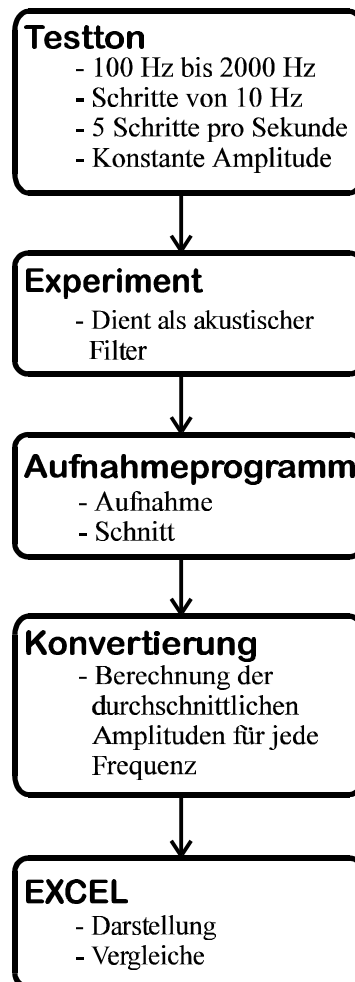
Der Lautsprecher wurde am Ende des Ansatzrohrs befestigt. Hinter dem Lautsprecher wurde ein Loch in das Styropor gesägt, damit der Lautsprecher Platz hat und frei zu schwingen kann. Die Einkerbung in Abbildung 8-3 dient dazu, daß sich der Schall nach Verlassen des simulierten Mundraums frei ausbreiten kann. Der Lautsprecher ist durch ein flaches Kabel mit einem Kassettenrecorder (Sony) verbunden, der ein Band mit dem Testton abspielt.



**Abbildung 8-3** Foto des nachgebauten Mundraums von oben

## 8.2 Meßablauf

Da der Meßablauf dieses Experiments von den zuvor beschriebenen Meßabläufen abweicht, werden in Abbildung 8-4 die Zusammenhänge des Meßablaufs durch ein Flußdiagramm dargestellt.



**Abbildung 8-4** Flußdiagramm des Meßablaufs

Weil der Testton eine Länge von 38 s hat, wurden die Aufnahmen sehr groß. Eine Aufnahme des Experiments hat einen Speicherbedarf von 3 MB. Von Interesse waren aber nur die durchschnittlichen Amplituden für jede Frequenz, und daher mußte ein kleines Konvertierungsprogramm geschrieben werden, das die einzelnen Aufnahmen in eine erträgliche Größe verkleinert. Zum Verständnis des Konvertierungsprogramms wird kurz die Mathematik erklärt, die diese Verkleinerung der Aufnahmen bewirkt.

Der aufgenommene und durch das Experiment gefilterte Ton besitzt die selbe Frequenzabfolge wie der Testton. Für eine Frequenz sollte der Ton einer Funktion

$$f(t) = a + b \sin(2\pi f t) \quad \text{Gl. 8-1}$$

folgen.  $a$  ist die mögliche Verschiebung der Nulllinie durch den Verstärker der Soundkarte und  $b$  die gesuchte Amplitude. Die Frequenz  $f$  ist durch den Testton festgelegt und ändert sich alle 0,2 s. Integriert man diese Funktion über ein Vielfaches der Periodenlänge  $T = \frac{1}{f}$ , so erhält man

$$\int_{nT} f(t) dt = anT. \quad \text{Gl. 8-2}$$

Integriert man das Quadrat der Funktion, so kommt man zur folgenden Gleichung

$$\int_{nT} [f(t)]^2 = a^2 nT + \frac{1}{2} b^2 nT. \quad \text{Gl. 8-3}$$

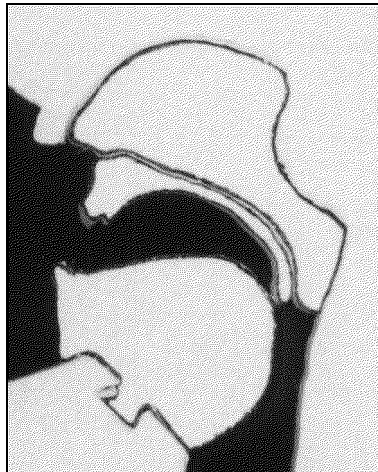
Ist die Frequenz also bekannt, kann man durch einfache numerische Integration die Amplitude und die mögliche Verschiebung des aufgenommenen Tons bestimmen. Für die Integration ist es sinnvoll nur etwa 0,1 s des 0,2 s langen Testtons zu betrachten, um Einschwingungseffekte an den Rändern des Frequenzintervalls nicht zu berücksichtigen. Das Programm faßt also jedes Frequenzintervall von 0,2 s in zwei Werten zusammen.

Die durch das Konvertierungsprogramm erzielten Werte werden wieder in Excel zu Diagrammen zusammengefaßt.

### 8.3 Meßergebnisse

Als erstes mußte die Apparatfunktion des Aufbaus gemessen werden, damit ein Fehler, der durch den Lautsprecher beziehungsweise durch das Abspielgerät verursacht werden könnte, ausgeschlossen werden konnte. Danach wurden die Mündern der verschiedenen Konfigurationen eingesetzt und der gefilterte Ton aufgenommen.

Da besonders zwei Mundstellungen unterschiedliche Ergebnisse lieferten, wird im weiteren auf die Mundstellungen der [a] und [i] Konfiguration eingegangen. In Abbildung 8-5 ist eine Fotografie der Mundstellung bei der Artikulation des Vokals [a]. Die Besonderheit der Zungenstellung des Vokals [a] sieht man in dieser Abbildung deutlich. Die Zunge ist sehr flach und in den Rachenraum gelagert, der dadurch sehr stark verengt wird.



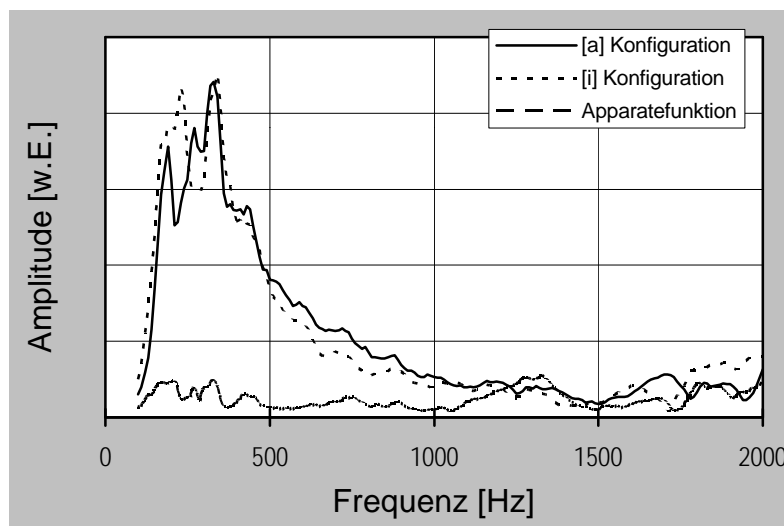
**Abbildung 8-5** Mundstellung der [a] Konfiguration



**Abbildung 8-6** Mundstellung der [i] Konfiguration

Die Stellung des Mundes der [i] Konfiguration wird in Abbildung 8-6 gezeigt. In dieser Abbildung sieht man den Unterschied zur [a] Konfiguration. Der Rachenraum ist nicht verengt und die Zunge liegt sehr nah am Gaumen.

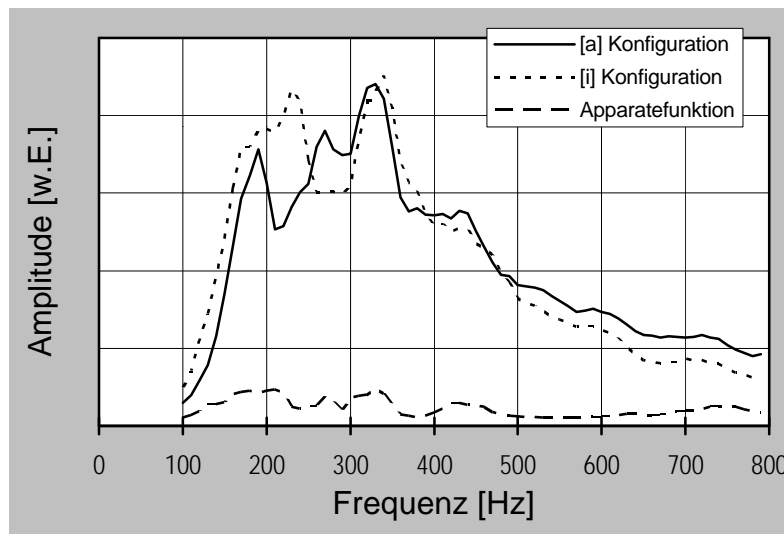
In Abbildung 8-7 sieht man die Frequenzverläufe bei den Mundstellungen. Die einzelnen Diagramme sind so normiert, daß sie zwischen 1300 Hz und 1400 Hz dasselbe Integral haben.



**Abbildung 8-7** Frequenzspektren der Vokalmundstellung [a] und [i]

Die Apparatfunktion hat eine sehr kleine Amplitude und verläuft im unteren Teil des Frequenzspektrums. Die Spektren der beiden Konfigurationen weisen einen unterschiedlichen Verlauf auf. Zwischen 100 Hz und 500 Hz haben beide Konfigurationen ihre Maxima und weichen sichtbar von der Apparatfunktion

ab. Um die Unterschiede der Konfigurationen in diesem Bereich zu sehen, ist in Abbildung 8-8 der Bereich bis 800 Hz vergrößert worden.



**Abbildung 8-8** Ausschnitt aus dem Frequenzspektrum der Konfigurationen

Bei der [a] Konfiguration kann man drei Maxima erkennen. Sie liegen bei 190 Hz, 270 Hz und 320 Hz. Danach fällt die Funktion ab. Die [i] Konfiguration hat nur zwei ausgeprägte Extrema, die bei 220 Hz und 330 Hz liegen.

Die beiden Frequenzspektren zeigen also bei verschiedenen Mundkonfigurationen unterschiedliche Verläufe und unterscheiden sich stark von der Apparatefunktion. Folglich sind die Verläufe der [a] und [i] Konfiguration durch den Aufbau zu erklären. Sie müssen durch Resonanzen im Ansatzrohr entstanden sein, deren Frequenzen durch die Maxima angezeigt werden. Die Lautstärkeschwankungen, die man im Frequenzspektrum sieht, kann man auch hören.

Durch die Ergebnisse können nur qualitative Aussagen über die Ausbildung der Resonanzfrequenzen gemacht werden. Man kann jedoch eindeutig feststellen, daß durch verschiedene Mundstellungen unterschiedliche Resonanzfrequenzen ausgebildet werden. Leider wurden keine Ergebnisse erzielt, die man mit denen aus Kapitel 6 vergleichen könnte. Trotzdem kann man anhand des Experiments die Vorgänge im menschlichen Stimmapparat verstehen. Möchte man die verschiedenartigen Mundstellungen illustrieren, so ist der Aufbau ein außergewöhnliches Anschauungsmaterial.

Um quantitative Aussagen machen zu können, müßte der Aufbau neu durchdacht werden. Der Stimmapparat sollte erstens dreidimensional (mit rundem Querschnitt) nachgebaut werden, damit sich die Schallwellen so wie im realen Mundraum ausbreiten können. Außerdem sollte man bessere Materialien verwenden. Denkbar wäre ein Aufbau mit einem knickbaren Rohr aus Schaumstoff, wie es zur Wärmeisolierung von Wasserleitungen verwendet wird. Man müßte das Rohr dann schrittweise der Form des Stimmapparats anpassen. Die Idee mit dem Lautsprecher kann übernommen werden, allerdings würde ich einen stärkeren mit besserer Dynamik empfehlen. Da die Nachbildung des menschlichen Stimmapparats nicht das Thema dieser Arbeit war, wäre es sicherlich eine interessante Aufgabenstellung für eine künftige Examensarbeit. Anregungen für den Nachbau finden sich unter anderem bei Flanagan [5].

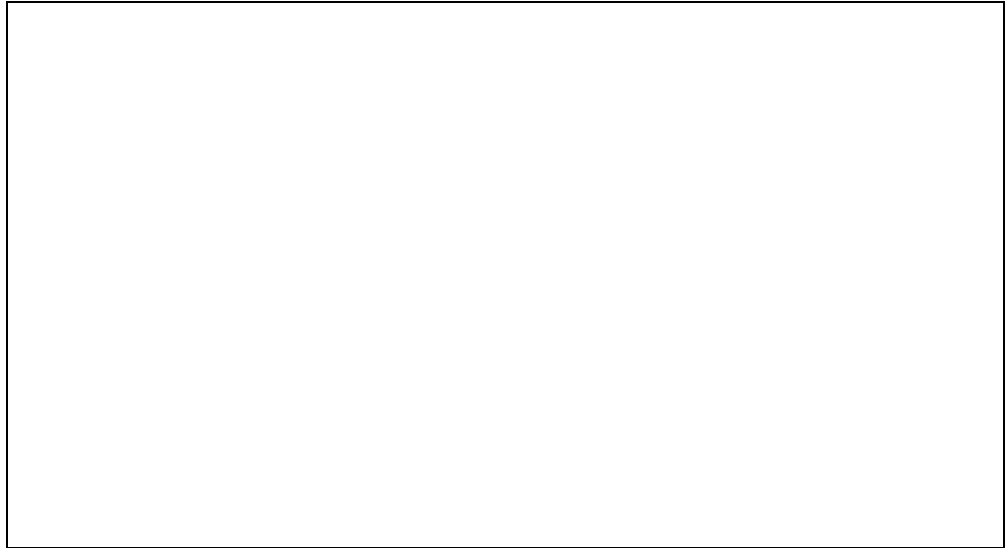
## 9 Alternative Meßmethode

Im Laufe meiner Arbeit wurde ich auf ein alternatives Meßsystem aufmerksam. Es handelt sich um das Transientenrekorder-Programm der Firma Leybold. Es kann nur in Verbindung mit einem Cassy-System betrieben werden. Das Cassy-System ist ein modulares Meßgerät für Schulen, das auch als Analog-Digital-Wandler genutzt werden kann.

Der Transientenrekorder ist in der Lage Klänge aufzunehmen, auszuwerten und graphisch darzustellen. Die Analyse des Signals erfolgt mit der Fast-Fourier-Transformation. Die FFT kann bis zu 2048 Stützpunkte verarbeiten. Das Programm weist einige Möglichkeiten, wie zum Beispiel drucken und Ausschnitte vergrößern, auf.

Zuerst wurde eine Prüfmessung mit dem System durchgeführt. Dafür wurde wieder die Stimmgabel mit 440 Hz benutzt. Das Ergebnis der Aufnahme wird in Abbildung 9-1 dargestellt.





**Abbildung 9-1** Abtastfunktion und Frequenzspektrum der Stimmgabel (440 Hz)

Oben im Bild ist die Abtastfunktion zu sehen und unten das dazugehörige Frequenzspektrum. Neben dem Maximum bei  $450 \text{ Hz} \pm 10 \text{ Hz}$  erscheint noch eine kleine Erhebung bei ungefähr 880 Hz.

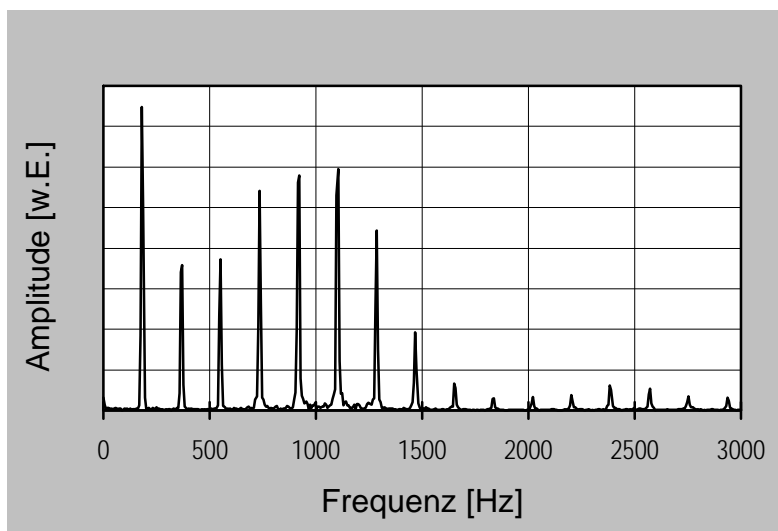
Anschließend wurden die verschiedenen Schallarten mit dem Transientenrekorder aufgenommen. Als Beispiel für die graphische Darstellung der Ergebnisse dient ein Vokal. In Abbildung 9-2 wird die Abtastfunktion und das daraus resultierende Frequenzspektrum des Kardinalvokals [a] der weiblichen Testperson dargestellt.



**Abbildung 9-2** Abtastfunktion und Frequenzspektrum des Kardinalvokals [a]

Die Grundfrequenz des gesprochenen Vokals [a] beträgt 175 Hz. Aufgrund der wenigen Stützpunkte, bedingt durch die hohe Grundfrequenz von Frauen, kann nur ein Formant bei etwa 900 Hz ermittelt werden.

Zum Vergleich mit dem Frequenzspektrum aus dem vorgestellten Analyseverfahren, ist in Abbildung 9-3 noch mal ein Spektrum der selben Testperson und des gleichen Vokals dargestellt.



**Abbildung 9-3** Frequenzspektrum ermittelt mit sFFT Programm

Vergleicht man die Grundfrequenzen der beiden Spektren (Abbildung 9-2 und Abbildung 9-3), so unterscheiden sie sich nur wenig. Bei den in dieser Arbeit vorgestellten Analyseverfahren liegt die Grundfrequenz bei 180 Hz. Dieser Unterschied kommt daher, daß die Aufnahmen an verschiedenen Tagen gemacht wurden und so der Vokal einmal höher und einmal etwas tiefer gesprochen wurde. Die Füße der scharfen Maxima beim Frequenzspektrum des Transientenrekorders sind durch den Leckeffekt sehr breit. In Abbildung 9-3 hingegen sind die Füße der Maxima schmäler. Nur wenn der Leckeffekt klein gehalten wird, ist die Höhe der Maxima exakt bestimmt. Überdies sieht man den dritten Formant im Frequenzspektrum des Transientenrekorders überhaupt nicht. Diese Tendenz zeigt sich auch bei den anderen Kardinalvokalen. Der dritte Formant ist beim Cassy-System nie zu sehen. Durch mein Analyseverfahren ist es möglich selbst den dritten Formanten noch zu bestimmen, was in Abbildung 9-3 zu sehen ist. Prinzipiell zeigen die Spektren das selbe Bild.

Der Transientenrekorder hat jedoch einige entscheidende Nachteile:

- Der Transientenrekorder arbeitet im Prinzip wie ein Speicheroszilloskop. Man kann also eine Messung aufnehmen, die im Speicher des Cassy-System abgelegt wird. Damit ist die Gesamtzahl der Stützpunkte auf die Größe des Speichers des Cassy-System begrenzt. Nach der Messung liest der Computer die gesamte Abtastfunktion aus dem Cassy-System aus. Durch dieses Meßprinzip kann man entweder die Zeit der Abtastfunktion verlängern, wobei man die Abtastrate verringern muß, oder man kann mit hoher Abtastrate einen kurzen Klang aufnehmen. In beiden Fällen ist es schwierig, Aussagen über hohe Frequenzen zu machen. Bei einer Aufnahme mit der Soundkarte besteht diese Einschränkung nicht. Es ist möglich, beliebig lange Aufnahmen mit beliebig hoher Abtastfrequenz zu analysieren.
- Die Aufnahmen können nur mit einer Auflösung von 8 Bit erfolgen. Das bedeutet, daß die gemessene Amplitude in ganzzahligen Werten von -127 bis 128 dargestellt werden kann. Die einzelnen Meßpunkte der Abtastfunktion weisen also einen relativ hohen systematischen Fehler von mindestens 0,8 % auf. Die Soundkarte hat eine Bandbreite von 16 Bit. Die Abtastfunktion wird also in Werten von -32767 bis 32768 dargestellt, was einem systematischen Fehler von 0,003 % entspricht.
- Da der Transientenrekorder nicht in der Lage ist die aufgenommenen Töne wieder abzuspielen, ist eine akustische Kontrolle der Aufnahmen nicht möglich. Es kann nicht festgestellt werden, ob die Aufnahme verrauscht oder durch Hintergrundgeräusche gestört wurde.
- Das Analyseprogramm des Transientenrekorders ist nicht in der Lage Ausschnitte aus der Abtastfunktion herauszuschneiden und separat zu analysieren.
- Die Frequenzspektren aus dem Transientenrekorder weisen einen großen Leckeffekt auf, weil nur mit FFT gearbeitet werden kann und kein Hamming-Fenster über die Abtastfunktion gelagert werden kann.
- Der Transientenrekorder kann nur Fourier-Transformieren. Weitere Auswertungsmethoden, wie die Lorenzabbildung und die Autokorrelationsfunktion, sind nicht möglich.

Das in dieser Arbeit verwendete System liefert bei der Analyse von Klängen qualitativ bessere Ergebnisse. Allerdings ist es auf die Bearbeitung von akustischen Signalen spezialisiert. Der große Vorteil dieses Analyseverfahrens

liegt darin, daß Schüler nicht nur sehen, was sie aufgenommen haben, sondern auch ihre Aufnahmen abspielen und akustisch wahrnehmen können. Überdies ist die Soundkarte ein interessantes Medium für den Physikunterricht. Mit ihr kann man nicht nur Experimente zur Akustik durchführen, sondern es wären auch Experimente zur Mechanik denkbar. Ein Beispiel wäre der Dopplereffekt: Auf einen Lautsprecher legt man ein Ton von 1000 Hz und schleudert ihn mit ausgestreckter Hand schnell im Kreis. Der erzeugte Ton wird dann mit einem Mikrofon und der Soundkarte aufgenommen und ausgewertet.

Zusammenfassend kann man sagen, daß der Transientenrekorder sicherlich eine alternative Meßmethode zu dem in dieser Arbeit vorgestellten Meßablauf darstellt. Möchte man jedoch quantitative Aussagen über die Physik des Stimmapparats machen, dann sollte das vorgestellte und ausgearbeitete Meßverfahren benutzt werden. Ferner ist dieses Verfahren aus didaktischer Sicht effektiver, weil Schüler nicht nur ihren Sehsinn, sondern auch ihr Gehör benutzen können. Hinzu kommt noch, daß alle benutzten Analyseprogramme kostenlos sind und Schüler so diese Versuche auch zu Hause ausprobieren könnten, denn heutzutage besitzt fast jeder Schüler einen Computer mit einer eingebauten Soundkarte.

## 10 Zusammenfassung und Ausblick

In dieser Arbeit wurde der menschliche Stimmapparat mit Hilfe verschiedener Experimente unter physikalischen Aspekten untersucht. Es wurde ein System für den Computer geschaffen, um die vorgestellten Experimente durchführen zu können. Mit diesem System können quantitative Aussagen zur Stimme getroffen werden.

Die Experimente umfaßten verschiedenartige Intentionen. In Kapitel 6 wurden die Grundschaallformen beschrieben, die der Mensch in der Lage ist zu artikulieren. Diese Kategorisierung der Schallformen wurde vorgenommen, um die bestehenden Unterschiede aufzuzeigen. Wichtig war dabei, daß der Entstehungsmechanismus der einzelnen Schallformen erläutert wurde. In der

Abtastfunktion des Explosionsschall [t], [k] läßt sich eine Zeitkonstante für das Öffnen eines Verschlusses im Rachenraum feststellen. Das Friktionsrauschen [f], [s] weist charakteristische Differenzen im Frequenzspektrum und in der 3D-Lorenzabbildung auf. Der periodische Schall findet sich in den Kardinalvokalen wieder. Es wurden die Kardinalvokale [a], [e], [i], [o], [u] untersucht. Aus den ermittelten Ergebnissen kann ein Zusammenhang zwischen der Zungenstellung und der Frequenz des ersten Formanten gezogen werden. Denn je höher die Zungenstellung bei der Artikulation eines Vokals im Mundraum ist, desto niedriger ist die Frequenz des ersten Formanten. Der zweite Formant ist mit der Lage der horizontalen Zungenspitze korreliert. Außerdem wurde das systematische Verhalten der Formanten der einzelnen Kardinalvokale in einem Diagramm zusammengefaßt.

Die Lorenzabbildung wurde zur Identifizierung einer Stimme verwendet. Wenn man Stimm aufnehmen eines bestimmten Sprechers hat, kann man anhand der Lorenzabbildung den Sprecherkreis eingrenzen. Es stellte sich während dieser Arbeit heraus, daß die Lorenzabbildung nur in Bezug auf die Kardinalvokale eingesetzt werden sollte, weil bei diesen Buchstaben die Individualität eines jeden Sprechers sichtbar wird. Die Lorenzabbildungen der verschiedenen Probanden zeigen einige Eigenarten, die in Kapitel 6 verdeutlicht wurden.

Im Kapitel „Untersuchungen am Stimmapparat“ (Kapitel 7) wurde durch gezielte Änderungen eines Parameters am Stimmapparat versucht, die Auswirkungen auf den Ton festzustellen. Das erste Experiment dieses Kapitels untersuchte die Auswirkung auf den Kardinalvokal [a] beim Öffnen und Schließen des Mundes. Dabei stellte sich heraus, daß sich die Grundfrequenz beim Öffnen des Mundes nicht ändert, aber die ersten beiden Formanten sich zu höheren Frequenz hin verschieben. Wie es zu dieser Formantenwanderung kommt, wurde durch ein Modell erklärt.

Das nächste Experiment beschäftigte sich mit der Variation der Tonhöhe und der daraus entstehenden Auswirkung auf das Frequenzspektrum. Es zeigte, daß sich die Formanten nur unwesentlich ändern, hingegen die Grundfrequenz ansteigt. Der Unterschied zwischen der Sing- und Sprechstimme wurde in einem weiteren Experiment untersucht. Einige Eigenheiten der Singstimme wurden anhand der Stimme einer Opernsängerin dargestellt. Die Kopf- und Bruststimme und das Vibrato sind einige Effekte, die eine Singstimme charakterisieren. Das letzte Experiment dieses Kapitels befaßte sich mit der Auswirkung von Helium auf die menschliche Stimme. Hier sieht man, daß sich

die Formanten unter Heliumeinfluß zu höheren Frequenzen hin verschieben, weil die Schallgeschwindigkeit in Helium größer ist als in Luft.

Anschließend wurde versucht durch ein Experiment den Stimmapparat nachzubauen, um an ihm reproduzierbare Experimente durchzuführen. Es ergaben sich zwar keine vergleichbaren Ergebnisse zum Menschen, jedoch kann man sehen, daß sich bei verschiedenen Mund- und Zungenstellungen unterschiedliche Resonanzfrequenzen ausbilden. Zum Schluß wurde eine alternative Meßmethode zu meinem Verfahren vorgestellt. Die Transientenrekorder-Software für das Cassy-System ist für die Schule sicherlich ein funktionsfähiges System, doch das hier vorgestellte System liefert exaktere Werte. Dadurch kann man detailliertere Aussagen treffen. Ferner ist es von der didaktischen Seite her sinnvoller das hier verwendete System mit der Soundkarte zu benutzen, weil die Schüler die Messungen nicht nur sehen, sondern auch akustisch überprüfen können. Durch die Möglichkeit von Aufnahme und Wiedergabe von unterschiedlichen Sprachsequenzen, kann die Aufmerksamkeit bei Schülern geweckt werden.

Diese Arbeit stellt einen Ausschnitt der möglichen Untersuchungen am menschlichen Stimmapparat dar. Mit dem zusammengestellten Meßsystem könnten nun weitere Examensarbeiten entstehen, die die Stimme unter anderen Gesichtspunkten untersuchen. Es wären viele Experimente denkbar, wie zum Beispiel die Untersuchung der Stimme von gesunden und kranken Testpersonen oder der Vergleich der Stimmen von kleinen und großen Probanden. Ferner wäre noch interessant, den Unterschied zwischen Kindern und Erwachsenen zu analysieren. Ein weiteres interessantes Thema für eine Examensarbeit wäre der Nachbau des menschlichen Stimmapparats. Die mit dieser „Maschine“ erzielten Werte könnten dann mit den Ergebnissen dieser Arbeit verglichen werden.

## 11 Anhang

### 11.1 Technische Hilfsmittel

#### Verwendete Computer:

- Intel 486DX-33, 8 MB Ram, Soundkarte Soundblaster 16 ASP.  
(Aufnahme der Sounddateien, Eigentum der Universität-GHS Essen)
- AMD 486DX4-120Plus, 20 MB Ram, Soundkarte Typhoon 16 Pro.  
(Analyse)
- Intel 486DX2-66, 8 MB Ram, Soundkarte Soundblaster 16 Pro.  
(Analyse)

#### Audio-CD:

- Stück: „La mamma Morta“ von Umberto Giordano; © EMI-Records

#### Cassy-System:

- Transientenrekorder, Leybold (524761), 1992.

### 11.2 Verwendete Programme

Für die vorliegende Arbeit habe ich folgende Programme verwendet:

Programm	Autor
sFFT v2.1	David C. Copley, © Colpleyright Software 1995
Spectrogram v2.3	R.S. Horne, ©1994
Excel v5.0 a	Microsoft, ©1985-1993 Microsoft Corporation
Visual Basic v4.0	Microsoft, ©1987-1995 Microsoft Corporation
Word v6.0 a	Microsoft, ©1989-1994 Microsoft Corporation

### 11.3 Nutzung der Disketten

Auf den beiden Disketten befinden sich lauffähige Versionen der verwendeten Analyseprogramme. Diskette 1 beinhaltet die Programme aus dem Internet (sFFT und Spectrogram) und einige Klänge im WAV-Format, die für meine Arbeit aufgenommen wurden. Die zweite Diskette enthält das Programm WAVcalc und weitere Klänge. Die Klänge werden weiter unten beschrieben.

Man kann die Programme von der Diskette starten, indem man im Programm-Manager von Windows im Menü „Datei“ den Befehl „Ausführen“ wählt. In dieser Form drückt man den Befehl „Durchsuchen“ und gelangt in eine weitere Form, in der man das jeweilige Programm von der Diskette starten kann.

#### Diskette 1:

Dateiname	Beschreibung
stimmga.wav	Aufnahme der Stimmgabel mit 440 Hz (3,2 s)
stimmga1.wav	Ausschnitt aus Stimmgabelaufnahme (30 Perioden)
uoaei.wav	Aufnahme der Vokale [a], [e], [i], [o], [u] für Sonographie-Programm (5,5 s)

#### Diskette 2:

Dateiname	Beschreibung
a_he_m.wav	Vokal [a] der männlichen Testperson 1 unter Heliumeinfluß (30 Perioden)
a_männl.wav	Vokal [a] der männlichen Testperson 1 (30 Perioden)
auf_zu_1.wav bis auf_zu_7.wav	Öffnen und Schließen des Mundes von der männlichen Testperson 2 gesprochen. (je 30 Perioden)
f_männl.wav	Friktionsrauschen [f] eines männlichen Probanden (220 ms)
k_männl.wav	Explosionsschall [k] eines männlichen Probanden (110 ms)
s_weibl.wav	Friktionsrauschen [s] der weiblichen Testperson (220 ms)
t_weibl.wav	Explosionsschall [t] der weiblichen Testperson (160 ms)
tonh_1_m.wav bis tonh_4_m.wav	Variation der Tonhöhen von der männlichen Testperson 3 gesprochen. (30 Perioden)
u_he_m.wav	Vokal [u] der männlichen Testperson 1 unter Heliumeinfluß (30 Perioden)



---

u_männl.wav	Vokal [u] der männlichen Testperson 1 (30 Perioden)
-------------	---

## 11.4 Abbildungsverzeichnis

Abb.	Titel	Seite
1-1	Teil einer Sprechmaschine von J. W. von Kempelen	2
2-1	Skizze eines Faden- und Federpendels	8
2-2	Illustration der Größen, die eine Sinusschwingung beschreiben	8
2-3	Überlagerung von Schwingungen verschiedener Amplituden	9
2-4	Überlagerung von Schwingungen mit Phasendifferenz	10
2-5	Überlagerung von Schwingungen mit doppeltem	11
2-6	Schwebung	11
2-7	Bewegung einer Welle (Wasserwelle)	12
2-8	Stehende Welle mit festen Enden	14
2-9	Stehende Welle mit einem festen und einem losen Ende	14
2-10	Frontalschnitt des Kehlkopfes von hinten	15
2-11	Schematische Aufsicht auf die Stimmlippen	16
2-12	Darstellung des aerodynamischen Paradoxons	17
2-13	Stimmbandschwingungen während des Luftdurchflusses	17
2-14	Frequenzspektrum der Stimmbandschwingungen	18
2-15	Schematische Schnittzeichnung des Ansatzrohrs	19
2-16	Filterfunktion des Ansatzrohrs	20
2-17	Analogie des Ansatzrohrs mit einem einseitig geschlossenen Rohr. Modell zur Erzeugung der Formanten.	21
2-18	Idealisiertes Frequenzspektrum des gesprochenen Tons	22
3-1	Sinusförmiges Signal und dessen Lorenzabbildung	25
3-2	1/f Rauschsignal und dessen Lorenzabbildung	26
3-3	Autokorrelationsfunktion einer Sinusschwingung	27
3-4	Autokorrelationsfunktion eines 1/f Rauschens	28
3-5	Darstellung der Funktion für die geklippte Autokorrelationsfunktion	30
3-6	Rechteckschwingung, deren Zerlegung in eine Summe von harmonischen Schwingungen und das dazugehörige...	34
3-7	Vergleich der Multiplikationen für die DFT und die FFT	35
3-8	Sinusfunktion überlagert mit Hamming-Fenster	36
3-9	DFT einer Sinusfunktion ohne Hamming-Fenster überlagert	37
3-10	DFT einer Sinusfunktion mit Hamming-Fenster überlagert	37
3-11	DFT einer Sinusfunktion mit ganzzahliger Periodenlänge und Hamming-Fenster	38

<b>Abb.</b>	<b>Titel</b>	<b>Seite</b>
4-1	Graphische Darstellung der ersten Form (WAVcalc v1.0)	40
4-2	Lorenzabbildung der Sinusfunktion aus Abbildung 4-1	41
4-3	Autokorrelationsfunktion der Sinusfunktion aus Abbildung 4-1	43
4-4	Darstellung des Aufnahmemodus (sFFT)	48
4-5	sFFT-Form mit geladener Sinusfunktion	49
4-6	Fourierspektrum der in Abbildung 4-5 gezeigten Sinusfunktion	50
4-7	Form für die Beschriftung der Bilder	50
4-8	Dialog-Box zum Analysieren einer Audiodatei in Spectrogram	52
4-9	Breitband-Sonagramm des Ausdrucks „Phonetik“	54
5-1	Flußdiagramm der Meßabläufe	58
5-2	Abtastfunktion der Stimmgabel	59
5-3	Lorenzabbildung der Abtastfunktion (Stimmgabel)	60
5-4	Autokorrelationsfunktion der Abtastfunktion (Stimmgabel)	60
5-5	Ausschnitt des Frequenzspektrums (Stimmgabel)	61
6-1	Abtastfunktion des Explosionsschalls [t] (weibl.)	64
6-2	Lorenzabbildung der Abtastfunktion [t] (weibl.)	65
6-3	Autokorrelationsfunktion des Explosionsschalls [t] (weibl.)	65
6-4	Frequenzspektrum des Explosionsschalls [t] (weibl.)	66
6-5	Abtastfunktion des Explosionsschalls [k] (weibl.)	66
6-6	Lorenzabbildung der Abtastfunktion aus Abbildung 6-5	67
6-7	Frequenzspektrum des Explosionsschalls [k] (weibl.)	68
6-8	Abtastfunktion des Explosionsschalls [k]	68
6-9	Lorenzabbildung der Abtastfunktion aus Abbildung 6-7 (männl.)	69
6-10	Frequenzspektrum des Explosionsschalls [k] (männl.)	69
6-11	Abtastfunktion des Friktionsrauschens [f] (weibl.)	71
6-12	Lorenzabbildung des Friktionsrauschens [f] (weibl.)	72
6-13	Autokorrelationsfunktion des Friktionsrauschens [f] (weibl.)	72
6-14	Frequenzspektrum des Friktionsrauschens [f] (weibl.)	73
6-15	Abtastfunktion des Friktionsrauschens [s] (männl.)	73
6-16	Lorenzabbildung des Friktionsrauschens aus Abbildung 6-13	74
6-17	Autokorrelationsfunktion der Abtastfunktion aus Abbildung 6-13 (männl.)	74
6-18	Frequenzspektrum des Friktionsrauschens [s] (männl.)	75
6-19	3D-Lorenzabbildung des Friktionsrauschens [f]	76
6-20	3D-Lorenzabbildung des Friktionsrauschens [s]	76
6-21	Mundstellung bei der Artikulation des Vokals [a]	77
6-22	Abtastfunktion des Vokals [a] (männl. 1)	78

<b>Abb.</b>	<b>Titel</b>	<b>Seite</b>
6-23	Lorenzabbildung des Vokals [a] (männl. 1)	79
6-24	Autokorrelationsfunktion der Abtastfunktion [a] (männl. 1)	79
6-25	Frequenzspektrum des Vokals [a] (männl. 1)	80
6-26	Ergebnisse der Analyseverfahren des Vokals [a] (weibl.)	81
6-27	Ergebnisse der Analyseverfahren des Vokals [a] (männl. 2)	82
6-28	Ergebnisse der Analyseverfahren des Vokals [a] (männl. 3)	83
6-29	Mundstellung für den Kardinalvokal [e]	85
6-30	Abtastfunktion und Frequenzspektrum des Vokals [e] (männl.2)	85
6-31	Lorenzabbildungen des Vokals [e]	87
6-32	Mund- und Zungenstellung bei der Artikulation des Vokals [i]	88
6-33	Abtastfunktion und Frequenzspektrum des Vokals [i] (männl.3)	88
6-34	Lorenzabbildung des Vokals [i]	90
6-35	Mundstellung für die Artikulation des Kardinalvokals [o]	91
6-36	Abtastfunktion und Frequenzspektrum des Vokals [o] (weibl.)	91
6-37	Lorenzabbildungen des Vokals [o]	93
6-38	Mundstellung für die Artikulation des Kardinalvokals [u]	94
6-39	Abtastfunktion und Frequenzspektrum des Vokals [u] (männl.3)	94
6-40	Lorenzabbildungen des Vokals [u]	96
6-41	Breitband-Sonagramm der Vokale [u], [o], [a], [e], [i]	97
6-42	Schema der Formantenbewegung der Vokale	98
7-1	Frequenzspektren des Vokals [a] beim Öffnen des Mundes (männl.	100
7-2	Formantenfrequenz in Abhängigkeit vom Öffnungsgrad des Mundes	101
7-3	Druckverteilung in einem einseitig geschlossenen Rohr	102
7-4	Auswirkung der Druckverteilung am offenen Ende eines einseitig geschlossenen Rohrs auf die Wellenlänge	103
7-5	Lage des dritten Formanten (männl. 3)	104
7-6	Variation der Tonhöhe des Vokals [a] (männl. 2)	105
7-7	Grundfrequenzänderung mit steigender Tonhöhe (männl. 2)	106
7-8	Verhältnisse der theoretischen und gemessenen Wert der Töne...	108
7-9	Abtastfunktion eines gesungenen Vokals [a]	109
7-10	Darstellungsarten des gesungenen Vokals [a]	110
7-11	Frequenzspektrum des gesungenen Vokals [a]	111
7-12	Obertonreiches Frequenzspektrum einer Bruststimme [a]	112
7-13	Breitbandsonagramm eines Ausschnitts aus „La mamma morta“	113
7-14	Frequenzspektren des Vokals [a] mit Luft und Helium	114
7-15	Frequenzspektren des Vokals [u] mit Luft und Helium	114
7-16	Verhältnis der Formanten von Helium- zu Lufteinfluß	115

---

<b>Abb.</b>	<b>Titel</b>	<b>Seite</b>
8-1	Schematischer Aufbau der Styroporsäge	118
8-2	Schematische Zeichnung des Aufbaus (Experiment)	119
8-3	Foto des nachgebauten Mundraums von oben	120
8-4	Flußdiagramm des Meßablaufs (Experiment)	121
8-5	Mundstellung der [a] Konfiguration	123
8-6	Mundstellung der [i] Konfiguration	124
8-7	Frequenzspektren der Vokalmundstellung [a] und [i]	124
8-8	Ausschnitt aus dem Frequenzspektrum der Konfigurationen	125
9-1	Abtastfunktion und Frequenzspektrum der Stimmgabel (440 Hz)	127
9-2	Abtastfunktion und Frequenzspektrum des Kardinalvokals [a]	128
9-3	Frequenzspektrum ermittelt mit sFFT Programm	129

## 11.5 Literaturverzeichnis

- [1] L. Mathelitsch, G. Friedrich, *Die Stimme Instrument für Sprache, Gesang und Gefühl*, Springer Verlag, Berlin Heidelberg, 1995.
- [2] B. Pompino-Marschall, *Einführung in die Phonetik*, de Gruyter, Berlin New York, 1995.
- [3] G. Fant, *Acoustic theory of speech production*, Mouton & Co., Netherlands, 1960.
- [4] G. Fant, *Speech sounds and features*, The Colonial Press Inc., USA, 1973
- [5] J.L. Flanagan, *Speech analysis Synthesis and Perception*, Springer-Verlag, Berlin Heidelberg New York, 1972.
- [6] H. Kuchling, *Physik*, VEB Fachbuchverlag, Leipzig, 1987.
- [7] M. Alonso, E.J. Finn, *Physik*, Addison-Wesley, Bonn, München, 1988.
- [8] G. Habermann, *Stimme und Sprache: Eine Einführung in ihre Physiologie und Hygiene*, Thieme, Stuttgart, 1978.
- [9] J. Neppert, M. Pétursson, *Elemente einer akustischen Phonetik*, Buske-Verlag, Hamburg, 1986.
- [10] C. Gerthsen, H.O. Kneser, H. Vogel, *Physik*, Springer-Verlag, Berlin Heidelberg, 1992.
- [11] L. Mathelitsch, *PhuD* **4**, 299-316 (1987)
- [12] P.Lieberman, S.E. Blumenstein, *Speech physiologie, speech perception, and acoustic phonetics*, Cambridge University Press, Cambridge, 1988.
- [13] R.F. Schmidt, G. Thewes, *Physiologie des Menschen*, Springer-Verlag, Berlin Heidelberg New York, 1986.
- [14] H.-H. Wängler, *Physiologische Phonetik*, N.G.Elwert Verlag, Marburg, 1972.
- [15] J. Sundberg, *Sientific American* **3**, 82-91 (1977).
- [16] R.D. Kent, *Journal of Voice* **2**, 97-117 (1993).
- [17] R.T. Sataloff, *Spektrum der Wissenschaft* **11**, 74-81 (1993).

- [18] J. Sundberg, *Folia phoniata*. **22**, 28-48 (1970).
- [19] E.N. Lorenz, *Journal of atmospheric science* **20**, 130-141 (1963).
- [20] H. Ney, *Untersuchungen zur automatischen Erkennung kooperativer Sprecher über Telefon*, Buske-Verlag, Hamburg, 1983.
- [21] K. Sickert, *Automatische Spracheingabe und Sprachausgabe: Analyse, Synthese und Erkennung menschl. Sprache mit digitalen Systemen*, Verlag Markt und Technik, Haar bei München, 1983.
- [22] E.O. Brigham, *FFT: schnelle Fourier-Transformation*, R. Oldenbourg Verlag GmbH, München, 1985.
- [23] P. Monadjemi, *Visual Basic 4: Das Kompendium*, Markt & Technik Verlag, Haar bei München, 1996.
- [24] P. Wollschlaeger, *Visual Basic*, Markt & Technik Verlag, Haar bei München, 1992.
- [25] H. Bußmann, *Lexikon der Sprachwissenschaft*, Kröner Verlag, Stuttgart, 1990.

## 11.6 Danksagung

Die vorliegende Arbeit wurde an der Fakultät der Physik der Universität-GHS Essen unter der Leitung von Herrn Prof. Dr. H. J. Schlichting erstellt. Ihm danke ich für die Themenstellung und die freundliche Unterstützung, sowie für viele Anregungen und die hilfreichen Diskussionen.

Außerdem danke ich Herrn Volkhard Nordmeier und Herrn Dr. Korpiun, die durch interessante Vorschläge meine Arbeit unterstützt haben.

Dem Fachbereich Physik, AG Didaktik, möchte ich für die Bereitstellung der verwendeten Geräte und des Arbeitsplatzes danken.

Ebenfalls möchte ich meinen Testpersonen einen Dank aussprechen, denn durch ihr Interesse und ihre Geduld konnten die Aufnahmen entstehen, die in dieser Arbeit dargestellt sind.

Zu guter Letzt möchte ich meinem Verlobten Armin und meiner Familie danken, die mir in der Zeit meiner Examensarbeit stets zur Seite standen.

## 11.7 Eidesstattliche Erklärung

Ich versichere, daß ich die schriftliche Hausarbeit einschließlich evtl. beigefügter Zeichnungen, Kartenskizzen, Darstellungen u.ä.m. selbständig angefertigt und keine anderen als die angegebenen Hilfsmittel benutzt habe.

Alle Stellen, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, habe ich in jedem einzelnen Fall unter genauer Angabe der Quelle deutlich als Entlehnung kenntlich gemacht.

.....



Andrea L. Moser