

# Akustik des Vokals Präliminarien

# subTexte

Die Reihe *subTexte* vereinigt Originaltexte zu jeweils einem Untersuchungsgegenstand aus den beiden Forschungsschwerpunkten «Performative Praxis» und «Filmwissen/Filmerfahrung». Sie bietet Raum für Texte, Bilder oder digitale Medien, die zu einer Forschungsfrage über, für oder mit Darstellender Kunst oder Film entstanden sind. Als Publikationsgefäß trägt die Reihe dazu bei, Forschungsprozesse über das ephemere Ereignis und die Einzeluntersuchung hinaus zu ermöglichen, Zwischenergebnisse festzuhalten und vergleichende Perspektiven zu öffnen. Vom Symposiumsband bis zur Materialsammlung verbindet sie die vielseitigen, reflexiven, ergänzenden, kommentierenden, divergierenden oder dokumentierenden Formen und Ansätze der Auseinandersetzung mit den Darstellenden Künsten und dem Film.

In der Reihe *subTexte* sind bisher erschienen:

**subTexte 01** Attention Artaud

**subTexte 02** Wirklich? — Strategien der Authentizität im aktuellen Dokumentarfilm

**subTexte 03** Künstlerische Forschung — Positionen und Perspektiven

**subTexte 04** research@film — Forschung zwischen Kunst und Wissenschaft

**subTexte 05** Theater Vermittlung Schule. Ein Dialog

**subTexte 06** Wirkungsmaschine Schauspieler —

Vom Menschendarsteller zum multifunktionalen Spielmacher

**subTexte 07** Ästhetische Kommunikation im Kindertheater —

Eine Studie zu Rezeptionsweisen und Erlebnisqualitäten

**subTexte 08** Akustik des Vokals — Präliminarien

**Z**

hdk

Zürcher Hochschule der Künste  
Institute for the Performing Arts and Film

# Akustik des Vokals

Präliminarien



# Einleitung

## **Thematik und Vorhaben des Buchs**

Die Stimmbänder erzeugen einen Klang, und die Hohlräume von Rachen, Mund und Nase formen diesen Klang durch ihre Resonanzen zu einem Laut. Je nach Stellung des Gaumens, der Zunge, der Lippen und des Kiefers sind die Hohlräume verschieden, und mit ihnen die ausgeübte Resonanzwirkung. Deshalb können verschiedene stimmhafte Laute gebildet und entsprechend verschieden wahrgenommen werden. Wird ein Laut wahrnehmend einer sprachlichen Einheit zugeordnet und findet sich im Hohlraum von Rachen und Mund kein Verschluss, so wird der Laut «Vokallaut» und die sprachliche Identität «Vokal» genannt.

Mit dieser oder einer ähnlichen allgemeinen Formulierung beginnt die derzeitige akustische Theorie des Vokals. Wir produzieren mit den Stimmbändern einen allgemeinen Klang und formen diesen über die Resonanzen des Vokaltraktes zu einem Laut eines spezifischen Vokals. Wir phonieren und artikulieren.

Weil dem so ist, weisen die Laute selbst, als Klänge, relative spektrale Energiemaxima in denjenigen Frequenzbereichen auf, welche den Resonanzen des Vokaltrakts bei der Produktion entsprechen. Diese spektralen Energiemaxima werden, ihrer Funktion gemäss, «Formanten» genannt.

Aus dieser Sichtweise ergibt sich der bestehende psychophysische Grundsatz des Vokals: Laute, welche als derselbe Vokal wahrgenommen und erkannt werden, weisen untereinander ähnliche Formantmuster auf, ähnliche Muster relativer spektraler Energiemaxima, und Laute, welche als verschiedene Vokale wahrgenommen und erkannt werden, weisen untereinander unähnliche Formantmuster auf.

Eine solche Auffassung, wie Vokale erzeugt und durch welche Merkmale sie in der Folge physikalisch repräsentiert werden, erscheint zunächst plausibel, wenn nicht in direkter Weise einleuchtend. Wir lassen beim Sprechen tatsächlich die Stimmbänder schwingen, wir bewegen den Mund zur Bildung verschiedener Laute, ja wir vermögen

oft gesprochene Wörter direkt an diesen Bewegungen «abzulesen», eine Fähigkeit, die insbesondere Gehörlose weit zu treiben im Stande sind.

Statistische Untersuchungen scheinen ihrerseits die Entsprechung von Vokalen und vokalspezifischen Formantmustern zu bestätigen.

Vokalsynthesen, welche Geräusche oder Klänge über Filter transformieren, erweisen sich ebenfalls als sehr potent, erkennbare «Vokallaute» zu produzieren.

Bestehende Probleme in der Beschreibung der physikalischen Merkmale von Vokalen stehen in einer solchen Perspektive nicht in Zusammenhang mit dem theoretischen Grundsatz selbst, sondern in Zusammenhang mit der tatsächlichen Dynamik und Komplexität der Sprache, ihre Produktion wie ihre Perzeption betreffend. Isolierte Vokallaute, für welche eine einfache und statische Entsprechung von erkanntem Vokal und ihm spezifisch zugehörigen Formantmustern vorzufinden wäre, spielen — wiederum dieser Perspektive gemäss — in der Alltagssprache eine marginale Rolle. Vokale in der Rede sind eingebunden in Laut- und Bedeutungskontexte, und die entsprechenden Klänge weisen ausgeprägte dynamische Vorgänge und insbesondere Transitionen von einem Laut zu einem anderen auf. So kommt es, dass Vokallaute in einer Rede auch dann erkannt werden können, wenn ausgeprägte statische Klanganteile fehlen. So kommt es auch, dass ein Vokallaut, aus der Rede als Klangfragment isoliert, in bestimmten Fällen einen anderen Vokal erkennen lassen kann, als dies in der Rede der Fall ist. So kommt es zudem, diese beiden Erfahrungen verlängernd, dass die Rede als solche auch bei massiven Störungen oder Transformationen ihrer Übermittlung verständlich bleiben kann. Und Ähnliches mehr.

In der Folge stehen derzeit verschiedene Arten der Phonation, dynamische Abläufe und Lautvariationen in Zusammenhang mit dem jeweiligen akustischen, syntaktischen und semantischen Kontext sowie die Komplexität der Wahrnehmung und des Verstehens der Sprache ganz allgemein im Vordergrund der Aufmerksamkeit.

Die nachfolgende Erörterung aber kehrt, trotz alledem, zum oben zitierten Anfangssatz der Akustik des Vokals zurück. Sie stellt eine kritische Lektüre des bestehenden Grundsatzes vor, bis hin zu seiner Falsifikation. Sie versucht aufzuzeigen, dass der Grundsatz zwar auf die tatsächliche physikalische Charakteristik des Vokals verweist, diese aber nicht zutreffend bezeichnet. Sie versucht also, auf ein bis anhin nicht gelöstes Grundlagenproblem des stimmhaften Sprachlauts, und damit der Stimme und der Sprache selbst, hinzuweisen und es — wieder — zur Frage zu bringen.

Die Form der Ausführungen ist für einen wissenschaftlichen Rahmen zum Teil unüblich. Dennoch verzichten wir, die nachfolgenden vier Hinweise ausgenommen, auf ausführliche einleitende Erklärungen. Der Sinn der Form sollte aus den Darstellungen selbst heraus einsichtig werden. Zudem finden sich im Nachwort einige ergänzende Erläuterungen.

Vier Hinweise seien wie angedeutet den Ausführungen dennoch vorangestellt. Sie betreffen den sprachlichen Stil, die Referenzen, den Anspruch und die eingenommene Perspektive der vorliegenden Schrift.

Die Ausführungen sind «abstrakt» beziehungsweise «technisch» gehalten. Dies mag die Lektüre erschweren. Hinzu kommt, dass der Text — die Kapitel 1.10 und 2.1 ausgenommen — nicht von illustrierenden Grafiken begleitet wird. Die Darstellungen ab dem dritten Teil setzen ausserdem ein Durchdenken des genannten theoretischen Grundsatzes, eine Sicherheit in der Verwendung der mit ihm verbundenen Begriffe und eine Sicherheit in der Bezugnahme auf derzeitige statistische Werte für Formantmuster seitens der Lesenden voraus. Doch erachten wir einen derart verfahrenen Text als notwendig: Er insistiert auf der Erörterung weniger grundsätzlicher Überlegungen und allgemeiner Sachverhalte sowie ihrer gegenseitigen Zusammenhänge, im erwähnten Versuch, ein Grundlagenproblem deutlich werden zu lassen.

Der grösste Teil der im Text erörterten Einzelaspekte ist bereits in der Literatur dargestellt, wenn auch häufig nicht mit derselben Interpretation verbunden, wie wir sie vornehmen. Die entsprechenden Veröffentlichungen stammen dabei wiederum zum grössten Teil nicht von uns. Wenn wir auf diese wie auch auf unsere eigenen Veröffentli-

chungen im vorliegenden Text nicht einzeln verweisen — die Erörterung bestehender Formantstatistiken in Teil zwei ausgenommen —, so nur deshalb, weil wir einen allgemeinen Gang der Argumentation verfolgen und deutlich werden lassen wollen und deshalb keine ausführliche Diskussion von Einzelaspekten und auch keine entsprechenden Referenzierungen vornehmen. In den Materialien hingegen finden sich zahlreiche exemplarisch ausgewählte Zitate, verbunden mit Verweisen auf bestehende Veröffentlichungen.

Hinzu kommt, dass wir in unseren eigenen bisherigen Veröffentlichungen auf die meisten der hier zur Frage stehenden Aspekte schon eingegangen sind. Die vorliegende Schrift ist zwar — abgesehen von wenigen aus früheren Veröffentlichungen übernommenen einzelnen Formulierungen — in der Art ihrer Zusammenstellung und ihres Vortrags neu, stellt aber nur wenig neue inhaltliche Aspekte vor.

Was die eigene empirische Grundlage anbelangt, auf welche sich die Aussagen insbesondere des vierten Teils beziehen, so besteht diese aus Aufnahmen sehr verschiedener Bereiche des Alltags, der Unterhaltung und der Kunst. Ein Teil dieser Aufnahmen bildet die Grundlage von unseren früher veröffentlichten wissenschaftlichen Einzelstudien, ein anderer Teil ist unveröffentlicht, und die entsprechenden Aufnahmen wurden noch keinen Identifikationstests unterzogen. Die Aussagen im vierten Teil erheben in ihrer gesamten Zusammenstellung in der Folge nicht den Anspruch auf eine durchgängige Verifikation im Rahmen bestehender wissenschaftlicher Standards, sondern sind als Thesen im Hinblick auf absehbare und zu erwartende allgemeine Befunde formuliert. In diesem Sinne werden sie auch in den Materialien exemplarisch illustriert.

Die Schrift als solche erhebt derart nur den Anspruch, diejenigen Argumente, Erfahrungen und Beobachtungen noch einmal zusammenzustellen und in einen gegenseitigen Zusammenhang zu bringen, welche es erlauben, die Behauptung zurückzuweisen, Vokale würden über Formantmuster physikalisch repräsentiert. Gelingt dieses Vorhaben, so stellt sich — wir wiederholen uns — die Frage nach dem stimmhaften Sprachlaut erneut als ein eigentliches Grundlagenproblem.

Die Zusammenstellung der Argumente ist ganz auf den Zusammenhang von Einzellaut, erkanntem Vokal, Lautspektrum und Formantfrequenzen ausgerichtet und zugleich beschränkt. Auf Formantampplituden und -bandbreiten etwa, um zwei Beispiele anderer Aspekte zu erwähnen, welche für die Erörterung spezieller Erscheinungen von Bedeutung sein können, wird nicht im Einzelnen eingegangen.

Es muss auffallen, dass die vorliegenden Darstellungen aus einer entschieden psychophysischen Perspektive heraus formuliert sind. Auf Produktion und Perzeption von Lauten wird zwar Bezug genommen, aber nur in allgemeiner Weise: Auf die Produktion wird Bezug genommen, weil der Formantbegriff selbst sich auf Resonanzen des Vokaltrakts bezieht und diese Bezugnahme in den Argumentationen immer wieder verdeutlicht werden muss. Auf die Perzeption wird Bezug genommen, weil vorausgesetzt wird, dass die erörterten Laute jeweils spezifischen Vokalen zugeordnet, das heisst wahrnehmend identifiziert werden können. Über diese allgemeinen Verweise hinaus bleiben aber Fragen der Produktion und der Perzeption in der Regel ausgeklammert.

Diese Ausklammerung entspricht in keiner Weise einer Auffassung, Aspekte von Produktion und Perzeption seien für die physikalische Beschreibung von Vokalen unwichtig. Die Ausklammerung hat nur den Sinn, die psychophysische Frage nach dem Vokal in der einfachsten Art zu stellen und deutlich zu machen: Gesetzt, eine lautliche Äußerung — oder eine Reproduktion von ihr, oder eine Klang- oder Geräuschsynthese — wird wahrnehmend als ein bestimmter Vokal erkannt, welches beschreibbare physikalische Merkmal oder welches Ensemble physikalischer Merkmale repräsentiert dann diesen Vokal?

Die genannte Ausklammerung geht mit der Beschränkung auf isoliert gebildete oder aus einem Lautkontext isolierte einzelne stimmhafte orale Vokallaute einher. Nasalisierung sowie syntaktischer und semantischer Kontext werden derart mit ausgeklammert, und im Hinblick auf verschiedene Arten der Phonation wird nur auf geflüsterte Vokallaute eingegangen, und auch dies nur am Rande, im Rahmen von Zusätzen. Auch diese zweite Beschränkung gehört zur Intention, die anstehende Frage in einfacher Weise anzugehen.

Wie am Ende der Erörterung deutlich wird, bestehen Gründe, die Psychophysik stimmhafter Sprachlaute als eine Phänomenologie zu verstehen und zu betreiben: nicht von einem Modell auszugehen und davon abgeleitete Einzelexperimente durchzuführen, sondern eine offene und sich fortlaufend erweiternde Zusammenstellung von stimmlichen Äusserungen anzugehen, verbunden mit einer sich gleichzeitig entwickelnden Art der Beschreibung ihrer physikalischen Merkmale.

Unter dieser Perspektive mag einsichtig werden, weshalb die vorliegende Schrift trotz enger Begrenzung auf das Phonetische nicht als Veröffentlichung eines entsprechenden universitären Instituts, sondern als Veröffentlichung eines Instituts einer Hochschule der Künste erfolgt. In Umkehrung vieler Ansätze der bestehenden Literatur gehen wir nicht von einem «Normalfall» des Sprechens aus und behandeln von ihm ausgehend «andere» Arten von Äusserungen als «Spezialfälle», wie beispielsweise emotional geprägte Äusserungen mit entsprechenden Variationen von Grundtonhöhe und Lautstärke, oder Äusserungen in der sogenannten «Kopfstimme» oder Rufen oder Singen. Eine solche Auffassung hält keiner Alltagserfahrung und keiner Erfahrung des künstlerischen Ausdrucks stand.

Stimmliche und sprachliche Äusserungen kennen zunächst keine allgemeine Norm, und die einzige verlässliche Repräsentation von Stimme und geäusselter Sprache, auf welche sich das Nachdenken und der Entwurf einer Empirie beziehen kann, ist die künstlerische oder interpretierende Äusserung. Nur sie, die Kunst vermag das «Künstliche» — sprich die Reduktion, die Normierung, die Kodierung — je spezifischer Äusserungen darzustellen und gleichzeitig, wenn auch immer nur ansatzweise, zu überwinden. Nur die Kunst betreibt eine quasi-systematische Variation stimmlicher und sprachlicher Äusserungen, ohne welche jedes Nachdenken über stimmliche und sprachliche Charakteristika und jede Untersuchung dieser Charakteristika der Gefahr ausgesetzt ist, Befunde von konkreten und je spezifischen Äusserungen als Befunde allgemeiner Charakteristika zu interpretieren. Wir werden darauf im Nachwort zurückkommen.

Vokallaute, als isolierte Klänge oder Geräusche wahrgenommen, sind der Möglichkeit nach für sich verständlich, und diese Tatsache gehört zum Kern der Stimme und der Sprache selbst. Von dieser Voraus-

setzung gehen die nachfolgenden Erläuterungen aus. Die Vokale — genauer formuliert die Differenzen verschiedener Vokale einer Einzelsprache — müssen sich demnach physikalisch abbilden. Es ist absehbar, so soll die Erörterung deutlich werden lassen, dass diese Abbildung nicht von einem physikalischen Modell her abgeleitet werden kann, sondern als eine Leistung der Stimme selbst beschrieben werden muss.

## **Gliederung**

Die vorliegende Veröffentlichung ist in zwei Teile, Text und Materialien, gegliedert. Der Text liegt hier in gedruckter Form vor, die Materialien sind aufgrund ihres Umfangs digital veröffentlicht und über das Internet via <http://subtexte08.ipf.zhdk.ch> zugänglich.

Der Text ist seinerseits in fünf Teile gegliedert, gefolgt von einem Nachwort:

- Der erste Teil formuliert den bestehenden theoretischen Grundsatz zur Frage, welche Merkmale einen Vokal physikalisch repräsentieren.
- Der zweite Teil stellt Überlegungen dar, wie sie sich unserer Lektüre der Fachliteratur gemäss im Hinblick auf den Grundsatz aufdrängen und einem intellektuellen Nachvollzug von ihm entgegenstehen.
- Der dritte Teil formuliert Thesen zur tatsächlichen Beziehung von Vokallaut, Spektrum und Formantmuster. Diese Thesen beziehen sich auf unsere im ersten Teil der Einleitung genannten Aufnahmen, Erfahrungen und Beobachtungen.
- Der vierte Teil erläutert, weshalb diese Überlegungen, Erfahrungen und Beobachtungen in ihrer Zusammenstellung eine Falsifikation des bestehenden theoretischen Grundsatzes darstellen.
- Der fünfte Teil kommentiert die sich dadurch ergebende Ausgangslage und verweist auf die Notwendigkeit einer anstehenden Phänomenologie und neuen Theoriebildung.
- Das Nachwort beinhaltet einige ergänzende Erläuterungen und Hinweise.

Die Materialien enthalten einerseits ausgewählte Zitate und Darstellungen von Einzelbefunden aus der Literatur, zum Teil verbunden mit zusätzlichen Hinweisen und Kommentaren, andererseits exemplarische Darstellungen akustischer Analysen von natürlichen Vokalisationen, Lautimitationen und Vokalsynthesen.

Weil in den Materialien englische Zitate nicht ins Deutsche übersetzt werden, bleiben auch die Literaturhinweise auf englischsprachige Veröffentlichungen unübersetzt.

### **Begriffe und Bezeichnungen**

Zur Erleichterung der Lektüre seien nachfolgend die Verwendung zentraler Ausdrücke des Texts mit ihren Schreibweisen und Abkürzungen erläutert.

**Vokallaut, Vokal** Der Ausdruck «Vokallaut» bezeichnet einen einzelnen konkreten Laut mit sprachlichem Wert, als Phon. Gemäss Literatur werden Vokallaute in eckiger Klammer angegeben, wie zum Beispiel [a]. Zum Teil werden zusätzlich auch suprasegmentale Merkmale mit angegeben, wie zum Beispiel in der Unterscheidung von [a:] in «Kahn» und [a] in «Kamm» (langer und kurzer Vokallaut).

Der Ausdruck «Vokal» bezeichnet eine Klasse von Vokallauten einer Einzelsprache, als Phonem. Konkrete einzelne Vokallaute als Phone werden derart zu abstrakten Klassen von Vokalen als Phoneme zugeordnet. Gemäss Literatur werden Vokale zwischen zwei Schrägstrichen angegeben, wie zum Beispiel /a/.

Vokale werden gemäss dem Internationalen Phonetischen Alphabet (Revision 2005) angegeben.

Insofern sich die dargestellten Überlegungen, Erfahrungen und Beobachtungen ab Teil 2 auf einzelne Vokale beziehen, so betrifft dies die langen Vokale der deutschen Sprache (Standard) /i, y, e, ø, ε, a, o, u/. Hinzu kommt der Vokal /ɑ/, um die Aussprache von Schweizern mit einzubeziehen, sowie der Vokal /ɔ/, um die spektralen Erscheinungen im Zwischenbereich von /a-ɑ/ und /o/ erörtern zu können. Diese Vokale werden im Text häufig zusammenfassend in drei Gruppen

aufgeführt, als vordere Vokale /i, y, e, ø, ε/, Vokale /a-ɑ/ und hintere Vokale /ɔ, o, u/. Die Bezeichnung «vordere Vokale» und «hintere Vokale» wird aus der Literatur übernommen, hat aber darüber hinaus keine weitere Bedeutung. Insbesondere spielt die in ihr assoziierte Beziehung zur Zungenlage bei der Lautbildung keine Rolle.

Deutsche Vokale stehen im Vordergrund, weil die meisten unserer bisherigen Erfahrungen und Beobachtungen die deutsche Sprache betrifft. Die entsprechenden Feststellungen sind aber in ihren Grundzügen auf andere Einzelsprachen übertragbar.

**Spektrum** Der Ausdruck «Spektrum» bezieht sich auf ein Schallspektrum eines Vokallauts, in der Regel als Ergebnis einer Fourier-Analyse. In bestimmten Fällen kann sich der Ausdruck auch auf ein Spektrogramm beziehen, weil in vielen empirischen Untersuchungen Formantwerte nicht anhand eines Fourier-Spektrums, sondern anhand eines Spektrogramms eingeschätzt oder gegengeprüft werden. Die beiden Arten der spektralen Darstellung weisen wichtige Unterschiede auf, doch ist unsere Erörterung mit wenigen Ausnahmen derart allgemein gehalten, dass diese Unterschiede in ihr keine substantielle Rolle spielen. In den genannten Ausnahmefällen werden Spektrum (als Fourier-Spektrum) und Spektrogramm einzeln bezeichnet und voneinander unterschieden.

Der Ausdruck «Klangspektrum» bezieht sich auf ein Spektrum eines stimmhaften Vokallauts.

Der Ausdruck «Partialtonspektrum» bezieht sich auf eine interpretierte Reihe von Harmonischen im Spektrum.

Der Ausdruck «relatives spektrales Energiemaximum» bezieht sich auf eng begrenzte Frequenzanteile eines Spektrums, für welche dieses im Vergleich mit den direkt vorangehenden beziehungsweise nachfolgenden Frequenzanteilen deutlich erhöhte Energie aufweist. In der Literatur werden solche relativen Maxima in der Regel anhand der Beurteilung einer spektralen Hüllkurve als spektrale Hüllkurvenspitzen bestimmt.

**Grundtonhöhe** Der Ausdruck «Grundtonhöhe» bezeichnet hier sowohl die wahrgenommene Tonhöhe eines Vokallauts wie die gemessene Grundfrequenz des Vokalklangs. (Eine Unterscheidung von Grundtonhöhe und Grundfrequenz wird nicht vorgenommen, weil sie in den Darstellungen nicht von Bedeutung ist.) In der Literatur wird die Grundtonhöhe häufig mit der Abkürzung F0 bezeichnet.

**Formant** Der Ausdruck «Formant» wird in der Literatur unterschiedlich verwendet. Insbesondere kann er entweder eine Resonanz als physikalische Eigenschaft des Vokaltrakts, eine spektrale Hüllkurvenspitze als physikalische Eigenschaft eines Vokallauts oder einen Filter («filter pole») als Wert einer Analyseverfahren der Sprachverarbeitung bezeichnen. Er kann auch zwei oder alle drei dieser Eigenschaften zugleich bezeichnen.

In der vorliegenden Schrift wird zunächst von einer grundsätzlichen Unterscheidung zwischen Resonanzen des Vokaltrakts und Formanten des erzeugten Vokallauts ausgegangen. Dies entspricht der eingenommenen Perspektive, nicht die Erzeugung des Lauts, sondern Letzteren selbst, als Klang- oder Geräuschereignis, zu erörtern.

Zu Beginn der vorliegenden Erörterung bezeichnet der Ausdruck «Formant» sowohl spektrale Hüllkurvenspitzen wie Filter von Analyseverfahren, weil in der Literatur eine solche Gleichsetzung in der Regel vorgenommen wird, wenn allgemeine Aussagen zu vokalspezifischen physikalischen Merkmalen formuliert werden. Im Verlauf der Argumentation wird sich dann erweisen, dass nach heutigen empirischen Untersuchungen und den in ihnen dargestellten Formantwerten der Begriff des vokalspezifischen Formanten auf Filter von Analyseverfahren eingeschränkt werden muss, mit dem Zusatz, dass diese Analyseverfahren Kontextwissen und regelbasierte Vergleiche von errechneten Formantwerten mit Spektren der analysierten Laute mit einbeziehen.

Abkürzungen für Formanten werden in der Literatur häufig nach Formantfrequenzen, -bandbreiten und -amplituden respektive -level (normierte Formantamplituden) differenziert. Auf eine solche Unterscheidung wird hier verzichtet. Einzelne Formanten werden nachfolgend in allgemeiner Weise als F1, F2, F3, ... F(i) bezeichnet und Konfigurationen von Formanten (Formantmuster) als F1-F2 oder

F1–F2–F3. Diese Abkürzungen beziehen sich in der Folge nur auf die jeweilige Ordnung und die Frequenzen der Formanten. Bandbreiten und Amplituden spielen in der Erörterung wie erwähnt keine substantielle Rolle.

Im Unterschied zu Formanten werden Resonanzen des Vokaltrakts mit den Abkürzungen R1, R2, R3 ... R(i) bezeichnet.

Teil 1  
Bestehender theoretischer Grundsatz,  
bestehende empirische Referenzen

# 1 Bestehender theoretischer Grundsatz

## 1.1 **Allgemeine akustische Charakteristik von Vokallauten**

Die Stimmbänder erzeugen einen Klang, und die Hohlräume von Rachen, Mund und Nase formen diesen Klang durch ihre Resonanzwirkung zu einem Laut. Je nach Stellung des Gaumens, der Zunge, der Lippen und des Kiefers sind die Hohlräume verschieden und mit ihnen die ausgeübte Resonanzwirkung. Deshalb können verschiedene stimmhafte Laute gebildet und wahrgenommen werden. Wird ein Laut wahrnehmend einer sprachlichen Einheit zugeordnet und findet sich im Hohlraum von Rachen und Mund kein Verschluss, so wird der Laut «Vokallaut» und die sprachliche Identität «Vokal» genannt (vgl. Einleitung).

Die Bildung eines Vokallauts erfolgt nach diesem Ansatz über zwei voneinander quasi-unabhängige Vorgänge: Klangbildung und Umformung des Klanges über Resonanzen oder Phonation und Artikulation. Die Klangbildung oder Phonation ist nicht vokalspezifisch, die jeweilige Resonanzwirkung oder Artikulation ist vokalspezifisch. Das diesem Verständnis nach entstehende zweiteilige Modell der Lautbildung wird als «Quelle-Filter Modell» bezeichnet.

Physiologisch gesehen entspricht die wahrgenommene sprachliche Identität eines Vokallauts einer bestimmten Artikulation im Sinne eines Ensembles möglicher Einstellungen des Vokaltrakts, welche quasi-identische (untereinander sehr ähnliche) Resonanzwirkungen ausüben.

Akustisch gesehen entspricht die wahrgenommene sprachliche Identität eines Vokallauts bestimmten Energiemaxima des Klangspektrums, welche ihrerseits quasi-identisch (sehr ähnlich) zu denjenigen anderer Laute desselben Vokals sind. In einer akustischen Analyse erscheinen diese Energiemaxima als spektrale Hüllkurvenmaxima, allgemein als Formanten bezeichnet.

Für Fälle geflüsterter Vokallaute besteht die Phonation nicht aus einem Klang, sondern aus einem Geräusch.

## **1.2 Im engen Sinne sprachspezifische akustische Charakteristik von Vokallauten**

Nicht alle, sondern in der Regel nur die ihrer Frequenz gemäss ersten zwei der in einem Spektrum eines Vokallauts vorgefundenen Formanten entsprechen der wahrgenommenen sprachlichen Identität (dem wahrgenommenen Phonem). Die darüber hinausgehenden höheren Formanten beziehen sich auf andere Qualitäten oder Identitäten des lautlichen Ausdrucks.

Ausnahmen bilden in bestimmten Sprachen «benachbarte» vordere sowie r-gefärbte Vokale, für welche die Frequenzen der ersten beiden Formanten von Einzellauten sich empirisch als quasi-identisch erweisen und nur der Unterschied in der jeweiligen Frequenz des dritten Formanten dem Unterschied der wahrgenommenen sprachlichen Identität entspricht.

## **1.3 Relativierung in Bezug auf allgemeine Sprechergruppen**

Der Vokaltrakt eines Kindes ist in der Regel wesentlich kleiner als derjenige eines Erwachsenen. Gleiches gilt für den Vokaltrakt einer Frau im Vergleich zu dem eines Mannes. In der Folge muss die allgemein formulierte akustische Entsprechung von Vokallauten und Formantmustern bezogen werden auf die allgemeinen alters- und genderbezogenen Sprechergruppen von Kindern, Frauen oder Männern: Für jeweils eine Sprechergruppe mit für sie durchschnittlicher Vokaltraktgrösse entsprechen Laute eines Vokals physiologisch gesehen einer bestimmten Artikulation mit einer bestimmten Resonanzwirkung und akustisch gesehen einem bestimmten Formantmuster.

## **1.4 Einschränkung auf stimmhafte orale Vokallaute**

Der Vokaltrakt weist für verschiedene Arten der Phonation im Bereich der Glottis nicht dieselbe Geometrie und also auch nicht dieselbe Re-

sonanzwirkung auf. Deshalb unterscheiden sich die Formantmuster von stimmhaften, geflüsterten und mit Knarrstimme (Laryngalisierung, «creak phonation») produzierten Lauten jeweils eines Vokals. In der Folge muss die allgemein formulierte akustische Entsprechung von Vokal und Formantmuster auch bezogen werden auf die Phonationsart: Für jeweils eine Sprechergruppe mit für sie durchschnittlicher Vokaltraktgröße und jeweils einer Art der Phonation entsprechen Laute eines Vokals physiologisch gesehen einer bestimmten Artikulation mit einer bestimmten Resonanzwirkung und akustisch gesehen einem bestimmten Formantmuster.

Bestehende empirische Referenzwerte für Formantmuster stimmhafter Vokallaute sind verbreiteter als entsprechende Werte für Vokallaute mit anderer Art der Phonation, und methodische Probleme der Formantbestimmung stellen sich für verschiedene Arten der Phonation nicht in gleicher Weise. Die nachfolgende Erörterung wird deshalb auf stimmhafte Vokallaute eingeschränkt. Auf Laute mit anderer Art der Phonation wird nur am Rande und im Rahmen von Zusätzen eingegangen.

Die akustische Analyse nasaler Laute stellt spezifische methodische Probleme, auf welche hier ebenfalls nicht eingegangen wird. Deshalb wird die nachfolgende Erörterung weiter eingeschränkt auf stimmhafte orale Vokallaute.

## **1.5 Einschränkung auf isolierte Vokallaute**

Die Wahrnehmung von Vokallauten kann von deren syntaktischem Kontext abhängen: In bestimmten Fällen kann ein Vokallaut in einem direkten Zusammenhang mit anderen Lauten eine bestimmte sprachliche Identität erkennen lassen, welche zu einer anderen wechselt, wird der Laut als isoliertes Klangfragment wahrgenommen.

Die Wahrnehmung von Vokallauten kann in ähnlicher Weise vom semantischen Kontext abhängen: In bestimmten Fällen kann ein Vokallaut in einem Bedeutungszusammenhang eine bestimmte sprachliche Identität erkennen lassen, welche wiederum zu einer anderen wechselt, wenn der Laut als isoliertes Klangfragment wahrgenommen wird.

In der Folge gilt die Entsprechung von wahrgenommenen sprachlichen Identitäten eines Vokallauts mit den ersten zwei oder drei Resonanzen des Vokaltrakts beziehungsweise Formanten nur für isoliert gebildete Vokallaute ohne unmittelbaren syntaktischen oder semantischen Kontext und für isolierte Klangfragmente, welche ausserhalb eines unmittelbaren syntaktischen oder semantischen Kontexts wahrgenommen und identifiziert werden können. Sie werden hier alle, unabhängig vom realen Kontext, in welchem sie gebildet werden, als Monophthonge bezeichnet.

### **1.6 Einschränkung auf Vokallaute mit quasi-konstanter Klangcharakteristik**

Stimmhafte Vokallaute mit quasi-konstanter Grundtonhöhe und quasi-konstanter Lautstärke weisen in der Regel eine quasi-periodische Charakteristik der Schallwelle auf. Im Hinblick auf die physikalische Repräsentation des Vokals lässt sich die entsprechende spektrale Charakteristik solcher Vokallaute anhand jeweils eines einzelnen für den Laut durchschnittlichen Partialtonspektrums beschreiben, mit einbezogen der entsprechenden spektralen Hüllkurve und ihrer Maxima sowie dem entsprechenden Formantmuster.

Dies gilt nicht für Vokallaute, für welche die Grundtonhöhe oder die Lautstärke im Laut selbst stark variieren. Um die sich dann ergebenden Fragen der spektralen Repräsentation und des möglichen Einflusses solcher Variationen auf die Wahrnehmung der Laute auszuklammern, wird die nachfolgende Erörterung weiter eingeschränkt auf Vokallaute mit quasi-konstanter Klangcharakteristik. Auf Laute ohne solche Klangcharakteristik wird wiederum nur am Rande und im Rahmen von Zusätzen eingegangen.

### **1.7 Einschränkung auf eine Sprachgemeinschaft**

Die Entsprechung von wahrgenommenen Lauten eines Vokals mit den ersten zwei oder drei tieferen Resonanzen des Vokaltrakts beziehungsweise tieferen Formanten des Vokalspektrums gilt zunächst nur innerhalb ein und derselben Sprachgemeinschaft von Sprechenden und Hörenden: Für diese besteht eine Quasi-Konstanz der Produk-

tion und Perzeption, gemäss welcher Lautvariationen entweder ein und derselben Vokalidentität oder aber verschiedenen Vokalidentitäten zugeordnet werden.

Auf die Methode einer konkreten empirischen Überprüfung dieser Konstanz der Zuordnung — und mit ihr auch deren Relativierung — wird hier nicht weiter eingegangen. Die vorliegende Erörterung setzt grundsätzlich voraus, dass die in ihr diskutierten Laute in einem konkreten Identifikationstest von Hörerinnen und Hörern, welche für einen solchen Wahrnehmungstest vorgängig geschult werden, eine übereinstimmende Zuordnung der jeweiligen Laute von  $\geq 50\%$  zu jeweils einem Vokal ergibt.

Inwiefern Entsprechungen im oben genannten Sinne formuliert werden können, welche den verschiedenen Sprachgemeinschaften übergeordnet sind, wird hier offen gelassen.

### **1.8 Theoretischer Grundsatz zur physikalischen Repräsentation des Vokals**

Gesetzt, es gilt:

- Vokallaute werden von Individuen einer der drei allgemeinen Sprechergruppen einer bestimmten Sprachgemeinschaft gebildet.
- Vokallaute werden entweder als stimmhafte isolierte orale Laute gebildet oder aus ihrem unmittelbaren syntaktischen und semantischen Kontext als stimmhafte orale Klangfragmente ohne Transitionen an deren Anfang und Ende herausgelöst.
- Vokallaute werden mit quasi-konstanter Grundtonhöhe und Lautstärke gebildet und weisen eine quasi-periodische Charakteristik der Schallwelle auf.
- Vokallaute werden von Individuen der oben genannten Sprachgemeinschaft gehört und dabei jeweils einem Vokal zugeordnet.

Dann gilt für jeweils einen einzelnen Vokallaut :

- Seine wahrgenommene sprachliche Identität als Vokal entspricht physiologisch gesehen einer spezifischen Einstellung des Vokaltrakts, welche eine spezifische Resonanzwirkung auf den Klang der Kehle ausübt, bezogen auf die (ihrer Frequenz nach) ersten beiden (in wenigen Fällen vorderer und r-gefärbter Vokale bestimmter Sprachen ersten drei) Resonanzen.
- Seine wahrgenommene sprachliche Identität entspricht (deshalb) akustisch gesehen den ersten beiden (oder ersten drei) tieferen Formanten des Spektrums.

Dies bedeutet für zwei als verschiedene Vokale wahrgenommene Laute, dieselben Voraussetzungen angenommen:

- Der Unterschied ihrer wahrgenommenen sprachlichen Identitäten entspricht physiologisch gesehen zwei verschiedenen Einstellungen des Vokaltrakts mit verschiedenen Resonanzwirkungen, bezogen auf die ersten beiden (oder ersten drei) tieferen Resonanzen.
- Der Unterschied ihrer wahrgenommenen sprachlichen Identitäten entspricht akustisch gesehen verschiedenen Mustern der ersten beiden (oder der ersten drei) tieferen Formanten ihrer Spektren.

Dies bedeutet für die Laute jeweils eines Vokals, aber verschiedener allgemeiner Sprechergruppen:

- Ihre wahrgenommene sprachliche Identität als derselbe Vokal entspricht physiologisch gesehen sprechergruppen-spezifisch verschiedenen Resonanzwirkungen der Vokaltrakte, bezogen auf die ersten beiden (oder ersten drei) tieferen Resonanzen.
- Ihre wahrgenommene sprachliche Identität entspricht (deshalb) akustisch gesehen sprechergruppen-spezifisch verschiedenen Mustern der ersten beiden (oder ersten drei) tieferen Formanten der jeweiligen Spektren.

Diese Formulierungen bilden den Kern des bestehenden theoretischen Grundsatzes zur physikalischen Repräsentation des Vokals.

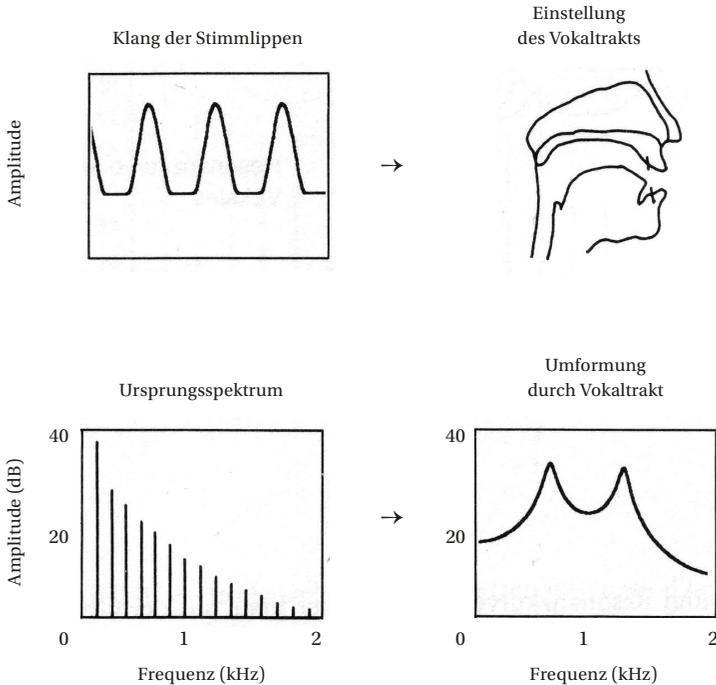
## 1.9 Formalisierung des theoretischen Grundsatzes

Für isolierte stimmhafte orale Vokallaute mit quasi-konstanter Klangcharakteristik, von Individuen einer Sprachgemeinschaft und einer Sprechergruppe gebildet, gilt:

- Laute derselben Vokalidentität = quasi-identische (untereinander ähnliche) R1-R2 (in wenigen Fällen vorderer Vokale bestimmter Sprachen R1-R2-R3) und gleichzeitig quasi-identische F1-F2 (respektive F1-F2-F3).
- Laute verschiedener Vokalidentitäten = untereinander grundsätzlich unähnliche R1-R2 (respektive R1-R2-R3) und gleichzeitig untereinander grundsätzlich unähnliche F1-F2 (respektive F1-F2-F3).

## 1.10 Illustration

Abbildung 1 zeigt ein Beispiel einer Illustration, wie sie als Typus in vielen Veröffentlichungen vorzufinden ist. (Die Illustration ist in dem Sinne vereinfacht, als dass in ihr die Differenzierung der tatsächlichen Charakteristik des Quellenspektrums einerseits und der Austrahlungsimpedanz bei der Abstrahlung des Klanges in den Raum andererseits fehlt. Auf diese Differenzierung wird in der vorliegenden Erörterung aber grundsätzlich nicht eingegangen, weil sie für die Argumentation keine substantielle Rolle spielt.)

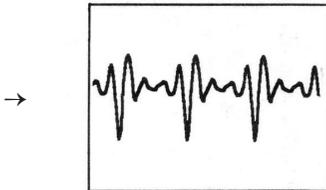


**Abbildung 1**

Illustration des theoretischen Grundsatzes. In Anlehnung an Fant (1960)<sup>1</sup>.

1 Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton, p. 20.

Ausgesandte Schallwelle



Vokalspektrum

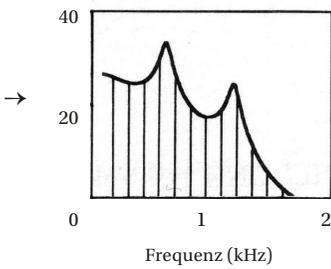
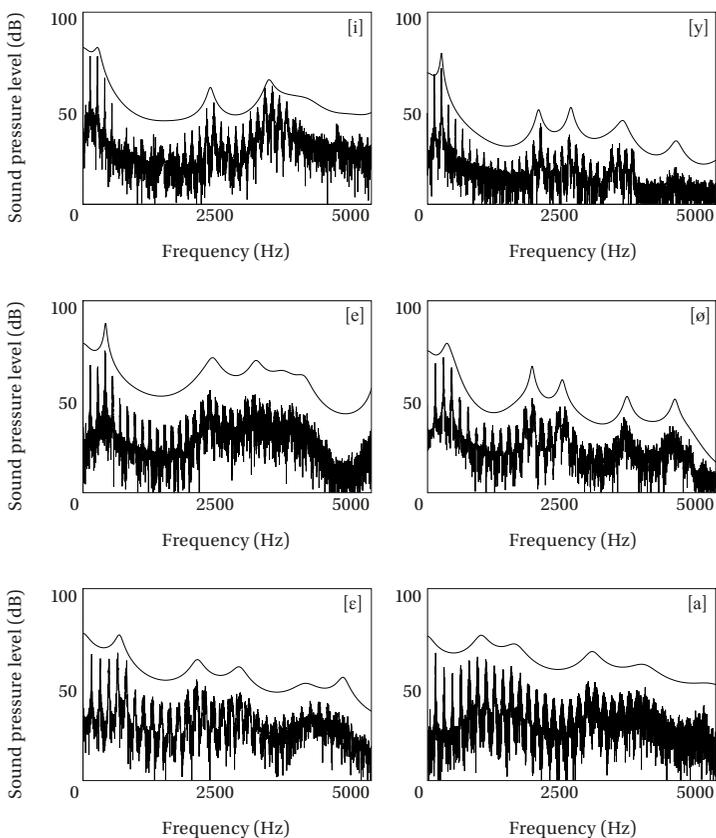
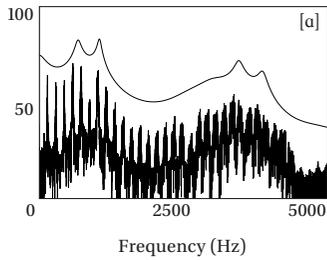
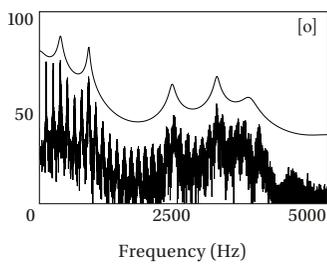
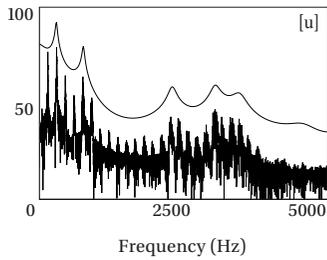


Abbildung 2 stellt Beispiele von Spektren, Filterkurven und Formantmustern (Maxima der Filterkurven) speziell ausgewählter Laute verschiedener Vokale dar. Auch diese Art der Illustration, sich auf die akustische Perspektive beschränkend, ist in der Literatur sehr verbreitet.



**Abbildung 2**

Speziell zur Illustration des theoretischen Grundsatzes ausgewählte Beispiele von Spektren und Filterkurven (LPC-Kurven) von isolierten Lauten verschiedener Vokale, von Männern auf Grundtonhöhen im Bereiche von 120-160 Hz erzeugt.



## 2 Bestehende empirische Referenzen

### 2.1 Allgemeine Referenzen

Die erste breit angelegte statistische Untersuchung der Entsprechung von Vokalen und Formantmustern, alle drei allgemeinen Sprechergruppen mit einbeziehend, stammt von Peterson and Barney (1952, siehe Tabelle 1 und Abbildung 3)<sup>2</sup>, bezogen auf die amerikanisch-englische Sprache. Sie wurde in der Folge zu einer Standardreferenz in der Fachliteratur.

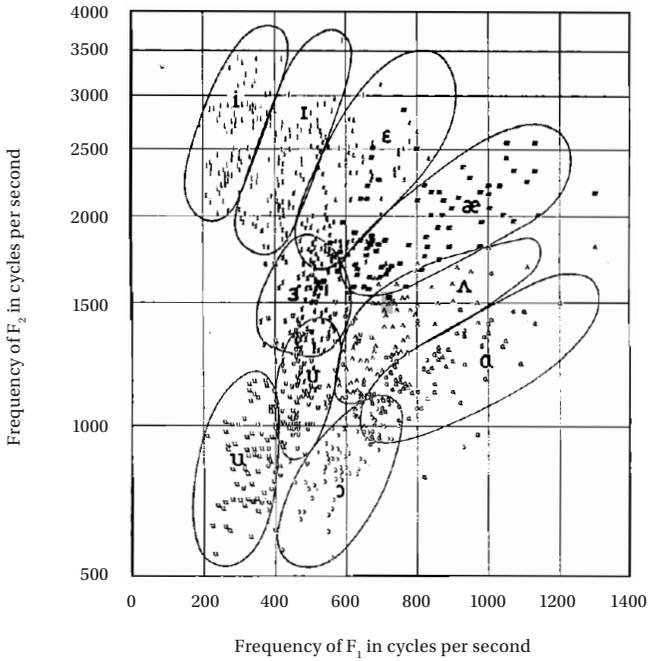
		i	I	ε
Fundamental frequencies (cps)	M	136	135	130
	W	235	232	223
	Ch	272	269	260
Formant frequencies (cps)	M	270	390	530
	W	310	430	610
	Ch	370	530	690
F <sub>1</sub>	M	2290	1990	1840
	W	2790	2480	2330
	Ch	3200	2730	2610
F <sub>2</sub>	M	3010	2550	2480
	W	3310	3070	2990
	Ch	3730	3600	3570
Formant amplitudes (db)	L <sub>1</sub>	-4	-3	-2
	L <sub>2</sub>	-24	-23	-17
	L <sub>3</sub>	-28	-27	-24

**Tabelle 1**

Formantstatistik von Peterson and Barney / (1952) für Vokale des Amerikanisch-Englischen. Mittlere Werte für Grundtonhöhen und Formantfrequenzen F1-F2-F3 in Hz sowie Formantamplituden (als Formantlevel) L1-L2-L3 in dB, getrennt nach allgemeinen Sprechergruppen von Männern (M), Frauen (W) und Kindern (Ch).

2 Peterson, G.E. and Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24/2, 175-184.

æ	a	ɔ	U	u	ʌ	ʒ
127	124	129	137	141	130	133
210	212	216	232	231	221	218
251	256	263	276	274	261	261
660	730	570	440	300	640	490
860	850	590	470	370	760	500
1010	1030	680	560	430	850	560
1720	1090	840	1020	870	1190	1350
2050	1220	920	1160	950	1400	1640
2320	1370	1060	1410	1170	1590	1820
2410	2440	2410	2240	2240	2390	1690
2850	2810	2710	2680	2670	2780	1960
3320	3170	3180	3310	3260	3360	2160
-1	-1	0	-1	-3	-1	-5
-12	-5	-7	-12	-19	-10	-15
-22	-28	-34	-34	-43	-27	-20



**Abbildung 3**

Illustration der Verteilung der ersten beiden Formanten für Vokale des Amerikanisch-Englischen, Peterson and Barney / (1952). Abszisse = Formantfrequenzen in Hz für  $F_1$ , Ordinate = Formantfrequenzen in Hz für  $F_2$ . (Reproduktion mit freundlicher Genehmigung von Peterson and Barney [1952]. Copyright 1952, Acoustical Society of America.)



Die Untersuchung von Peterson and Barney wurde von Hillenbrand et al. (1995)<sup>3</sup> mit neuer Aufnahme- und Messmethode (Digitalisierung, LPC-Analyse) und auf zwölf Vokale erweitert repliziert. Die entsprechenden Werte sind in Tabelle 2 aufgeführt.

		/i/	/I/	/e/	/ε/	/æ/
F0	M	138	135	129	127	123
	W	227	224	219	214	215
	C	246	241	237	230	228
F1	M	342	427	476	580	588
	W	437	483	536	731	669
	C	452	511	564	749	717
F2	M	2322	2034	2089	1799	1952
	W	2761	2365	2530	2058	2349
	C	3081	2552	2656	2267	2501
F3	M	3000	2684	2691	2605	2601
	W	3372	3053	3047	2979	2972
	C	3702	3403	3323	3310	3289
F4	M	3657	3618	3649	3677	3624
	W	4352	4334	4319	4294	4290
	C	4572	4575	4422	4671	4409

**Tabelle 2**

Formantstatistik von Hillenbrand et al. (1995) für Vokale des Amerikanisch-Englischen. Mittlere Werte für Grundtonhöhen F0 und Formantfrequenzen F1-F2-F3-F4 in Hz, getrennt nach allgemeinen Sprechergruppen von Männern (M), Frauen (W) und Kindern (C).

3 Hillenbrand, J., Getty, L.A., Clark, M.J. and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97/5, 3099-3111.

/a/	/ɔ/	/o/	/U/	/u/	/ʌ/	/ɜ/
123	121	129	133	143	133	130
215	210	217	230	235	218	217
229	225	236	243	249	236	237
768	652	497	469	378	623	474
936	781	555	519	459	753	523
1002	803	597	568	494	749	586
1333	997	910	1122	997	1200	1379
1551	1136	1035	1225	1105	1426	1588
1688	1210	1137	1490	1345	1546	1719
2522	2538	2459	2434	2343	2550	1710
2815	2824	2828	2827	2735	2933	1929
2950	2982	2987	3072	2988	3145	2143
3687	3486	3384	3400	3357	3557	3334
4299	3923	3927	4052	4115	4092	3914
4307	3919	4167	4328	4276	4320	3788

Parallel zu Peterson and Barney veröffentlichte Fant (1959)<sup>4</sup> eine ähnlich breit angelegte statistische Untersuchung, bezogen auf die schwedische Sprache, aber nur die beiden allgemeinen Sprechergruppen von Frauen und Männern mit einbeziehend (siehe Tabelle 3). Auch diese Statistik wurde in der Folge zu einer Standardreferenz in der Fachliteratur.

V	S	F <sub>0</sub> Hz	L <sub>0</sub> dB	F <sub>1</sub> Hz	L <sub>1</sub> dB
[u]	Gj-n	125	0.5	325	3
	M-r	256	6	270	6
	M	127	-0.5	307	4.5
	F	222	2	340	5
[o]	Gj-n	125	2	405	6
	M-r	256	5	380	8
	M	132	-0.5	402	6
	F	223	1	433	7
[o1]	Gj-n	125	1.5	500	7
	M-r	257	5	510	8
	M	123	-0.5	487	6
	F	217	-2	518	5
[a]	Gj-n	125	1.5	600	7
	M-r	250	4	650	6
	M	126	-1	582	7.5
	F	225	-1	682	4
[a1]	Gj-n	125	2.5	680	6
	M-r	255	3	770	8
	M	124	-1	680	6
	F	215	-1	860	4
[a3]	Gj-n	125	2.5	560	6
	M-r	257	3	600	6
	M	125	0.5	606	7
	F	213	0	785	5
[ä]	Gj-n	125	2.5	480	5
	M-r	255	4	535	7
	M	125	0	438	6
	F	214	0	545	5

Diese drei Auflistungen von vokalspezifischen Formantmustern stellen vermutlich die in der Literatur am weitestverbreiteten Referenzen dar, auf welche in allgemeinen — nicht auf Einzelsprachen bezogenen — Erörterungen der physikalischen Merkmale von Vokalen Bezug genommen wird.

F <sub>2</sub>	L <sub>2</sub>	F <sub>3</sub>	L <sub>3</sub>	F <sub>4</sub>	L <sub>4</sub>
Hz	dB	Hz	dB	Hz	dB
640	2.5	2400	-40	3500	-40
740	-5	2550	-37	(3200)	(-35)
730	-8	2230	-37	3300	-38
690	-6	2900	-43	(4000)	(-45)
700	-1	2450	-32	3200	-29
850	-10	2800	-38	---	---
708	-2	2460	-31	3150	-33
815	-9	2840	-38	(3600)	(-37)
800	-1	2530	-26	3150	-25
900	0	2800	-35	(3000)	(-35)
825	1	2560	-26	3250	-28
840	-6	2825	-36	(3500)	(-40)
935	0	2620	-18	3150	-22
1125	0	2800	-18	---	---
940	4	2480	-21	3290	-20
1075	4	2930	-22	(3800)	(-28)
1075	-1	2720	-18	3350	-18
1250	1	2800	-15	---	---
1070	1	2520	-10	3345	-20
1195	4	2830	-23	---	---
1740	-7	2470	-12	3200	-16
1740	-7	2900	-20	---	---
1550	-3	2450	-12	3400	-15
1820	-6	2950	-18	(3600)	(-17)
1870	-7	2480	-9	3250	-18
1870	-8	2600	-18	---	---
1795	-9	2385	-12	3415	-19
2140	-11	2860	-20	---	---

V	S	F <sub>0</sub> Hz	L <sub>0</sub> dB	F <sub>1</sub> Hz	L <sub>1</sub> dB
[e]	Gj-n	125	1	325	4
	M-r	256	7	320	7
	M	124	0	334	6
	F	215	1	365	6
[i]	Gj-n	140	2	275	2
	M-r	256	7	270	7
	M	128	0	256	3
	F	218	3	278	5
[ü]	Gj-n	140	3	275	2
	M-r	257	8	260	8
	M	128	1	257	4.5
	F	215	5	270	6
[u2]	Gj-n	125	1	290	3
	M-r	256	6	300	6
	M	126	0	283	5.5
	F	217	4	300	5
[u3]	Gj-n	125	2.5	375	4
	M-r	257	9	370	10
	M	125	-0.5	416	6
	F	216	2	410	7
[ö]	Gj-n	125	2.5	345	5
	M-r	257	8	350	8
	M	126	0	363	6.5
	F	215	2	372	5
[ö1]	Gj-n	125	2.5	470	5
	M-r	257	4	500	10
	M	124	-0.5	524	6
	F	217	1	565	8

**Tabelle 3 (S. 34–37)**

Formantstatistik von Fant (1959) für Vokale des Schwedischen. Werte für Grundtonhöhen F<sub>0</sub> und Formantfrequenzen F<sub>1</sub>–F<sub>2</sub>–F<sub>3</sub>–F<sub>4</sub> in Hz sowie Formantamplituden (als Formantlevel) L<sub>1</sub>–L<sub>2</sub>–L<sub>3</sub>–L<sub>4</sub> in dB. Gj-n = Werte für einen einzelnen erwachsenen Sprecher; M-r = Werte für eine einzelne erwachsene Sprecherin; M = mittlere Werte für Männer; F = mittlere Werte für Frauen.

4 Fant, G. (1959). *Acoustic analysis and synthesis of speech with applications to Swedish*. Ericsson Technics No. 1.

F <sub>2</sub>	L <sub>2</sub>	F <sub>3</sub>	L <sub>3</sub>	F <sub>4</sub>	L <sub>4</sub>
Hz	dB	Hz	dB	Hz	dB
2210	-11	2650	-12	3400	-20
2200	-13	2700	-12	---	---
2050	-12	2510	-13	3400	-16
2540	-15	2950	-18	---	---
2205	-17	3100	-12	3500	-17
2200	-23	3100	-15	---	---
2066	-23	2960	-20	3400	-23
2520	-24	3450	-24	(3900)	(-28)
2050	-12	2300	-15	3325	-21
2070	-22	2820	-17	3300	-24
1928	-17	2421	-19	3300	-24
2480	-21	2920	-23	3575	-26
1690	-12	2170	-15	3300	-22
1760	-14	2270	-14	3100	-34
1633	-13	2140	-17	3314	-26
1910	-18	2600	-22	3450	-34
1070	-12	2500	-20	3500	-27
1050	-14	2400	-22	---	---
1070	-7	2315	-24	3300	-29
1175	-11	2700	-31	3600	-35
1735	-10	2250	-12	3400	-22
1800	-2	2250	-8	---	---
1690	-9	2200	-11	3390	-20
2000	-14	2610	-18	3650	-28
1195	-7	2550	-16	3300	-24
1300	-11	2600	-22	3500	-35
1103	-4	2430	-22	3250	-19
1290	-6	2730	-21	3700	-29

## 2.2 Referenz für die deutsche Sprache (Standard)

In Hinsicht auf Vokale der standarddeutschen Sprache liegt eine statistische Untersuchung von Pätzold and Simpson (1997)<sup>5</sup> vor, welche die Laute von zwei der drei allgemeinen Sprechergruppen, Frauen und Männer, dokumentiert (siehe Tabelle 4, beschränkt auf Monophthonge).

(a)	Vowel	F1	lq	uq	F2
	i:	329	292	385	2316
	I	391	350	442	2136
	y:	342	312	401	1667
	Y	406	369	466	1612
	e:	431	382	495	2241
	ε	592	517	687	1944
	ø:	434	391	482	1646
	œ	509	452	584	1767
	a:	779	665	880	1347
	a	751	651	838	1460
	o:	438	395	487	953
	ɔ	573	509	660	1174
	u:	350	319	405	1048
	U	450	387	504	1184
	ɐ	590	494	685	1608
	ə	420	369	482	1746

lq	uq	F3	lq	uq	n
2125	2496	2796	2644	3000	719
1905	2348	2867	2660	3026	1014
1485	1833	2585	2437	2691	125
1475	1735	2631	2518	2779	105
1949	2472	2871	2691	3055	579
1774	2100	2867	2679	2997	607
1551	1739	2573	2440	2708	108
1620	1870	2640	2488	2757	48
1236	1439	2785	2644	2941	452
1346	1583	2841	2679	2983	810
789	1102	2835	2673	2990	269
1055	1279	2825	2668	2965	279
885	1220	2760	2624	2877	299
1074	1302	2749	2570	2960	434
1430	1754	2829	2679	2968	610
1554	1948	2811	2649	2968	1338

(b)	Vowel	F1	lq	uq	F2
	i:	290	266	337	1986
	ɪ	343	303	380	1803
	y:	310	278	349	1505
	Y	374	333	401	1431
	e:	372	328	436	1879
	ɛ	498	443	552	1639
	ø:	375	333	414	1458
	œ	437	407	501	1504
	a:	639	570	700	1225
	ɑ	608	529	674	1309
	o:	380	352	429	907
	ɔ	506	455	550	1060
	u:	309	283	343	961
	U	382	332	439	1058
	ɐ	503	440	561	1372
	ə	370	321	424	1521

**Tabelle 4 (S. 38–41)**

Formantstatistik von Pätzold and Simpson (1997) für Vokale des Deutschen (Standard). Mittlere Werte für Formantfrequenzen F1-F2-F3 in Hz, mit zusätzlichen Angaben für untere und obere Quartile (lq und uq). Teil (a) Werte für Frauen, Teil (b) Werte für Männer.

- 5 Pätzold, M. and Simpson, A. P. (1997). Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. In: Simpson, A. P., Kohler, K. J. and Rettstadt, T., eds., *The Kiel Corpus of Read/Spontaneous Speech — Acoustic data base, processing tools and analysis results*, Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK) 32, pp. 215–247.

lq	uq	F3	lq	uq	n
1813	2106	2493	2328	2668	710
1640	1956	2483	2309	2632	1009
1362	1624	2205	2117	2321	126
1345	1529	2284	2131	2445	102
1700	2006	2486	2324	2614	580
1517	1755	2451	2299	2599	613
1383	1505	2220	2104	2319	107
1376	1598	2179	2121	2327	49
1166	1292	2477	2316	2613	452
1224	1386	2466	2317	2618	831
774	1009	2415	2269	2570	265
992	1127	2415	2295	2546	283
835	1145	2366	2247	2520	291
966	1165	2363	2225	2522	435
1253	1463	2430	2288	2570	610
1391	1660	2368	2219	2547	1286

### 2.3 Zusatz: Verweise auf andere statistische Angaben

Verweise auf andere statistische Untersuchungen und zusätzliche Angaben, welche für die vorliegende Erörterung von Interesse sind, finden sich in den Materialien. Hierzu zählen insbesondere modellartige Werte für Formantfrequenzen, Vergleiche von Formantmustern verschiedener allgemeiner Sprechergruppen sowie Angaben für Formantfrequenzen bei verschiedenen Grundtonhöhen der Vokallaute.

Teil 2  
Überlegungen

## 3 Vokale und Anzahl Formanten

### 3.1 **Inkonstanz der Anzahl vokalspezifischer spektraler Energiemaxima bei hinteren Vokalen und /a-a/**

Bestimmte Vokallaute von hinteren Vokalen und /a-a/ können im Spektrum nur eine ausgeprägte vokalspezifische Hüllkurvenspitze aufweisen.

Andere Laute derselben Vokale können im Spektrum zwei deutlich ausgeprägte vokalspezifische Hüllkurvenspitzen aufweisen.

Empirisch erweist sich die Anzahl vokalspezifischer spektraler Energiemaxima als inkonstant.

### 3.2 **Inkonstanz der Entsprechung von vokalspezifischen spektralen Energiemaxima mit vokalspezifischen Formantmustern**

Weisen Vokallaute von hinteren Vokalen und /a-a/ im Spektrum nur eine ausgeprägte vokalspezifische Hüllkurvenspitze auf, so ergibt die Formantanalyse gemäss Literatur häufig zwei nahe beieinanderliegende Formantfrequenzen, weshalb solche Erscheinungen als «Formantverschmelzungen» («formant merging») bezeichnet werden. Für diese Vokallaute entsprechen sich in der Folge die spektrale Hüllkurvenspitzen und die ersten beiden Formanten nicht, weder in der Anzahl noch in den Frequenzen.

Die Inkonstanz der Anzahl vokalspezifischer Hüllkurvenspitzen von Vokallauten jeweils eines Vokals stellt damit die grundsätzliche Beziehung von Formanten und spektraler Hüllkurve zur Frage.

Formanten können nicht in direkter Weise mit relativen spektralen Energiemaxima parallel gesetzt werden, und sie entsprechen auch keinem anderen spektralen Merkmal in direkter Weise.

Formanten erweisen sich in der Folge als ein Konstrukt einer Analyse-  
methode (vgl. dazu Kapitel 6.1).

### **3.3 Inkonstanz der Anzahl vokalspezifischer Formanten bei vorderen und r-gefärbten Vokalen**

Wie in Teil 1 dargestellt, können Paare von Vokallauten zweier verschiedener vorderer Vokale in einzelnen Sprachen dieselben ersten beiden Formantfrequenzen aufweisen. Solche Paare von Vokallauten unterscheiden sich physikalisch erst anhand des dritten Formanten.

Auch für r-gefärbte Vokale kann sich F3 als vokalspezifisch erweisen.

Für solche Einzelsprachen gilt in der Folge, dass die hinteren und der eine Teil der vorderen Vokale über unterschiedliche Muster F1–F2 physikalisch beschreibbar sind, der andere Teil vorderer Vokale und allenfalls auch der r-gefärbten Vokale aber nur über unterschiedliche Muster F1–F2–F3.

Empirisch erweist sich die Anzahl vokalspezifischer Formanten ihrerseits als inkonstant.

### **3.4 Zusatz: «Unechte» Formanten**

Bei einzelnen Sprecherinnen oder Sprechern kann bei bestimmten Vokallauten zwischen dem ersten und dem zweiten (in bestimmten Fällen von vorderen und r-gefärbten Vokalen auch zwischen dem zweiten und dritten) Formanten ein zusätzliches spektrales Maximum auftreten. Dieses Maximum wird nach methodischen Regeln der Formantbestimmung nicht als vokalspezifisch, sondern als Merkmal der

jeweils spezifischen Stimmcharakteristik der Sprecherin oder des Sprechers interpretiert und deshalb als «unechter Formant» («spurious formant») bezeichnet.

Solche «unechten» spektrale Hüllkurvenspitzen, welche zwischen den als vokalspezifisch interpretierten Formanten beobachtet werden können, sind ihrerseits in Zusammenhang mit der Inkonstanz der Anzahl vokalspezifischer spektraler Energiemaxima zu bedenken.

### **3.5 Zusatz: Inkonstanz der Anzahl vokalspezifischer Formanten in Synthesen**

Synthetisch erzeugte Vokalklänge können für die meisten Vokale sowohl über eine Dreiformanten- wie über eine Zweiformanten-Synthese gut erkennbar erzeugt werden. Für bestimmte Vokale, insbesondere für hintere Vokale und Vokale /a-ɑ/, ist dies auch über eine Einformant-Synthese möglich.

Der Vergleich der jeweils verwendeten Formantmuster der ersten beiden Synthesearten weist dabei Unterschiede für F2 insbesondere bei vorderen Vokalen auf. Gleiches gilt für F1 im Vergleich der verwendeten Formantmuster in der zweiten und dritten Syntheseart. (Allerdings wurden in entsprechenden Studien die Grundtonhöhen bei den Synthesen nicht systematisch variiert.)

Synthesen bestätigen in dieser Weise die Inkonstanz der Anzahl zu beobachtender vokalspezifischer Formanten. Darüber hinaus verwenden Synthesen mit verschiedener Anzahl von Formanten (verwendeten Filtern) auf Unterschiede einzelner Formantfrequenzen für Klänge jeweils derselben Vokale.

## 4 Vokale und Grundtonhöhe

### 4.1 **Grundtonhöhe, erster Formant und «Qualität» von einzelnen Vokalen**

Vokalspezifische Formantmuster sind, dem theoretischen Grundsatz entsprechend, unabhängig von der Grundtonhöhe jeweiliger Einzellaute.

Die Frequenzen des ersten Formanten aller Vokale, wie sie in derzeitigen Formantstatistiken aufgeführt werden, liegen im Bereich möglicher Grundtonhöhen für Sprecherinnen und Sprecher jeweils einer allgemeinen Sprechergruppe. (Ausnahmen bilden in vereinzelt Statistiken die Angaben für die ersten Formanten von /a-a/.) Tiefste Werte für die erste Formantfrequenz weisen dabei /i, y, u/, mittlere Werte /e, ø, o/ und nachfolgend /ε, ɔ/, höchste Werte /a-a/ auf.

Steigt bei der Erzeugung von Vokallauten die Grundtonhöhe über die Frequenz des jeweils ersten Formanten von /i, y, u/ und nähert sie sich den ersten Formantfrequenzen von /e, ø, o/ an, so müssten die drei Vokale /i, y, u/ nach theoretischem Grundsatz unverständlich werden, weil der jeweils erste für sie als vokalspezifisch angenommene Formant physikalisch nicht mehr repräsentiert werden kann. Die Vokale /i, y, u/ wären also «schlechter» — in ihrer Bildung und physikalischen Repräsentation eingeschränkter — als die übrigen Vokale, und gleiches gälte für /e, ø, o/ im Vergleich zu /ε, a, α, ɔ/ sowie für /ε, ɔ/ im Vergleich zu /a-a/.

Die Möglichkeit, dass die Grundtonhöhe oberhalb der (nach Angabe von Formantstatistiken) ersten Formantfrequenz jeweils eines Vokals liegen kann, bedeutet nach theoretischem Grundsatz, eine akustisch bedingte unterschiedliche Qualität von Vokalen anzunehmen.

Nur schon die Alltagserfahrung erlaubt es aber, einer solchen verallgemeinernden Folgerung zu widersprechen. Bilden Sprecherinnen oder Sprecher jeweils einer allgemeinen Sprechergruppe Vokallaute,

übersteigt dabei die Grundtonhöhe die (gemäss Formantstatistiken) ersten Formantfrequenzen von /i, y, u/ und nähert sie sich den ersten Formantfrequenzen von /e, ø, o/ an, so können bei entsprechend guter stimmlicher Fähigkeit der Sprecherinnen und Sprecher die ersten drei genannten Vokale durchaus deutlich erkennbar gebildet werden. Gleiches gilt, wenn auch im Alltag weniger offensichtlich und nur bei sehr guten Stimmen nachvollziehbar, für die Vokale /e, ø, o/ auf Grundtonhöhen oberhalb von statistischen Werten ihrer ersten Formantfrequenzen.

In diesem Zusammenhang sind insbesondere im Alltag zu beobachtende Ausdrucksweisen zu bedenken, welche eine Variation der Grundtonhöhe von über einer Oktave aufweisen, weil sie die Bedeutung der sich eröffnenden Problematik für eine akustische Theorie des Vokals deutlich machen. Bei einzelnen Sprecherinnen und Sprechern betrifft dies nicht nur emotionale Äusserungen oder (anderweitig) lautstarke Äusserungen, sondern ihre Prosodie ganz allgemein.

In diesem Zusammenhang sind auch stimmliche und sprachliche Äusserungen im Rahmen des Musik- und Sprechtheaters sowie des Films und des Fernsehens besonders zu bedenken, weil die ausgeprägte Variation der Grundtonhöhe zu den zentralen Merkmalen des Singens und Sprechens im Bereich der Unterhaltung und des Ästhetischen gehört.

#### **4.2 Grundtonhöhe, spektrale Hüllkurve, Formantmuster und «Qualität» von Vokallauten**

Steigt die Grundtonhöhe eines Lauts, so nimmt der Frequenzabstand der Partialtöne im Spektrum zu, und die Bestimmung der spektralen Hüllkurve und ihrer Maxima wird schwierig. Gleiches gilt für die Berechnung von Formantfrequenzen. Dies lässt nach dem theoretischen Grundsatz den Schluss zu, dass die Verständlichkeit der Laute aller Vokale in einer allgemeinen Beziehung zu ihrer Grundtonhöhe steht: der Tendenz nach je tiefer, desto besser, je höher, desto schlechter.

Über diese allgemeine Beziehung von Grundtonhöhe und Bestimmbarkeit der spektralen Hüllkurve beziehungsweise der Formantfrequenzen hinaus besteht zudem eine je frequenzspezifische und va-

riable Beziehung: Je nach statistisch angegebenen Frequenzen von vokalspezifischen Formanten ergeben sich Fälle von Vergleichen, für welche dem theoretischen Grundsatz nach angenommen werden kann, dass Einzellaute eines bestimmten Vokals auf höheren Grundtonhöhen in ihrer Vokalidentität besser erkannt werden als Einzellaute desselben Vokals auf tieferen Grundtonhöhen. Diese Fälle betreffen Laute auf höheren Grundtonhöhen, für welche die Frequenzen einzelner Partialtöne mit relativen Amplitudenmaxima nahe der statistischen Werte der Formantfrequenzen für den zur Frage stehenden Vokal liegen, verglichen mit Lauten auf tieferen Grundtonhöhen, für welche alle Partialtöne innerhalb des vokalspezifischen Frequenzbereichs zwischen diesen Formantfrequenzen liegen. Für letztere können sich in der Folge die Formanten nur undeutlich oder gar nicht als Hüllkurvenspitzen ausprägen, und der Vokal wird, so wäre zu folgern, deshalb schlechter erkennbar. Dies aber würde zur Annahme führen, dass die Verständlichkeit von Vokallauten nicht nur in einer allgemeinen, sondern auch einer diskontinuierlichen Beziehung zu ihrer Grundtonhöhe steht: Tiefere Vokallaute wären zwar in der Regel besser verständlich als höhere Vokallaute, aber die Verständlichkeit würde zusätzlich vom jeweiligen Verhältnis von Grundtonhöhe, vokalspezifischem Formantmuster und sich ergebendem Partialtonspektrum abhängen.

Der Zusammenhang der Grundtonhöhe mit der Bestimmung von spektralen Hüllkurven und Formantmustern bedeutet nach theoretischem Grundsatz, eine akustisch bedingte unterschiedliche Qualität von Vokallauten auf verschiedenen Grundtonhöhen anzunehmen.

Die mögliche Alltagserfahrung, auf welche hingewiesen wurde, erlaubt, auch diesen Verallgemeinerungen zu widersprechen, und eine Theorie von Vokalen als Elementen einer Sprache, welche in sich einen grundsätzlichen qualitativen und zugleich diskontinuierlichen Aspekt gemäss Grundtonhöhe formuliert, steht in Kontrast zu der — möglicherweise als «Sensation» zu bezeichnenden — Charakteristik eines stimmhaften Elements einer Sprache, innerhalb eines Bereichs von Grundtonhöhen verständlicher sprachlicher Äusserungen unabhängig von ihr zu sein.

## 5 Formantmuster und allgemeine Sprechergruppen

### 5.1 Grundtonhöhe, spektrale Hüllkurve, Formantmuster und «Qualität» von Vokalen von Kindern, Frauen und Männern

Die Folgerung gemäss theoretischem Grundsatz weiter auslegend, dass die Verständlichkeit der Laute eines Vokals in direkter Beziehung zu ihrer Grundtonhöhe steht, führt zur Auffassung, dass die Qualität der Vokale selbst grundsätzlich auch abhängig ist von der jeweiligen allgemeinen Sprechergruppe: am besten verständlich bei Lauten von Männern, mittelmässig bei Lauten von Frauen und am schlechtesten bei Lauten von Kindern.

Die erläuterten Zusammenhänge der Grundtonhöhe mit spektralen Hüllkurven und mit Formantmustern bedeutet nach theoretischem Grundsatz, eine grundsätzlich unterschiedliche Qualität von Vokalen verschiedener allgemeiner Sprechergruppen anzunehmen.

Die Alltagserfahrung erlaubt es, auch dieser Verallgemeinerung zu widersprechen, und eine Theorie von Vokalen als Elemente einer Sprache, welche in sich einen grundsätzlichen qualitativen Aspekt gemäss Alter und Gender (genauer: gemäss Grösse der Stimm- und Sprechorgane) formuliert, steht in Kontrast zur — möglicherweise erneut als «Sensation» zu bezeichnenden — Charakteristik eines stimmhaften Elements einer Sprache, quasi-unabhängig von den individuellen Konstitutionen oder Ausdrucksweisen der je Sprechenden zu sein.

Vokale als solche haben weder Alter noch Gender noch Geschlecht. Wenn Individuen im gegenseitigen Vergleich Vokallaute «besser» oder «schlechter» bilden, so liegt dies — funktionelle und im engeren Sinne pathologische Störungen ausgenommen — an den Individuen, nicht an den Vokalen. Vokale sind, als Sprachlaute, in der Regel für alle entweder der Möglichkeit nach bildbar oder nicht bildbar. Vo-

kale sind kein Attribut eines Individuums, sondern ein Element der Sprache. Vokale sind vom Individuum «abstrahiert».

## 5.2 Verschiedene Formantmuster für gleiche Vokale

Empirische Referenzwerte für vokalspezifische Formantmuster werden in der Literatur als für jeweils eine allgemeine Sprechergruppe — Kinder, Frauen und Männer — unterschiedlich und also gruppenspezifisch dargestellt. Erklärt werden diese Unterschiede in erster Linie nicht mit den je verschiedenen durchschnittlichen Grundtonhöhen, sondern mit den je verschiedenen durchschnittlichen Vokaltraktgrößen.

Dieser Auffassung gemäss ist anzunehmen, dass drei verschiedene Formantmuster nicht nur auf jeweils verschiedenen Grundtonhöhen, sondern auch auf ein und derselben Grundtonhöhe denselben Vokal repräsentieren können: Frauen und Männer sind in der Regel imstande, auf einer Grundtonhöhe von Kindern — beispielsweise auf einer Grundtonhöhe von 300 Hz — deutlich erkennbare Vokallaute zu produzieren. Drei Laute desselben Vokals — gebildet von einem Kind, einer Frau und einem Mann — hätten dann, dem theoretischen Grundsatz folgend, drei vollständig verschiedene Formantmuster, obwohl sie als derselbe Vokal wahrgenommen werden.

Der nach theoretischem Grundsatz bestehende Zusammenhang von vokalspezifischen Formantmustern und allgemeinen Sprechergruppen lässt eine physikalische Repräsentation eines Vokals anhand verschiedener Formantmuster annehmen.

Die Überlegung verlängernd, liesse sich auch annehmen, dass es Fälle gibt, bei denen Frauen und Männer imstande sind, Einzellaute eines Vokals auf höheren Grundtonhöhen als Kinder, aber mit tieferen Formantmustern zu erzeugen, ohne Änderung der Vokalidentität. Die sich ergebende Problematik für die hier diskutierten Lautkonfigurationen wird besonders deutlich, wenn für Laute von /a, ɑ, ɔ, o, u/ (statistische Werte für  $F1-F2 < 2$  kHz) die höheren Frequenzen  $> 2$  kHz in Wahrnehmungstests gefiltert werden, so dass die

manipulierten Klänge keine deutliche Unterscheidung im Hinblick auf die Wahrnehmung von stimmlichen Eigenheiten der allgemeinen Sprechergruppen mehr erlauben.

All dies gilt, so versteht sich, auch für den eingeschränkten Vergleich von Frauen und Männern.

Eine solche Auffassung beziehungsweise Folgerung widerspricht aber erneut der Anforderung einer psychophysischen Parallele von wahrgenommener Identität und physikalischer Repräsentation: Entweder sind Formantmuster als solche vokalspezifisch, dann repräsentieren deutlich verschiedene Formantmuster — unabhängig von der Grundtonhöhe — nicht denselben Vokal, oder Formantmuster, für sich genommen, sind nicht direkt vokalspezifisch, dann ist die Annahme von sprechergruppenspezifischen und zugleich vokalspezifischen Formantmustern in Frage zu stellen.

### 5.3 Gleiche Formantmuster für verschiedene Vokale

Die im vorangehenden Kapitel angestellte Überlegung erneut verlängert, verleitet des Weiteren zur Annahme, dass ein einzelnes Formantmuster zwei verschiedene Vokale repräsentieren kann: Wenn Laute eines Vokals, von einem Kind erzeugt, höhere vokalspezifische Formantmuster aufweisen als Laute desselben Vokals, von einem Mann erzeugt, wenn die Grundtonhöhe dabei keine substantielle Rolle spielt und wenn das Formantmuster des Kindes im Frequenzbereich möglicher Formantmuster des Mannes liegt, dann müssen in bestimmten Fällen die Formantmuster dieser Laute des Kindes bei Lauten eines anderen Vokals des Mannes zu finden sein.

Der nach theoretischem Grundsatz bestehende Zusammenhang von vokalspezifischen Formantmustern und allgemeinen Sprechergruppen lässt ein einzelnes Formantmuster als physikalische Repräsentation zweier verschiedener Vokale annehmen.

Diese Folgerung ist ihrerseits kritisch. Zu den bereits genannten Gründen kommt hinzu, dass die Annahme der Möglichkeit einer doppelten Repräsentation, gemäss welcher ein einzelnes Formantmuster den Lauten zweier verschiedener Vokale akustisch entsprechen kann, dem theoretischen Grundsatz in direkter Weise widersprechen würde.

Diese Überlegungen führen zu einer entschiedenen Skepsis gegenüber der These, vokalspezifische Formantmuster seien grundsätzlich und in durchgängiger Weise abhängig von der allgemeinen Sprechergruppe beziehungsweise der Vokaltraktgrösse. Eine solche grundsätzliche Abhängigkeit ist nur schon intellektuell nicht nachvollziehbar. Darüber hinaus mag die einfache Tatsache, dass eine Vokalsynthese, auf einer bestimmten Grundtonhöhe und mit zwei Formanten  $\leq 1,5$  kHz durchgeführt, einen hinteren Vokal oder /a/ oder /a/ erkennen lassen kann, und dass dieser Vokal ändert, wenn F1 und F2 substanziiell geändert werden, in paradigmatischer Weise die hier erörterte Problematik verdeutlichen.

Zugleich deutet sich eine alternative Erklärung für die bestehenden empirischen Befunde an, welche grundsätzlich sprechergruppenspezifische Formantmuster nachzuweisen scheinen: Vokalspezifische spektrale Frequenzanteile, und mit ihnen je berechnete vokalspezifische Formantmuster, können von der Grundtonhöhe abhängen.

Es fällt im vorliegenden Zusammenhang auf, dass keine der in der Literatur als Referenzen bewerteten Formantstatistiken Frequenzwerte von Formantmustern für systematisch variierte Grundtonhöhen innerhalb jeweils einer Sprechergruppe aufführt. Der empirische Nachweis der These, allgemeine Sprechergruppen würden grundsätzlich je verschiedene vokalspezifische Formantmuster aufweisen, und diese Unterschiede seien nicht auf die — jeweils gleichzeitig beobachteten — verschiedenen Grundtonhöhen der Laute, sondern auf je verschiedene Vokaltraktgrössen zurückzuführen, fehlt derart in der Literatur, und er wird, so lässt sich aufgrund der vorliegenden Ausführungen vermuten, auch nicht geführt werden können.

## 5.4 Lücke in der Begründung

Formantstatistiken von Frauen und Männern lassen darauf schließen, dass Erwachsene imstande sind, unabhängig von der Grundtonhöhe Laute für alle Varianten von F1-F2, wie sie für Kinder als statistische Werte für /y, ø, a, ɑ, ɔ, o, u/ aufgeführt werden, zu bilden. (In diesem Zusammenhang ist es bedeutungslos, als welche Vokale die Laute der Erwachsenen dabei jeweils konkret erkannt werden.) Auch wenn Frauen und Männer grössere Vokaltrakte als Kinder besitzen, so wäre es ihnen also zumindest in Hinsicht auf F1-F2 von /y, ø, a, ɑ, ɔ, o, u/ dennoch möglich, Laute mit denselben Formantmuster zu bilden, wie letztere für Kinder aufgeführt werden.

Wenn dem so ist, wenn Frauen und Männer grundsätzlich dieselben F1-F2 wie die Kinder zu produzieren imstande sind, wenn also zumindest für einen substanziellen Teil der Vokale gleiche vokalspezifische Formantmuster für alle allgemeinen Sprechergruppen möglich wären, weshalb dann die Unterschiede?

Auch wenn die bestehenden statistischen Werte in der Regel die These bestätigen, dass die Formantmuster jeweils eines Vokals bei Kindern höher liegen als bei Erwachsenen und bei Frauen höher als bei Männern, so bestehen dennoch Ausnahmen, für welche einzelne gemessene Formantfrequenzen oder gar Formantmuster F1-F2 oder F1-F2-F3, wie sie in einzelnen Formantstatistiken erscheinen, für Vokallaute von Männern oberhalb derjenigen von Frauen liegen. Dies wirft dieselbe Frage auf.

Der nach theoretischem Grundsatz beschriebene Zusammenhang von vokalspezifischen Formantmustern und allgemeinen Sprechergruppen antwortet nicht auf die Frage, weshalb gleiche vokalspezifische Formantmuster trotz verschiedener Vokaltraktgrössen zwar möglich wären, aber nicht realisiert werden.

Es könnte sich darüber hinaus erweisen, dass diese Formulierung auch in ganz allgemeiner Weise zutrifft und alle vokalspezifischen Formantmuster F1-F2 beziehungsweise F1-F2-F3, zu beobachten

bei Kindern, der Möglichkeit nach auch von Frauen und Männern produziert werden können.

### **5.5 Zusatz: Formantmuster stimmhafter und geflüsterter Vokallaute**

Empirische Studien, in welchen stimmhafte und geflüsterte Vokallaute miteinander verglichen werden, verweisen auf substanzielle Unterschiede in den je ermittelten vokalspezifischen Formantmustern. Insbesondere die ersten beiden Formantfrequenzen sind bei geflüsterten Lauten jeweils eines Vokals in zum Teil ausgeprägtem Masse höher als bei stimmhaften Lauten.

Dies relativiert seinerseits die direkte Entsprechung von Formantmustern und Vokalen: Laute desselben Vokals können nicht nur aufgrund verschiedener Vokaltraktgrößen, sondern auch aufgrund verschiedener Arten der Phonation verschiedene Formantmuster aufweisen.

Darüber hinaus deuten Vergleiche von veröffentlichten Formantfrequenzen geflüsterter und stimmhafter Vokallaute darauf hin, dass alle  $F1$  sowie  $F2 \leq 1,5$  kHz von geflüsterten Lauten von Männern in der Regel über den entsprechenden  $F1$  und  $F2$  von stimmhaften Lauten von Frauen liegen, bei jeweils gleicher erkannter Vokalidentität und trotz grösserer Vokaltraktgrösse der Männer. Ähnliches gilt für den Vergleich von geflüsterten Lauten von Frauen mit stimmhaften Lauten von Kindern. Auf  $F1$  beschränkt trifft dies auch für den Vergleich von geflüsterten Lauten von Männern mit stimmhaften Lauten von Kindern zu.

Diese Feststellung relativiert ihrerseits die Entsprechung von Vokaltraktgrösse und vokalspezifischen Formantmustern: Eine solche Entsprechung ist anhand der in der Literatur dargestellten Werte nur für jeweils eine identische Art der Phonation, nicht aber für den Vergleich verschiedener Arten der Phonation dokumentiert. Zu beachten ist zudem, dass Formantunterschiede für Laute jeweils eines Vokals in Zusammenhang mit der Variation der Phonation von stimmhaft versus geflüstert in der Regel grösser sind als entsprechende Formantunterschiede der verschiedenen allgemeinen Sprechergruppen.

Ein einzelnes Formantmuster kann also, so scheint es, nicht nur dann verschiedene Vokale physikalisch repräsentieren, wenn die entsprechenden Laute von Sprecherinnen und Sprechern verschiedener allgemeiner Sprechergruppen erzeugt werden, sondern auch dann, wenn eine einzelne Sprecherin oder ein einzelner Sprecher die Phonation variiert.

Auf solche Überlegungen wird aber in Teil 3 zurückzukommen sein: Vergleiche von Formantmustern stimmhafter und geflüsterter Laute, wie sie in der Literatur dokumentiert werden, beziehen sich nur auf jeweils eine durchschnittliche (tiefere) Grundtonhöhe der stimmhaften Laute, nicht aber auf einen Vergleich, welcher eine systematische Variation ihrer Grundtonhöhe mit einbezieht. (Eine solche experimentelle Anlage geht erneut davon aus, dass Formantmuster unabhängig von der Grundtonhöhe sind und in Vergleichen stimmhafter und geflüsterter Laute deshalb vernachlässigt werden können.)

## 6 Begriffe, Messmethode der Formantbestimmung

### 6.1 Formant und Lautspektrum

Die Unterscheidung der beiden Ausdrücke «Resonanz» und «Formant» als Unterscheidung einer Charakteristik des Vokaltrakts und einer Charakteristik des Lautspektrums lässt annehmen, dass das Spektrum eines Vokallauts grundsätzlich bestimmbar und auf vokalspezifische Frequenzbereiche bezogene Energiemaxima aufweist und dass die Frequenzen dieser spektralen Maxima den erwarteten beziehungsweise je berechneten Formantfrequenzen entsprechen.

Dabei ist zu bedenken, dass Formantfrequenzen als numerische Werte heute nicht (mehr) von der spektralen Hüllkurve abgeleitet, sondern als Filter eines analytischen Modells berechnet werden.

Wie in den Kapitel 3.1 und 3.2 erläutert, können Spektren von Lauten hinterer Vokale und /a–ɑ/ nur ein vokalspezifisches Energiemaximum aufweisen, auch wenn in der Formantanalyse — zum Teil unter Bezugnahme auf ein «phonetisches Wissen» und entsprechender manueller Anpassungen von Parametereinstellungen — zwei frequenzmässig nahe beieinanderliegende Formanten F1 und F2 berechnet werden. Dies widerspricht der Annahme, relative spektrale Energiemaxima beziehungsweise Hüllkurvenmaxima würden analytisch ermittelten Formanten grundsätzlich in Anzahl und Frequenz entsprechen.

Weist das Spektrum eines Lauts eines hinteren Vokals oder eines Vokals /a/ oder /ɑ/ nur ein vokalspezifisches Energiemaximum  $\leq 1,5$  kHz auf und werden die Parameter einer Formantanalyse so gewählt, dass sich dennoch zwei Formantfrequenzen bestimmen lassen, so liegt in einem substantziellen Teil solcher Fälle von Vokallauten F1 tiefer und F2 höher als das im Spektrum jeweils vorfindbare Energiemaximum. Dies bedeutet, dass das vokalspezifische spektrale Hüllkurvenmaximum dann einem relativen Minimum der sich bei der Formantanalyse ergebenden Filterkurven (etwa LPC-Kurven) entspricht. Auch dies steht in Kontrast zur Annahme einer Parallele von spektralen Energiemaxima und Formanten.

Mit steigender Grundtonhöhe wird die Bestimmung eines spektralen Energiemaximums beziehungsweise einer Hüllkurvenspitze zunehmend kritisch, und zwar aufgrund der zunehmenden Abstände der Frequenzen der Partialtöne (Einzelheiten siehe Kapitel 6.4). Dies verhindert seinerseits die Formulierung einer allgemeinen Entsprechung von spektralen Energiemaxima und Formantfrequenzen.

Bezogen auf derzeitige Verfahren der Formantanalyse und auf entsprechende numerische Werte von Formantmustern bezeichnet der Begriff des Formanten in der Folge nicht in allen Fällen zu berücksichtigender Vokallaute — und damit grundsätzlich nicht — ein reales Merkmal des Spektrums, sondern ein Artefakt der jeweiligen Analysemethode (siehe auch Kapitel 3).

Der Ausdruck «Formant» bezeichnet in der Literatur in der Regel nicht nur keine Charakteristik des Vokaltrakts — wenn er sich denn vom Ausdruck «Resonanz» unterscheidet —, sondern auch keine grundsätzliche Charakteristik des Lautspektrums.

Damit verliert die Annahme einer direkten Entsprechung von Resonanzen als physikalischer Eigenschaft des Vokaltrakts, spektralen Energiemaxima als physikalischer Eigenschaft des erzeugten Vokallauts und Frequenzen von Filtern, abgeleitet von Methoden der akustischen Analyse lautlicher Äusserungen, ihre Plausibilität.

## 6.2 Allgemeine Sprechergruppe und Vokaltraktgrösse

Der nach theoretischem Grundsatz bestehende Zusammenhang von vokalspezifischen Formantmustern und allgemeinen Sprechergruppen wird, wie bereits erläutert, mit dem durchschnittlichen Unterschied der jeweiligen Vokaltraktgrössen begründet.

Es ist anzunehmen, dass bestimmte Frauen grössere Vokaltrakte aufweisen als bestimmte Männer. Werden die beiden Gruppen entsprechenden Vokallaute miteinander verglichen, so ist im vorliegenden Zusammenhang folgende Konstellation von besonderem Interesse: Die Grundtonhöhen der Laute dieser Sprecherinnen entsprechen den

bestehenden statistischen Durchschnittswerten für Grundtonhöhen von Frauen allgemein, und sie liegen gleichzeitig über denjenigen der Laute dieser Sprecher. Für solche Vergleiche müsste nach theoretischem Grundsatz gelten, dass die vokalspezifischen Formantmuster der Laute dieser Sprecherinnen — trotz vergleichsweise höheren Grundtonhöhen — frequenzmässig tiefer liegen als die entsprechenden Formantmuster dieser Sprecher.

Gleichzeitig stellt sich die Frage einer systematischen Untersuchung der Beziehung von Vokaltraktgrösse und vokalspezifischen Formantmustern innerhalb einer allgemeinen Sprechergruppe.

Offen bleibt hier, ob — für sehr spezielle Fälle von Vergleichen — Kinder grössere Vokaltrakte als Erwachsene haben können.

Abgesehen von den fehlenden empirischen Grundlagen zu den hier aufgeworfenen Fragen verweisen diese Erläuterungen darauf, dass der theoretische Grundsatz nicht sprechergruppenspezifische, sondern für verschiedene Grössen von Vokaltrakten spezifische Formantmuster behauptet.

Der Ausdruck «allgemeine Sprechergruppe» ist nicht gleichzusetzen mit dem Ausdruck «Vokaltraktgrösse».

Zur Erleichterung der Lektüre werden wir in diesem zweiten Teil auch nachfolgend die beiden Ausdrücke «allgemeine Sprechergruppe» und «Vokaltraktgrösse» parallel verwenden, der Literatur entsprechend, auf welche sich die Ausführungen beziehen. In Teil 3 wird die Problematik als solche in anderem Zusammenhang wieder aufgegriffen, und die beiden Ausdrücke werden getrennt voneinander verwendet.

### **6.3 Objektivität von Formantanalysen im Allgemeinen**

Derzeit verwendete Analysemethoden zur Formantbestimmung gehen von einem natürlichen Vokallaut als realem Klang aus und berechnen modellartig diejenige Umformung eines Quellenklangs durch

eine Reihe von Filtern, welche dem realen Klang am besten entspricht. (Für Flüstervokale gilt dasselbe, bezogen auf jeweilige Geräusche.)

Dieses Verfahren muss sowohl eine bestimmte Charakteristik des Quellenklangs wie eine bestimmte Anzahl und Charakteristik von Filtern für den jeweils zu analysierenden Frequenzbereich annehmen. Wie nahe die Charakteristik des Quellenklangs der realen Phonation kommt, bleibt dabei offen. Gleiches gilt für die Frage, ob die Anzahl und Charakteristik der Filter der realen Artikulation und ihrer Resonanzwirkung tatsächlich entsprechen.

Aus einem Vokalklang allein kann man Formanten in der Folge nicht verlässlich bestimmen. Die Analyse verlangt im Minimum ein vorgängiges Wissen darüber, ob der zu analysierende Laut von einem Kind, einer Frau oder einem Mann vokalisiert wurde, in der Annahme, dass daraus zumindest annähernd die Vokaltraktgröße und mit ihr die zu setzende Anzahl der Filter für die Formantanalyse abgeleitet werden können.

Analytisch gewonnene Werte für Formantmuster werden zudem häufig mit den realen Energiemaxima des Lautspektrums oder -spektrogramms verglichen: Stimmen die in einem ersten Schritt — auf der Grundlage von Analyseparametern gemäss bestehender Standards — automatisch ermittelten Formantwerte nicht mit den realen spektralen Energiemaxima eines analysierten Lauts überein, so wird die Anzahl der Filter variiert und die Analyse so lange durchgeführt, bis sich eine solche Entsprechung einstellt. Die Charakteristik des Quellenklangs wird dabei in der Regel nicht verändert. Dies gilt allerdings nur für diejenigen Fälle, für welche übereinstimmende numerische Formantwerte und spektrale Energiemaxima innerhalb von Frequenzbereichen ermittelt werden können, welche sowohl dem allgemeinen theoretischen Grundsatz wie den bestehenden statistisch ermittelten Referenzwerten entsprechen. Ergeben sich auch bei einer solchen Variation keine Werte, welche erwarteten Formantmustern in erwarteten Frequenzbereichen entsprechen, so werden die jeweiligen Vokallaute häufig aus weiterführenden Untersuchungen ausgeschlossen. Ausnahmen bilden die in Kapitel 3.1 erläuterten «Formantverschmelzungen».

Formantanalysen setzen demzufolge individuelle analytische Fähigkeiten (ein Wissen um bestehende phonetische Grundsätze und Interpretationsregeln sowie eine analytische Erfahrung) und also ein vorgängiges Training der Analysierenden voraus, weil sie Kontextwissen, Vergleiche numerischer Werte mit dem Lautspektrum, verbunden mit einer visuellen Interpretation dieses Spektrums, sowie zum Teil manuell vorzunehmende Variationen von Filtereinstellungen und Wiederholungen von numerischen Analysen mit einschliessen. Formantanalysen sind also nicht vollständig objektivierbar. Wären sie dies, so besässen die Analysierenden als Individuen keine Funktion.

Methoden der Formantanalyse sind nicht in strengem Sinne objektivierbar, und sie sind dementsprechend auch nicht automatisierbar.

#### **6.4 Abhängigkeit und Beschränkung von Formantanalysen aufgrund ihrer Beziehung zu Grundtonhöhe und Sprechergruppe beziehungsweise Vokaltraktgrösse**

Zur grundsätzlichen Unmöglichkeit einer vollständigen Objektivierung der Formantanalyse kommt, wie in Kapitel 6.1 angedeutet, deren Abhängigkeit von der Grundtonhöhe hinzu: Mit steigender Grundtonhöhe wird nicht nur die Bestimmung von spektralen Energiemaxima beziehungsweise von spektralen Hüllkurvenspitzen zunehmend kritisch, weil die Frequenzabstände der Partialtöne zu gross werden, um das Energiemaximum beziehungsweise die Hüllkurve zu definieren, sondern auch die Bestimmung von Formanten im Rahmen bestehender Analyseverfahren.

Als allgemeine Richtwerte für kritische Grenzen der Frequenzen der Grundtonhöhe, oberhalb derer Formantanalysen unzuverlässig werden, sind dabei zwei Referenzwerte zu bedenken: die Hälfte der Frequenz des tiefsten ersten Formanten für jeweils eine Sprechergruppe beziehungsweise für einen Bereich von Vokaltraktgrössen und die Frequenz dieses tiefsten Formanten selbst.

Die erste Referenz ergibt sich aus der Überlegung, dass oberhalb einer Grundtonhöhe von  $\frac{1}{2} \times F1$  die Frequenzabstände der ersten beiden Partialtöne bereits derart gross sind, dass die Festlegung einer spektralen Hüllkurve und mit ihr die Einschätzung errechneter numerischer Formantfrequenzen problematisch werden. Entsprechend dieser Referenz und Bezug nehmend auf Standardwerte von Hillenbrand et al. (1995) für F1 von /i/ (tiefster Durchschnittswert für F1 in dieser Referenzstatistik) würden Formantanalysen kritisch für Grundtonhöhen ab:

- 226 Hz bei Kindern beziehungsweise entsprechend kleineren Vokaltraktgrössen
- 219 Hz bei Frauen beziehungsweise entsprechend mittleren Vokaltraktgrössen
- 171 Hz bei Männern beziehungsweise entsprechend grösseren Vokaltraktgrössen

Die zweite Referenz ergibt sich aus der Überlegung, dass oberhalb einer Grundtonhöhe des tiefsten Wertes von F1 für eine Sprechergruppe grundsätzlich nicht mehr alle F1 aller Vokale dieser Sprechergruppe unterschieden werden können, ganz abgesehen von der verschärften Problematik der Bestimmung einer spektralen Hüllkurve. Entsprechend dieser zweiten Referenz, und wiederum auf die oben genannte Statistik Bezug nehmend, würden Formantanalysen aussagegelos für Grundtonhöhen ab:

- 452 Hz bei Kindern beziehungsweise entsprechend kleineren Vokaltraktgrössen
- 437 Hz bei Frauen beziehungsweise entsprechend mittleren Vokaltraktgrössen
- 342 Hz bei Männern beziehungsweise entsprechend grösseren Vokaltraktgrössen

Die erste Aufstellung lässt zunächst deutlich werden, dass die Analyse von Vokallauten von Kindern und Frauen methodisch grundsätzlich als problematisch eingeschätzt werden muss. Der genannte kritische Frequenzwert liegt für Kinder deutlich unterhalb der empirisch ermittelten durchschnittlichen Grundtonhöhe ihres entspannten Sprechens, und also weisen die meisten Vokallaute insbesondere jüngerer Kinder deutlich höhere Grundtonhöhen auf. Der genannte

kritische Frequenzwert für Frauen entspricht gemäss Formantstatistiken ihrer durchschnittlichen Grundtonhöhe bei entspanntem Sprechen; in der alltäglichen Ausdrucksweise aber sind Vokallaute in einem Bereich von Grundtonhöhen bis eine Oktave über diesem Wert die Regel.

Die erste Aufstellung ist darüber hinaus auch auf die Möglichkeit von Männern zu bedenken, Vokallaute auf höheren Grundtonhöhen zu bilden. Die Problematik erscheint für Männer zwar weniger ausgeprägt als für Kinder und Frauen, betrifft aber dennoch einen substantziellen Teil ihrer stimmhaften sprachlichen Äusserungen.

Methoden der Formantanalyse erweisen sich als abhängig von der Grundtonhöhe und der Sprechergruppe beziehungsweise der Vokaltraktgrösse.

Die zweite Aufstellung lässt deutlich werden, dass Formantmuster für Vokallaute mit höheren Grundtonhöhen aus methodischen Gründen grundsätzlich nicht mehr ermittelt werden können. In diesem Zusammenhang sind insbesondere prosodische Variationen der Grundtonhöhe bis zu eineinhalb Oktaven sowie die mögliche Bandbreite verschiedener Stimmlagen von Sprecherinnen und Sprechern — und mit ihnen besonders Stimmlagen mit spontan hohen Grundtonhöhen — zu bedenken, wie sie sowohl im Alltag wie im Bereich des unterhaltenden, künstlerischen und interpretierenden Ausdrucks beobachtet werden können.

Formantanalysen sind in ihrem Aussagewert für Grundtonhöhen ab 350 Hz grundsätzlich kritisch, und sie sind für Grundtonhöhen ab 500 Hz methodisch nicht mehr abzustützen. Dies bedeutet, dass Formantanalysen nicht auf alle Fälle deutlich verstehbarer Vokallaute anwendbar sind.

In einem Teil der Literatur besteht die Tendenz, die methodische Problematik mit einer Charakteristik der Wahrnehmung gleichzusetzen, was zu zwei bereits in Kapitel 5.1 beziehungsweise 4.1 dargestellten Annahmen zurückführt: dass Vokale von Frauen und Kindern grundsätzlich weniger deutlich verständlich wären als Vokale von Männern und dass Vokallaute auf einer Grundtonhöhe von 500 Hz grundsätzlich nicht mehr deutlich voneinander unterscheidbar seien. Beide Annahmen aber widersprechen, wie erläutert, der tatsächlich erfahrbaren Charakteristik von Vokallauten.

### **6.5 Zusatz: Anpassungen von Parametern in der Formantanalyse und Inkonsistenz der Bezugnahme auf eine Vokaltraktgröße**

Wie in Kapitel 6.3 ausgeführt, werden bei derzeitigen Analysen von Formanten einerseits vor der Analyse deren Parameter in Abhängigkeit der einzelnen Sprechergruppen beziehungsweise (angenommenen) Vokaltraktgrößen eingestellt und andererseits während der Analyse diese Einstellungen verändert, wenn ermittelte numerische Werte im Vergleich mit dem jeweiligen Spektrogramm oder dem Spektrum nicht die erwartete Anzahl von Formanten in den erwarteten je vokalspezifischen Frequenzbereichen aufweisen.

So kommt es, dass in bestimmten Fällen für eine einzelne erwachsene Sprecherin die Analyse des einen Teils von Vokallauten mit den für Frauen eingestellten Standardparametern erwartete Werte ergibt, bei der Analyse des anderen Teils von Vokallauten aber die Zahl der Filter (respektive Pole) einer LPC-Analyse erhöht werden muss, um die erwarteten Werte zu erhalten. Die Erhöhung kann dabei zu einer Anzahl von Filtern (Polen) führen, welche zum Standard der Analyse von Vokallauten von Männern gehört. Gleiches gilt für bestimmte Fälle einzelner Kinder, für welche teilweise Parametereinstellungen vorgenommen werden, wie sie den Standards für Frauen entsprechen.

Dies drückt eine Inkonsistenz in der Begründung von Parametereinstellungen aus: Für die Voreinstellungen werden spezifische Analyseparameter mit spezifischen Vokaltraktgrößen der Sprecherinnen oder Sprecher, für Korrekturen dieser Einstellungen aber mit dem jeweiligen Grad der «Auflösung» der Analyse begründet.

## 6.6 Zusatz: Spektrum, Formantmuster und Resynthesen

Wie in Kapitel 6.1 erläutert, besteht kein durchgängiger direkter Zusammenhang von Spektrum beziehungsweise seiner Hüllkurve einerseits und analytisch ermittelten Formantfrequenzen andererseits. In der Folge steht zur Frage, welche allgemeine Beziehung zwischen dem natürlichen Vokallaut, dem ermitteltem Formantmuster und der Resynthese besteht.

Ein Teil von analytisch ermittelten Formantmustern wird denn derzeit auch anhand von Resynthesen auf ihre Aussagekraft hin überprüft. Aber diese Überprüfung vermag die beschriebene allgemeine Problematik der Analysemethoden nicht substantziell zu relativieren. Resynthesen sind nur dann sinnvoll, wenn Formantanalysen nicht grundsätzlich, sondern nur im Hinblick auf eine beschränkte Variation von Analyseparametern zur Frage stehen.

Hinzu kommt, dass die Frage von Resynthesen in einen Zusammenhang mit der in Kapitel 3.1 beschriebenen Variabilität der Anzahl von Formanten in Synthesen, welche sich nicht auf natürliche Einzellaute und deren Analysen beziehen, gestellt werden muss: Wenn ein bestimmtes analytisch ermitteltes Formantmuster, in einer Resynthese angewendet, eine «erwartete» Vokalidentität erkennen lässt, bedeutet dies nicht, dass bei einer Variation der Analyseparameter ein anderes ermitteltes Formantmuster, in einer zweiten Resynthese angewendet, nicht dieselbe Vokalidentität erkennen lassen kann. Umgekehrt ist denkbar, dass für bestimmte Einzellaute Resynthesen anhand «unerwarteter» Formantmuster eine bessere Annäherung an den zu analysierenden natürlichen Vokallaut ergeben als «erwartete» Formantmuster.



Teil 3  
Erfahrungen und Beobachtungen

## 7 Unsystematische Entsprechung von Vokalen mit Mustern spektraler Energiemaxima und mit Formantmustern

### 7.1 **Inkonstanz der Anzahl vokalspezifischer spektraler Energiemaxima und mögliche Inkongruenz von Formantmustern**

Laute von hinteren Vokalen und /a-a/ können, wie in Kapitel 3.1 erläutert, nur ein spektrales Energiemaximum innerhalb des für sie vokalspezifischen Frequenzbereichs  $\leq 1,5$  kHz aufweisen, im Unterschied zu anderen Lauten desselben Vokals mit zwei solchen Maxima. Die Anzahl vokalspezifischer spektraler Energiemaxima ist in der Folge inkonstant.

Nicht in allen Fällen entsprechen dabei spektrale Hüllkurven und Formantmuster solcher Vokallaute einer Interpretation von «Formantverschmelzungen» («formant merging»): Vergleiche zweier Laute jeweils desselben Vokals können im einen Falle nur ein tieferes Energiemaximum, im anderen Falle derer zwei aufweisen, wobei die Frequenzen der ersten Hüllkurvenspitzen sich in beiden Fällen entsprechen. Die zweite Hüllkurvenspitze  $\leq 1,5$  kHz des zweiten Lautes mag dann mit der «Färbung» des Klangs in einem Zusammenhang stehen, erweist sich aber als marginal für die Vokalerkennung selbst. Sie besitzt keinen vokaldifferenzierenden Wert.

Formantanalysen mögen in beiden Fällen zwei tiefere Formanten ergeben, doch erweist sich dabei die Berechnung von F2 für den Vokallaut mit nur einer tieferen Hüllkurvenspitze häufig als direkt abhängig von der Wahl der Anzahl Filter, und seine Amplitude und Bandbreite können sich als ausgesprochen tief beziehungsweise gross erweisen.

Verwandtes gilt für Laute vorderer Vokale: Laute eines vorderen Vokals können einerseits in dem als vokalspezifisch angenommenen höheren Frequenzbereich  $> 2$  kHz nur ein spektrales Energiemaximum aufweisen, im Unterschied zu anderen Lauten desselben Vokals mit zwei solchen Maxima, wobei Formantanalysen in den meisten Fällen auch für die erstgenannten Laute Werte für F2 und F3 ergeben.

Laute eines vorderen Vokals können andererseits Frequenzen des jeweils zweiten spektralen Energiemaximums beziehungsweise von F2 aufweisen, welche stark variieren, bis hin zu Werten, welche nahe bei oder gar über den Frequenzen des jeweils dritten spektralen Energiemaximums beziehungsweise von F3 anderer Laute desselben Vokals liegen (letzteres insbesondere zu beobachten im Vergleich von Vokallauten verschiedener allgemeiner Sprechergruppen auf jeweils sich entsprechender Grundtonhöhe). Damit stellt sich die Frage der Rolle von F2 für die physikalische Repräsentation des jeweiligen Vokals: Im einen Falle mögen nur F1–F2, im anderen Falle aber erst F1–F2–F3 dem Vokal selbst physikalisch entsprechen, was in Filterexperimenten zum Ausdruck kommt. (Auch Fälle, für welche F1–F3 sich als physikalische Entsprechung erweisen, sind in diesem Zusammenhang nicht auszuschliessen.)

Eine einheitliche Angabe der Anzahl sich abfolgender spektraler Energiemaxima, welche jeweils einen Vokal repräsentieren, ist in der Folge nicht möglich. Gleiches gilt, wenn auch nicht in derselben offensichtlichen Weise, für die Formanten. Hinzu kommen Formantmuster, für welche die gegenseitige Entsprechung von vokalspezifischen Formantfrequenzen zur Frage steht.

Die Anzahl vokalspezifischer spektraler Energiemaxima ist inkonstant, und es gibt Fälle inkongruenter Formantmuster.

## 7.2 Teilweise fehlende Ausprägung vokalspezifischer spektraler Energiemaxima

Laute hinterer Vokale und /a–ɑ/ auf Grundtonhöhen  $\leq 350$  Hz können in dem für sie vokalspezifischen Bereich des Spektrums  $\leq 1,5$  kHz Partialtöne mit durchgängig quasi-identischen Amplituden aufweisen. Diese Partialtonspektren erscheinen dann als «flach», ohne deutliche Ausprägung von relativen Energiemaxima im Sinne deutlicher Hüllkurvenspitzen. Von besonderem Interesse sind dabei Fälle von Lauten von /a, ɑ, ɔ, o/, in welchen die ersten drei bis fünf Partialtöne keine deutlichen Unterschiede ihrer Amplituden aufweisen.

Laute der vorderen Vokale /i, e, ε/ auf Grundtonhöhen  $\leq 350$  Hz können in dem für sie vokalspezifischen höheren Anteil des Spektrums  $> 2$  kHz gleichermaßen Partialtöne mit durchgängig quasi-identischen Amplituden aufweisen. Was für hintere Vokale und Varianten von /a-a/ grundsätzlich für den gesamten für sie vokalspezifischen Frequenzbereich gilt, gilt für vordere Vokale parziell, für den höheren für sie vokalspezifischen Frequenzbereich.

Darüber hinaus bestehen Fälle solch «flacher» spektraler Anteile auch für Vokallaute mit Grundtonhöhen  $> 350$  kHz, auch wenn dies in der Regel in Zusammenhang mit dem grossen Frequenzabstand der Partialtöne auf Laute der Vokale /i, e, ε, a, a/ beschränkt bleibt.

In diesem Zusammenhang sind auch Laute der Vokale /a, a, ɔ, o, u/ zu erwähnen, für welche die Amplituden der Partialtöne im Spektrum  $\leq 1,5$  kHz in konstanter Weise leicht oder stark abnehmen. Für alle diese Laute bildet der erste Partialton das tatsächliche spektrale Maximum.

In dieser Weise erweitert sich die Problematik, keine allgemeine Formulierung des Zusammenhangs von erkanntem Vokal und seiner physikalischen Repräsentation anhand einer bestimmten Anzahl einander abfolgender spektraler Hüllkurvenspitzen leisten zu können.

Spektrale Hüllkurvenmaxima, wie sie in der Literatur beschrieben werden, sind keine Bedingung für die physikalische Repräsentation von Vokalen.

Der Zusammenhang von fehlenden Ausprägungen vokalspezifischer Hüllkurvenspitzen im Spektrum und numerisch ermittelten Werten für Formantfrequenzen ist nicht in einfacher und allgemeiner Weise zu beschreiben. Entsprechende Einzelheiten werden hier deshalb offen gelassen. Die Erscheinung als solche erweitert aber auch die bis anhin aufgeführten methodischen Probleme der Formantanalyse.

### **7.3 Zusatz: Nachvollzug anhand von Resynthesen und Synthesen**

Inkonstanz der Anzahl vokalspezifischer spektraler Energiemaxima, mögliche Inkongruenz von Formantmustern und Vokallaute mit «flachen» oder konstant abfallenden spektralen Anteilen sind, für Grundtonhöhen  $\leq 350$  Hz, in Resynthesen nachvollziehbar.

Gleiches gilt für Resynthesen der ersten und dritten der genannten Erscheinungen auf Grundtonhöhen  $> 350$  Hz, welche sich direkt auf Partialtonspektren natürlicher Vokallaute beziehen.

Gleiches gilt auch für Synthesen über Formantmuster oder Partialtonspektren, welche nicht in direkter Weise von natürlichen Vokallaute abgeleitet werden.

## 8 Fehlende Entsprechung von Vokalen mit Mustern spektraler Energiemaxima oder mit Formantmustern

### 8.1 Abhängigkeit der vokalspezifischen spektralen Energiemaxima und der tieferen Formanten $\leq 1,5$ kHz von der Grundtonhöhe

Die erste spektrale Hüllkurvenspitze — soweit vorhanden — und der erste Formant von Vokallauten auf Grundtonhöhen  $\leq 350$  Hz erweisen sich, in systematischer Weise empirisch untersucht, als abhängig von der Grundtonhöhe selbst: Ihre Frequenzen steigen mit zunehmender Grundtonhöhe. Ausnahme bilden bestimmte Fälle von Lauten der Vokale /a-a/.

Diese Abhängigkeit fehlt zwar oder ist marginal innerhalb nur eines Bereichs der Änderung der Grundtonhöhe  $\leq 200$  Hz, sie erweist sich hingegen als sehr ausgeprägt innerhalb eines Bereichs der Änderung der Grundtonhöhe zwischen 250 Hz und 350 Hz.

Vokallaute mit höheren Grundtonhöhen  $> 350$  Hz bestätigen anhand von Vergleichen von Partialtonspektren ihrerseits eine Abhängigkeit spektraler Energiemaxima von der Grundtonhöhe.

Für die hinteren Vokale auf Grundtonhöhen  $\leq 350$  Hz finden sich zudem auch Anzeichen einer zumindest teilweisen — für einzelne Formantmuster spezifischen — Abhängigkeit des zweiten Energiemaximums und des zweiten Formanten von der Grundtonhöhe.

Offen gelassen wird hier die Beziehung von Grundtonhöhen und vokalspezifischen Hüllkurvenspitzen beziehungsweise vokalspezifischen Formanten  $> 1,5$  kHz bei Lauten vorderer Vokale.

Diese Beobachtungen gelten für Vokallaute jeweils ein und derselben Sprecherin beziehungsweise ein und desselben Sprechers. Sie gelten also grundsätzlich für Vokale und ihre physikalische Repräsentation.

Von besonderer Bedeutung ist in diesem Zusammenhang die Beobachtung, dass die Abhängigkeit der tieferen Hüllkurvenspitzen und der tieferen Formanten  $\leq 1,5$  kHz keine Erscheinung des «Übersingens» des ersten Formanten darstellt: Die zu beobachtenden Formantverschiebungen für Laute der Vokale /e, ø, o/ sind grundsätzlich bereits auf Grundtonhöhen unterhalb der in der Literatur für «normales» Sprechen aufgeführten statistischen Werte für F1 zu beobachten, und sie erweisen sich im Vergleich mit Lauten der Vokale /i, y, u/ in vielen Fällen als stärker ausgeprägt, obwohl für letztere in der Literatur deutlich tiefere statistische Werte für F1 angegeben werden als für erstere.

Von besonderer Bedeutung ist auch die Beobachtung, dass bei ausgeprägter Variation der Grundtonhöhe die auftretenden Unterschiede von F1 für Laute jeweils eines Vokals in vielen Fällen die in der Literatur statistisch für «normales» Sprechen aufgeführten Unterschiede von F1 für Laute zweier benachbarter Vokale deutlich übertreffen.

Vokalspezifische spektrale Hüllkurvenspitzen und vokalspezifische Formantmuster  $\leq 1,5$  kHz sind in der Regel abhängig von der Grundtonhöhe.

## **8.2 Erkennbare Vokallaute auf Grundtonhöhen oberhalb statistischer Werte der ersten Formantfrequenz**

Laute der Vokale /i, e, a, o, u/ können von ausgewählten Sprecherinnen und Sprechern mit grossem Stimmumfang und guter Artikulation beziehungsweise sehr deutlicher Lautbildung bis auf Grundtonhöhen von 600 Hz deutlich erkennbar und voneinander unterscheidbar gebildet werden, und die jeweiligen Spektren zeigen vokalspezifische Unterschiede, auch wenn diese im Einzelnen anders als über Hüllkurven und Formanten beschrieben werden müssen. Einfach zu erfahren, weil im Alltag häufig zu hören, sind solche Laute bei Kindern und Frauen; sie können aber auch für Männer nachgewiesen werden. Eine Grundtonhöhe von 600 Hz liegt oberhalb aller statistischer Werte für F1 aller deutschen Vokale bei «normalem» Sprechen, für Frauen wie für Männer, mit Ausnahme von /a/.

Vokallaute auf Grundtonhöhen oberhalb statistischer Werte der ersten Formantfrequenz bei «normalem» Sprechen sind erkennbar.

Die Möglichkeit einer solchen Vokalerkennung und -unterscheidung widerspricht der Bezeichnung von vorliegenden, statistisch ermittelten Formantmustern als «vokalspezifischen» Mustern, ganz unabhängig von methodischen Problemen der Bestimmung von Hüllkurvenspitzen und Formantfrequenzen. Zugleich bestätigt der Befund seinerseits die Abhängigkeit von tieferen spektralen Energiemaxima und von tieferen Formanten von der Grundtonhöhe.

Die Feststellung, dass Laute der Vokale /i, e, a, o, u/ bis auf eine Grundtonhöhe von 600 Hz erkannt werden können, wird hier «konservativ» getroffen. Die Frage, ob bis auf diese Grundtonhöhe grundsätzlich alle Vokale erkennbar gebildet werden können, wie auch die Frage erkennbarer und voneinander unterscheidbarer Vokale auf deutlich darüber liegenden Grundtonhöhen, werden hier offen gelassen.

Hohe Laute hinterer Vokale auf Grundtonhöhen  $> 350$  Hz können nur über die ersten zwei oder drei Partialtöne und deren Amplituden physikalisch repräsentiert werden. Dies verdeutlicht die grundsätzliche Problematik der Annahme, spektrale Energiemaxima beziehungsweise Hüllkurvenspitzen in enger begrenzten Frequenzbereichen seien in durchgängiger Weise das physikalische Merkmal des Vokals.

Entsprechendes gilt, wenn auch in Hinsicht auf die physikalische Repräsentation nur parziell, für den vokalspezifischen tieferen Frequenzanteil vorderer Vokale.

### **8.3 «Inversionen» spektraler Maxima und Minima und «inverse» Formantmuster bei Lauten eines einzelnen Vokals**

Weil sich spektrale Energiemaxima beziehungsweise Hüllkurvenverläufe  $\leq 1,5$  kHz, soweit sie bestimmbar sind, als von der Grundtonhöhe

abhängig erweisen, können bestimmte Vergleiche zweier Einzellaute desselben hinteren Vokals, aber auf verschiedenen Grundtonhöhen erzeugt, «inverse» spektrale Maxima und Minima als «inverse» Hüllkurvenverläufe  $\leq 1,5$  kHz aufweisen. Entsprechendes gilt für den Vergleich der jeweiligen Formantmuster: Im Frequenzbereich relativ minimaler spektraler Energie beziehungsweise im Frequenzbereich zwischen den Formanten des einen Lauts kann ein relatives spektrales Maximum beziehungsweise eine Formantfrequenz des anderen Lauts vorgefunden werden und vice versa.

Für vordere Vokale sind solche «Inversionen» für den höheren vokalspezifischen Frequenzbereich  $> 2$  kHz ebenfalls zu beobachten, auch wenn die Frage eines Zusammenhangs solcher «Inversionen» mit der Variation der Grundtonhöhe hier offen gelassen wird.

Diese Beobachtung bestätigt erneut die fehlende allgemeine Entsprechung von Vokalen mit für sie spezifischen Verläufen spektraler Hüllkurven und entsprechenden Formantmustern.

Spektrale Hüllkurvenverläufe zweier Laute desselben Vokals, auf verschiedenen Grundtonhöhen erzeugt, können sich in den für sie vokalspezifischen Frequenzbereichen «invers» zueinander verhalten. Gleiches gilt für Formantmuster.

#### **8.4 Zusatz: Flüstervokale, Grundtonabhängigkeit vokalspezifischer spektraler Merkmale, «Inversionen»**

Wie in Kapitel 5.5 erläutert, verweisen empirische Studien auf erhöhte vokalspezifische Formantfrequenzen F1 und F2 bei geflüsterten Lauten im Vergleich zu stimmhaften Lauten. Wie ebenfalls erwähnt, bezieht sich eine solche Beobachtung aber nur auf eine tiefere Grundtonhöhe der stimmhaften Laute bei entspanntem Sprechen in einem geschlossenen Raum.

Gesetzt, ein geflüstertes Laut weist einen höheren ersten und zweiten Formanten als ein stimmhafter Laut desselben Vokals auf, und gesetzt, die Grundtonhöhe von letzterem wird bei seiner Bildung schritt-

weise erhöht, so wird sich in vielen Fällen eine bestimmte Grundtonhöhe festlegen lassen, für welche sich F1 und F2 von geflüstertem und stimmhaftem Laut entsprechen.

Ob dies darüber hinaus eine eigentliche Regel darstellt, wird hier offen gelassen.

Wird die Grundtonhöhe des stimmhaften Lauts nachfolgend weiter erhöht, so werden sich Fälle finden lassen, für welche F1 oder F1-F2 des geflüsterten Lauts tiefer liegen als F1 oder F1-F2 des stimmhaften Lauts.

In jedem Fall trifft die allgemeine Aussage nicht zu, geflüsterte Laute würden grundsätzlich höhere vokalspezifische Formantmuster als stimmhafte Laute aufweisen.

Im Laufe einer derartigen Untersuchung werden sich zudem Fälle von Vergleichen eines geflüsterten und eines stimmhaften Lauts mit den beschriebenen «Inversionen» finden lassen.

### **8.5 Zusatz: Nachvollzug anhand von Resynthesen und Synthesen**

Alle erläuterten Aspekte der fehlenden Entsprechung von Vokalen mit spektralen Energiemaxima beziehungsweise Hüllkurvenspitzen oder mit Formantmustern sind, ausgehend von Analysen natürlicher Vokallaute, deren Bestimmung bestehenden Standards entspricht, in Resynthesen nachvollziehbar.

Gleiches gilt für Resynthesen auf Grundtonhöhen  $> 350$  Hz, welche sich direkt auf Partialtonspektren natürlicher Vokallaute beziehen.

Gleiches gilt auch für Synthesen über Formantmuster oder Partialtonspektren, welche nicht in direkter Weise von natürlichen Vokallaute abgeleitet werden.

## 9 «Ambige» Entsprechung von Vokalen mit Mustern spektraler Energiemaxima oder mit Formantmustern oder mit vollständigen spektralen Hüllkurven

### 9.1 «Ambige» Muster spektraler Energiemaxima und «ambige» Formantmuster

All dies läuft auf die Beobachtung und Feststellung hinaus, dass zwei Einzellaute auf zwei verschiedenen Grundtonhöhen, welche als zwei verschiedene Vokale erkannt werden, innerhalb des für sie vokalspezifischen Frequenzbereichs quasi-identische spektrale Energiemaxima im Sinne von spektralen Hüllkurvenspitzen und quasi-identische Formantmuster aufweisen können. Viele Muster von spektralen Energiemaxima beziehungsweise viele Formantmuster erweisen sich in der Folge als «ambig», als mehrdeutig, sie repräsentieren für sich genommen physikalisch zwei oder mehrere verschiedene Vokale.

Muster spektraler Energiemaxima erweisen sich empirisch in vielen Fällen nicht als vokalspezifisch, sondern im Gegenteil als «ambig». Gleiches gilt für Formantmuster.

Besonders deutlich wird diese Feststellung, wenn Vokallaute auf allen Grundtonhöhen ihrer Erkennbarkeit miteinander verglichen werden, und Entsprechungen spektraler Energiemaxima und -minima auch abseits von Bestimmungen spektraler Hüllkurven und Formantfrequenzen im direkten Vergleich der Partialtonspektren geprüft werden.

### 9.2 «Ambige» spektrale Hüllkurven

Diese «Ambiguität» betrifft in bestimmten Fällen über die spektralen Energiemaxima beziehungsweise Hüllkurvenspitzen hinaus auch den gesamten Verlauf der spektralen Hüllkurve.

Spektrale Hüllkurvenverläufe können gleichermaßen «ambig» sein.

### 9.3 «Ambiguität» und Einzelvokale

Für alle hier erörterten deutschen Vokale können Fälle von Lauten mit «ambigen» spektralen Energiemaxima oder mit «ambigen» Formantmustern in den vokalspezifischen Frequenzbereichen auftreten.

Inwieweit dies auch für die vollständigen Hüllkurvenverläufe gilt, wird hier offen gelassen.

Werden Vokallaute auf allen Grundtonhöhen ihrer Erkennbarkeit miteinander verglichen und werden trotz methodischer Probleme der Bestimmung von spektralen Hüllkurven und von Formantmustern Beurteilungen anhand direkter Vergleiche der Partialtonspektren vorgenommen, so lässt sich die beschriebene «Ambiguität» nicht nur für benachbarte Vokalpaare, sondern auch für mehr als zwei Vokale beobachten. Dies gilt insbesondere dann, wenn Laute der drei allgemeinen Sprechergruppen miteinander verglichen werden.

Die beschriebene «Ambiguität» ist weder auf einzelne Vokale noch auf benachbarte Vokalpaare beschränkt, und sie kann mehr als zwei Vokale gleichzeitig betreffen.

Die Frage, ob für bestimmte Vokale — insbesondere für /a-a/ — auch Fälle von Lauten mit in strengem Sinne vokalspezifischen spektralen Energiemaxima beziehungsweise Hüllkurvenspitzen und mit in strengem Sinne vokalspezifischen Formantmustern bestehen, welche bei keinem Laut eines anderen Vokals vorzufinden sind, wird hier offen gelassen.

#### **9.4 Zusatz: «Ambiguität» spektraler Charakteristika von Flüsterlauten**

Wie in Kapitel 5.5 erläutert und in Kapitel 8.4 kritisch kommentiert, verweisen empirische Studien zwar auf erhöhte vokalspezifische Formantfrequenzen F1 und F2 bei geflüsterten Lauten im Vergleich zu stimmhaften Lauten, aber solche Unterschiede können sich aufheben oder sich gar in ihr Gegenteil verkehren, wenn die Grundtonhöhe der jeweils in einen Vergleich mit einbezogenen stimmhaften Laute substantiell variiert wird.

Diese Feststellung weiter ausgelegt, bedeutet, dass höhere F1 oder F1-F2 für geflüsterte Laute im Vergleich zu stimmhaften Lauten — bei tieferen Grundtonhöhen von letzteren — in erster Linie nicht als Merkmal des Unterschieds der Phonationsart, sondern im Rahmen der grundsätzlichen «Ambiguität» von Mustern spektraler Energiemaxima beziehungsweise Hüllkurvenspitzen und von Formantmuster zu diskutieren sind.

#### **9.5 Zusatz: Nachvollzug anhand von Resynthesen und Synthesen**

Für diejenigen Fälle «ambiger» Formantmuster von natürlichen Vokallauten, deren Bestimmung bestehenden Standards entspricht, bestätigen Formantsynthesen in der Regel die erläuterte «Ambiguität»: Werden ausgehend von einem einzelnen Formantmuster, welches sich bei natürlichen Vokallauten als «ambig» erweist, verschiedene Klänge synthetisch erzeugt, indem die Grundtonhöhe den Beobachtungen natürlicher Vokale entsprechend deutlich variiert wird, ändert der erkannte Vokal des synthetischen Klanges.

Gleiches gilt für Resynthesen auf Grundtonhöhen  $> 350$  Hz, welche sich direkt auf Partialtonspektren natürlicher Vokallaute beziehen. Gleiches gilt auch für Synthesen über Formantmuster oder Partialtonspektren, welche nicht in direkter Weise von natürlichen Vokallauten abgeleitet werden.

Auch die zusätzlichen Feststellungen zum Vergleich von stimmhaften mit geflüsterten Lauten lassen sich in der Resynthese bestätigen.

## 10 Fehlende Entsprechung von Mustern spektraler Maxima oder von Formantmustern mit allgemeinen Sprechergruppen beziehungsweise Vokaltraktgrössen

### 10.1 Gleiche Muster spektraler Maxima und gleiche Formantmuster $\leq 1,5$ kHz für verschiedene allgemeine Sprechergruppen wie für verschiedene Vokaltraktgrössen

Gesetzt, Laute eines Vokals werden von Sprecherinnen und Sprechern verschiedener allgemeiner Sprechergruppen auf derselben Grundtonhöhe gebildet, und die Laute entsprechen einander nicht nur in der allgemeinen Zuordnung zum jeweiligen Vokal, sondern auch in der jeweiligen Variante der «Vokalfärbung» — die wahrnehmungsmässige Entsprechung ist dann höchstmöglich —, so erweisen sich sowohl spektrale Energiemaxima beziehungsweise Hüllkurvenspitzen wie Formantmuster für den tieferen Frequenzbereich  $\leq 1,5$  kHz empirisch als quasi-identisch, abgesehen von möglichen Unterschieden, welche aufgrund verschiedener Parametereinstellungen bei der Formantanalyse auftreten können. Der tiefere Anteil des Spektrums erweist sich derart als unabhängig von der allgemeinen Sprechergruppe und, so folgern wir, auch von der Vokaltraktgrösse. Dieser Frequenzbereich wird in der Literatur als in vollständigem Sinne vokalspezifisch für alle hinteren Vokale, /a/ mit eingeschlossen, bezeichnet. Gleiches gilt parziell, im Hinblick auf F1 aller anderen Vokale.

Dies bedeutet insbesondere für hintere Vokale, dass Einzellaute von Männern mit höheren Grundtonhöhen als Einzellaute von Frauen in vielen Fällen auch höhere spektrale Energiemaxima beziehungsweise Hüllkurven sowie höhere Formantmuster F1-F2 aufweisen, was einer Umkehrung des in der Literatur aufgeführten Zusammenhangs von vokalspezifischen Formantmustern und allgemeinen Sprechergruppen beziehungsweise Vokaltraktgrössen entspricht. Gleiches gilt für F1 von Lauten vorderer Vokallaute. Gleiches gilt auch für entsprechende Vergleiche von Vokallaute von Erwachsenen mit denjenigen von Kindern.

Aussagen zu /a-a/ werden hier nicht gemacht, da die Systematik des Zusammenhangs von Grundtonhöhe und Mustern spektraler Energiemaxima sowie Formanmustern hier nicht formuliert werden kann (siehe Kapitel 8.1).

In jedem Fall trifft die allgemeine Aussage, Laute von Kinder würden grundsätzlich die höchsten, Laute von Frauen mittlere, und Laute von Männern die tiefsten vokalspezifischen Formantfrequenzen aufweisen, nicht zu.

Muster spektraler Energiemaxima beziehungsweise Hüllkurvenspitzen wie auch Formantmuster erweisen sich für einen Frequenzbereich  $\leq 1,5$  kHz empirisch als quasi-identisch für die verschiedenen allgemeinen Sprechergruppen beziehungsweise für verschiedene Vokaltraktgrößen. Auftretende Unterschiede beziehen sich, strenge Lautentsprechungen vorausgesetzt, in direkter Weise auf Unterschiede von Grundtonhöhen.

Für höhere Frequenzanteile  $> 1,5$  kHz steht wie erwähnt noch zur Frage, ob diese grundsätzlich oder nur für bestimmte Fälle von Lauten unabhängig von der Grundtonhöhe sind. Damit zusammenhängend steht deshalb auch zur Frage, ob höhere Frequenzanteile grundsätzlich spezifisch für Vokaltraktgrößen sind.

## 10.2 «Dichotomie» des Vokalspektrums

Wie bereits mehrfach erwähnt, ist die Abhängigkeit der vokalspezifischen spektralen Charakteristik und der Formanten von der Grundtonhöhe für den tieferen Frequenzbereich  $\leq 1,5$  kHz empirisch leicht nachzuvollziehen, nicht aber für den höheren Frequenzbereich  $> 1,5$  kHz. Zugleich sind die tieferen spektralen Anteile und die tieferen Formantfrequenzen unspezifisch in Hinsicht auf Sprechergruppen und Vokaltraktgrößen, was für die höheren spektralen Anteile und die höheren Formantfrequenzen zur Frage steht. Das Spektrum eines Vokallauts ist somit nicht einheitlich, sondern zweigeteilt zu erörtern.

Das Spektrum eines Vokals erweist sich als «dichotom».

Von besonderer Bedeutung ist in diesem Kontext die Möglichkeit, dass sich spektrale Energiemaxima beziehungsweise Hüllkurvenspitzen sowie höhere Formanten insbesondere  $> 2$  kHz bei Lauten vorderer Vokale in gleichzeitigem Zusammenhang mit der Vokalidentität und der wahrgenommenen Sprechergruppe stehen können: Systematische Unterschiede in diesem höheren Frequenzbereich können für Laute jeweils eines Vokals von Kindern, Frauen und Männern auch dann beobachtet werden, wenn diese die Laute auf jeweils gleichen Grundtonhöhen bilden.

Allerdings wird hier offen gelassen, ob dies auch dann der Fall ist, wenn Männer sogenannte «Frauenstimmen» beziehungsweise Erwachsene sogenannte «Kinderstimmen» imitieren.

### **10.3 Zusatz: Geflüsterte Vokallaute im Vergleich verschiedener allgemeiner Sprechergruppen beziehungsweise verschiedener Vokaltraktgrößen**

Vergleichende Untersuchungen der Formantmuster von geflüsterten Vokallauten von Kindern, Frauen und Männern, deren Befunde denselben Status beanspruchen können, wie dies für die in Teil 2 aufgeführten Referenzstatistiken stimmhafter Vokallaute gilt, bestehen bis anhin keine. Aber diejenigen Studien (in der Anzahl sehr beschränkt und meist nicht alle Vokale systematisch mit einbeziehend), welche geflüsterte Laute verschiedener allgemeiner Sprechergruppen miteinander vergleichen, verweisen auf Unterschiede in den Formantmustern, welche trotz aller bisheriger Erläuterungen als Indiz für einen allgemeinen Zusammenhang von Mustern spektraler Maxima und von Formantmustern mit der allgemeinen Sprechergruppe beziehungsweise der Vokaltraktgröße verstanden werden können, die tieferen Frequenzbereiche mit betreffend.

Die Klärung dieses Aspekts sowie seine Einbettung in den bisher erläuterten Zusammenhang von Vokalen und ihrer spektralen Charakteristik wird hier offen gelassen.

#### 10.4 Zusatz: Vokalimitationen durch Vögel

Im Zusammenhang mit der Erörterung von Vokallauten, Spektrum, Formantmustern, wahrgenommenen Sprechergruppen und Vokaltraktgrößen sind auch Äusserungen von Tieren von Bedeutung, welche menschliche Laute imitieren.

Zunächst stellt sich grundsätzlich die Frage, wie es dazu kommt, dass Vögel imstande sind, Vokale, wie sie Menschen erzeugen, zu imitieren, obwohl sie über keine dem Menschen entsprechende Phonation und Artikulation und insbesondere über keine entsprechenden Resonanzräume eines Vokaltrakts verfügen.

Eine eingehende Untersuchung solcher Imitationen, bezogen auf Vögel, welche sich darin besonders auszeichnen, lässt Folgendes feststellen: Imitieren Vögel Wörter, werden daraus imitierte Einzellaute als Klangfragmente isoliert, und besitzen diese Klangfragmente quasi-statischen Charakter im Sinne von quasi-statischen spektralen Charakteristika, so können solche Klänge als isolierte Vokallaute erkennbar sein. Für einen Teil solcher Lautfragmente lassen sich vollständige Formantmuster F1-F2-F3 interpretieren, wie sie für Laute von Menschen in Formantstatistiken aufgeführt sind. Für einen anderen Teil sind die Entsprechungen nur parziell.

Die Tatsache, dass Vögel imstande sind, menschliche Vokale zu imitieren, welche mit letzteren vergleichbare vokalspezifische spektrale Merkmale und Formantmuster aufweisen, widerspricht seinerseits einer strengen Entsprechung von spektralen Charakteristika des produzierten Klanges mit der Vokaltraktgröße. Vokale müssen in der Folge so erörtert werden, dass die Möglichkeit, unter verschiedenen physischen und physiologischen Bedingungen dieselben Klangcharakteristika erzeugen zu können, mit ins Zentrum der Untersuchung und Reflexion gestellt wird.

## **10.5 Zusatz: Nachvollzug anhand von Resynthesen und Synthesen**

Auch die erläuterten Aspekte einer fehlenden Entsprechung von Mustern spektraler Maxima und von Formantmustern mit allgemeinen Sprechergruppen beziehungsweise mit Vokaltraktgrößen sind anhand von Resynthesen und Synthesen im oben genannten Sinne nachvollziehbar.

## 11 Fehlende Entsprechung von methodischen Begrenzungen der Formantbestimmung mit Begrenzungen der Vokalerkennung

### 11.1 Erkennbarkeit von Vokallauten auf Grundtonhöhen $> 350$ Hz

Erkennbare Vokallaute können, wie in Kapitel 8.2 dargelegt, auf Grundtonhöhen gebildet werden, welche weit über der kritischen Grenze liegen (vgl. dazu Kapitel 6.4), ab welcher Formanten aus methodischen Gründen nicht mehr verlässlich bestimmt werden können.

Vokallaute mit Grundtonhöhen  $> 350$  Hz sind erkennbar, Formantmuster aber aus methodischen Gründen nicht bestimmbar. Die methodische Begrenzung der Bestimmung von Formantmustern für Vokallaute auf Grundtonhöhen  $> 350$  Hz geht also nicht grundsätzlich mit einer Minderung der Vokalerkennung einher.

Die Formulierung einer allgemeinen Theorie der physikalischen Repräsentation von Vokalen anhand von Formantmustern erweist sich in der Folge im Hinblick auf ihre methodische Beschränkung als kritisch.

### 11.2 Fehlende Entsprechung von methodischen Problemen der Bestimmung von Formantmustern auf Grundtonhöhen $\leq 350$ Hz mit einer Abnahme der Erkennbarkeit von Vokallauten

Vokallaute auf Grundtonhöhen  $\leq 350$  Hz, für welche die Bestimmung von Formanten sich aus anderen Gründen als der Grundtonhöhe als kritisch erweist — beispielsweise bei unerwartet «fehlenden» einzelnen spektralen Energiemaxima oder bei «flachen» Spektren —, sind ihrerseits nicht schlechter erkennbar als Vokallaute, für welche die Bestimmung von Formanten als unproblematisch gilt.

Methodische Probleme bei der Bestimmung von Formantmustern für Vokallaute auf Grundtonhöhen  $\leq 350$  Hz gehen ihrerseits nicht grundsätzlich mit einer Minderung der Vokalerkennung einher.

### **11.3 Zusatz: Fehlende methodische Grundlage für die Bestimmung von Formantmustern bei Vokalimitationen durch Vögel**

Imitationen menschlicher Vokallaute durch Vögel sind auf der Grundlage bestehender methodischer Standards in strengem Sinne nicht auf Formantmuster hin untersuchbar. Wie in Kapitel 6.3 erläutert, verlangt die Berechnung von Formanten eine Parametereinstellung für die maximale Anzahl von Filtern der Analyse, welche sich auf die Vokaltraktgröße bezieht. Vögel aber haben keinen Vokaltrakt, welcher demjenigen eines Menschen vergleichbar wäre, und die Frage, wie viele Filter in einer Analyse eines von einem Vogel erzeugten vokalartigen Klanges zur Ermittlung von vokalspezifischen Formanten zu wählen sei, ist nicht zu beantworten.

Für Vergleiche von Äusserungen von Menschen und Vögeln bleibt also nur der direkte Vergleich der jeweiligen Spektren.

Aber auch diese methodische Beschränkung, die genannten Vokalimitationen betreffend, geht nicht mit einer grundsätzlichen Minderung der Erkennung von ihnen einher.

Teil 4  
Falsifikation

## 12 Empirische Falsifikation trotz methodischer Beschränkungen der Bestimmung von spektralen Hüllkurvenmaxima oder von Formantmustern

### 12.1 Fehlende methodische Grundlage für eine Verifikation des theoretischen Grundsatzes

Es ist nicht möglich, innerhalb einer bestimmten Sprache für isolierte und wahrnehmungsmässig erkennbare Vokallaute mit quasistatischer Charakteristik des Spektrums allgemeine und einheitliche Regeln für die Formantbestimmung zu formulieren, gemäss welchen in allen Fällen solcher Einzellaute jeweils eines Vokals einheitliche Formantmuster in je vokalspezifischen Frequenzbereichen zu ermitteln wären.

Es ist in der Folge nicht möglich, allgemeine statistische Werte für Formantfrequenzen erkennbarer Vokallaute, den gesamten Bereich ihrer Erscheinungsweise betreffend, zu erheben.

Der theoretische Grundsatz kann aus methodischen Gründen nicht verifiziert werden.

Bestehende Statistiken lassen diese Problematik nur deshalb nicht erkennen, weil sie einerseits die untersuchten Sprecherinnen und Sprecher in der Regel keiner qualitativen Selektion im Hinblick auf ihre stimmlichen Fähigkeiten unterziehen, und andererseits keine systematische Variation der Grundtonhöhe mit einbeziehen. Beides aber bildet eine notwendige Voraussetzung dafür, den möglichen Bereich von Grundtonhöhen erkennbarer Vokallaute zu untersuchen. Darüber hinaus erlaubt eine qualitative Selektion von Sprecherinnen und Sprechern auch eine Untersuchung weiterer wichtiger Aspekte der Variation der Produktion, wie insbesondere Laustärke, Registerwechsel und Phonationsart.

## 12.2 Systematisches Abweichen empirischer Befunde von Voraussagen des theoretischen Grundsatzes

Tiefere spektrale Energiemaxima beziehungsweise Hüllkurven, tiefere Formanten wie ganz allgemein tiefere Anteile von Partialtonspektren  $\leq 1,5$  kHz erweisen sich in ihrer Beziehung zum je erkannten Vokal in den meisten Fällen als abhängig von der Grundtonhöhe.

Es ist für Sprecherinnen und Sprecher einer Sprachgemeinschaft, aber verschiedener allgemeiner Sprechergruppen oder verschiedener Vokaltraktgrößen grundsätzlich möglich, Vokallaute desselben Vokals mit quasi-identischen Grundtonhöhen und quasi-identischen tieferen Formantfrequenzen  $\leq 1,5$  kHz zu bilden. Darüber hinaus ist es möglich, dass Sprecherinnen und Sprecher mit vergleichsweise grösseren Vokaltrakten einen bestimmten Vokal mit höheren Grundtonhöhen und höheren F1–F2 produzieren können als Sprecherinnen und Sprecher mit vergleichsweise kleineren Vokaltrakten.

Diese empirischen Befunde stehen in gegenseitigem Zusammenhang. Sie weichen in systematischer Weise sowohl von der Voraussage grundtonhöhenunabhängiger vokalspezifischer Formantmuster als auch von der Voraussage durchgängig spezifischer Formantmuster für Vokale verschiedener Sprechergruppen beziehungsweise verschiedener Vokaltraktgrößen ab.

Empirische Befunde weichen in systematischer Weise von Voraussagen des theoretischen Grundsatzes ab.

Der theoretische Grundsatz erweist sich in dieser Weise aus empirischer Sicht als unzutreffend.

### 12.3 Dem theoretischen Grundsatz direkt widersprechende empirische Befunde

Es ist für Sprecherinnen und Sprecher einer Sprachgemeinschaft, unabhängig von ihrer Zugehörigkeit zu einer Sprechergruppe beziehungsweise unabhängig von ihrer Vokaltraktgrösse, grundsätzlich nicht nur möglich, isolierte Einzellaute verschiedener Vokale zu bilden, welche dieselben Formantmuster F1-F2 oder F1-F2-F3 aufweisen, sondern diese «Ambiguität» von Formantmustern stellt ihrerseits eine systematische Erscheinung von Vokallauten dar: Für Fälle solcher «Ambiguitäten» können die Sprecherinnen und Sprecher nicht gleichzeitig das Formantmuster beibehalten, die Grundtonhöhe aber ausgeprägt variieren, sollen alle dabei produzierten Einzellaute denselben Vokal erkennen lassen.

Diese Feststellung gilt gleichermassen für Muster spektraler Energiemaxima. Darüber hinaus erweisen sich in bestimmten Fällen, wie bereits erwähnt, auch die gesamten interpretierbaren spektralen Hüllkurvenverläufe als «ambig».

Empirische Befunde können in direktem Widerspruch zu Voraussagen des theoretischen Grundsatzes stehen.

Der theoretische Grundsatz ist in der Folge falsifiziert, weil für einen substanziellen Teil von Vokallauten das Gegenteil von dem zutrifft, was er behauptet: Einzellaute mit sehr verschiedenen Formantmustern müssen in bestimmten Fällen denselben Vokal und Einzellaute mit demselben Formantmuster müssen in bestimmten Fällen verschiedene Vokale erkennen lassen.

Teil 5  
Kommentar

## 13 Präliminarien

### 13.1 **Hindernisse einer Anpassung oder Ausdifferenzierung des theoretischen Grundsatzes**

Mehrere Gründe sprechen gegen eine Anpassung oder Ausdifferenzierung des bestehenden theoretischen Grundsatzes und der ihm entsprechenden Methoden der akustischen Analyse, als Antwort auf die dargestellten Schwierigkeiten der Plausibilität und auf die erläuterten empirischen Befunde.

Gemäss dem theoretischen Grundsatz werden Formantmuster von Resonanzmustern des Vokaltrakts abgeleitet. Die Formulierung eines substanziellen Zusammenhangs dieser Resonanzen mit der Grundtonhöhe bei der Bildung von Vokallauten würde dem zweiteiligen Modell von Phonation und Artikulation, als Produktion eines allgemeinen Quellenklangs und dessen Umformung durch die Resonanzen des Vokaltrakts, direkt widersprechen. Die Grundtonhöhe ist ein Merkmal der Quelle, die Resonanzen sind das Merkmal des Vokaltrakts, und letztere sind unabhängig von den Klängen oder Geräuschen, auf welche sie wirken. Dies stellt ein grundsätzliches konzeptionelles Hindernis für eine Anpassung oder Ausdifferenzierung des bestehenden theoretischen Ansatzes dar.

Methoden der Formantanalyse selbst verfahren in einer Weise, nach welcher die Grundtonhöhe als Charakteristik der Quelle bei der Berechnung von Filtern (respektive Polen) nicht mit berücksichtigt wird. Dies ist innerhalb der bestehenden Verfahren kaum zu ändern.

Wie Formanten im Sinne von spektralen Energiemaxima für alle erkennbaren Laute zu bestimmen wären, alle in Teil 3 erläuterten Aspekte mit einbeziehend, ist zudem schwer vorstellbar. Dies stellt seinerseits ein Hindernis für eine Anpassung oder Ausdifferenzierung des bestehenden methodischen Ansatzes dar.

Die tatsächlich festzustellende Systematik des Verhaltens der vokalspezifischen spektralen Charakteristika und der entsprechenden Formantmuster kann nicht in einfachem Sinne ausformuliert werden, beispielsweise als allgemeine Regel des Zusammenhangs von For-

mantmustern und Grundtonhöhen. Empirisch erweisen sich die Verschiebungen von Hüllkurvenspitzen und Formanten für Laute jeweils eines Vokals als unsystematisch: Sie sind erst ab einer Grundtonhöhe von 200 Hz deutlich zu beobachten, sie betreffen die tieferen Frequenzbereiche in ganz anderem Ausmass als die hohen Frequenzbereiche und damit die vokalspezifischen Frequenzbereiche hinterer Vokale insgesamt, die vokalspezifischen Frequenzbereiche vorderer Vokale aber nur parziell, und sie sind in ihrem Ausmass sowohl abhängig vom jeweiligen Vokal als auch von der Frequenzlage der Hüllkurvenspitzen oder Formanten. Das Fehlen einer Systematik der empirischen Befunde bietet somit seinerseits keine robuste Grundlage für eine Ausdifferenzierung der Beschreibung vokalspezifischer spektraler Charakteristika im Rahmen des bestehenden Ansatzes.

Diese Überlegungen begründen unsere Skepsis gegenüber einem Versuch der Anpassung oder Ausdifferenzierung des bestehenden theoretischen Grundsatzes und der mit ihm verbundenen Methoden der akustischen Analyse vokalspezifischer Merkmale.

### **13.2 Der theoretische Grundsatz als Index**

Wird ein Laut isoliert gebildet, wird der entsprechende Klang nachfolgend untersucht, weist letzterer quasi-periodischen Charakter auf und wird er als ein bestimmter Vokal einer bestimmten Sprache wahrgenommen und identifiziert, so ist das ihm entsprechende Partialtonspektrum, über die gesamte Dauer des jeweiligen Klangs berechnet, vokalspezifisch: In Hinsicht auf einen Frequenzbereich, welcher die physikalische Repräsentation aller Vokale der jeweiligen Sprache betrifft, findet sich kein gleiches Partialtonspektrum bei Lauten eines anderen Vokals dieser Sprache. Dies wird hier als These formuliert.

Gleiches gilt für Laute, welche aus einem Lautkontext isoliert und nachfolgend als entsprechende Klangfragmente untersucht werden.

Partialtonspektren, so lässt sich behaupten, sind vokalspezifisch und als Abbildung orthogonal: Sie zeichnen sich durch eine systematische Formung aus, nach der die Amplituden der Partialtöne in vokalspezifischer gegenseitiger Relation stehen, von welchen ausgehend sich der jeweilige Vokal wieder reproduzieren lässt.

Die in den vorangehenden Teilen erläuterten Einzelprobleme lassen sich in ihren grundsätzlichen Aspekten deshalb weder auf dynamische Abläufe innerhalb des Klangs, noch auf Eigenheiten des syntaktischen und semantischen Kontextes, noch auf spezielle Prozesse der Wahrnehmung zurückführen, und sie sind auch nicht entsprechend zu relativieren. Sie stellen im Gegenteil einen erst noch zu erklärenden Komplex von Einzelproblemen dar und mit ihm ein erst noch zu erklärendes Phänomen der physikalischen Repräsentation des Vokals selbst.

Für eine gegebene Grundtonhöhe sind derart — abgesehen von möglichen Bestimmungsproblemen — auch die spektrale Hüllkurve, und mit ihr allfällige Hüllkurvenspitzen, vokalspezifisch. Gleiches gilt für berechnete Formantmuster, unter der Voraussetzung, dass die Parametereinstellungen bei entsprechenden Prüfungen jeweils identisch sind, und dass auch Formantamplituden und -bandbreiten in die Beurteilungen mit einbezogen werden.

Für jeweils eine bestimmte Grundtonhöhe und innerhalb methodisch unkritischer Berechnungen von Formantmustern treffen die Voraussetzungen des theoretischen Grundsatzes zu. Dieser verweist damit auf eine grundsätzliche — wenn auch im Einzelnen erst noch zu formulierende — Charakteristik der physikalischen Repräsentation des Vokals. Er erweist sich damit als ein Index dieser Charakteristik.

### **13.3 Einschub: Voraussetzung einer einheitlichen Erkennung von Vokalen**

Ein einzelner isolierter Laut kann von zwei verschiedenen Hörerinnen oder Hörern derselben Einzelsprache in bestimmten Fällen als zwei verschiedene Vokale erkannt werden. Dieser mögliche wahrnehmungsmässige Unterschied ist zwar für die These vokalspezifischer Partialtonspektren mit zu berücksichtigen, er widerspricht ihr aber nicht.

Schwerhörige, um ein erstes Beispiel zu geben, können Vokale anders erkennen als Normalhörende. Dies liegt insbesondere am Verlust der Fähigkeit, höhere Frequenzanteile mit minderer Energie ausreichend wahrnehmen zu können.

In sehr beschränkter Masse lassen sich solche Unterschiede, um ein zweites Beispiel zu geben, auch im Vergleich ausschliesslich Normalhörender beschreiben, was darauf verweist, dass individuelle Eigenarten der Gewichtung von tieferen und höheren Frequenzanteilen in der Wahrnehmung von Vokalen bestehen.

Ob im Bereich der Pathologie, höhere kortikale Prozesse betreffend, weitere Fälle inkonstanter Vokalererkennung beschrieben werden können oder bereits beschrieben sind, wird hier offen gelassen beziehungsweise entzieht sich unserer Kenntnis.

Die These vokalspezifischer Partialtonspektren gilt derart nur unter der Voraussetzung, dass die jeweils zur Frage stehenden Vokallaute in systematischer und einheitlicher Weise erkannt werden. Dies berücksichtigend, lässt sich die entsprechende Formulierung des vorangehenden Kapitels wie folgt differenzieren: Wird ein Laut isoliert gebildet, wird der entsprechende Klang nachfolgend untersucht, weist letzterer quasi-periodischen Charakter auf und wird er als ein bestimmter Vokal einer bestimmten Sprache von einer Gruppe von Hörenden in einheitlicher und wiederholbarer Weise identifiziert, so ist das ihm entsprechende Partialtonspektrum vokalspezifisch.

### **13.4 Vorfeld**

Die empirische Falsifikation des bestehenden theoretischen Grundsatzes, verbunden mit der These, dass der Grundsatz dennoch indexikalischen Charakter besitzt, lässt folgern, dass wir derzeit noch über keine Theorie der akustischen Repräsentation des Vokals verfügen, aber über robuste empirische Hinweise zu einer möglichen solchen Theorie. Wir befinden uns in ihrem Vorfeld.

### **13.5 Zwei Ansätze**

Der bestehende theoretische Grundsatz zeichnet sich dadurch aus, dass er Vokallaute im Rahmen eines für die Sprache unspezifischen physikalischen Modells beschreibt. Klänge und Geräusche aller Art werden durch Filter immer in derselben Weise umgeformt, unabhängig davon, ob sie sprachliche Ereignisse betreffen.

Ein Versuch, auf die sich ergebenden gedanklichen Schwierigkeiten des Nachvollzugs des bestehenden Grundsatzes und auf die empirischen Widersprüche zu seinen Voraussagen zu reagieren, kann darin bestehen, das bestehende Quelle-Filter-Modell zu ergänzen oder durch ein anderes, gleichermassen physikalisches und also der Sprache externes Modell zu ersetzen.

Ein anderer Versuch kann darin bestehen, von einer der Sprache spezifischen Geräusch- und Klangformung auszugehen und dafür in einem ersten Schritt eine Methode der formalen Beschreibung zu entwickeln. Dieser Versuch geht von der Annahme aus, dass die Erscheinungsweise des Vokallauts sich nicht im Rahmen eines rein physikalischen Modells beschreiben lässt.

Ob damit alle grundsätzlich möglichen Ansätze genannt sind, wird hier offen gelassen.

Wie in Kapitel 13.1 erläutert, bestehen Gründe für eine Skepsis gegenüber der Möglichkeit einer Anpassung und Ausdifferenzierung des bestehenden theoretischen Grundsatzes und der mit ihm verbundenen Methoden der akustischen Analyse. Zu den genannten Argumenten kommt folgende Überlegung: Es wäre grundsätzlich möglich, dass Menschen einen allgemeinen Klang produzieren, diesen über die Resonanzen des Vokaltrakts umformen und dadurch unabhängig von der Grundtonhöhe und in eindeutiger Weise das jeweils vokalspezifische physikalische Merkmal produzieren, gemäss welchem Hörende Vokale erkennen. Aber es gehört zu den tatsächlich produzierten Vokallauten, systematisch von dieser Möglichkeit *abzuweichen*. Die Empirie der Vokallaute verweist darauf, dass Menschen Vokallaute nicht in systematischer Weise so erzeugen, wie es von der Physik und der Physiologie her als vorgegeben erscheint. Dieser *Kontrast* könnte sich für die zukünftige Theoriebildung als grundlegend erweisen.

Wir haben die sich hier ergebende Ausgangslage an anderer Stelle wie folgt formuliert: Entweder stellen die Resonanzwirkungen und entsprechenden Resonanzmuster von Rachen, Mund und Nase nicht die richtige physikalische Grösse dar, auf welche sich die Sprache in direkter Weise bezieht, doch es lässt sich eine andere solche physikalische Grösse finden. Dann steht bloss eine Ersetzung eines beste-

henden (physikalischen) Modells durch ein anderes, nicht aber ein grundsätzlicher Perspektivenwechsel an. Oder aber es findet sich in der Physik abseits der Sprache keine Anordnung, kein Instrument, keinen Vorgang, mit deren Hilfe alle möglichen Erscheinungsformen der Vokale erklärt und erzeugt werden können. Erweist sich dies als der Fall, so lässt sich die menschliche Stimme nicht auf eine einfache physikalische Grösse beziehen, sondern die Stimme würde diese «Grösse» erst produzieren.

Aufgrund unserer Interpretation der sich ergebenden Ausgangslage gehören unsere Arbeiten zum zweiten Vorhaben. Dieses verlangt nach einer Phänomenologie und, in deren Verlauf, nach einer Theoriebildung.

### **13.6 Phänomenologie**

Die Erläuterungen in den vorangehenden Teilen verweisen darauf, dass wir einerseits nur über fragmentarische Dokumentationen von Vokallauten verfügen und dass andererseits die Methode der Beschreibung ihrer akustischen Merkmale substanzielle Mängel und Beschränkungen aufweist.

Es steht deshalb eine Phänomenologie an, eine schrittweise aufbauende und sich fortlaufend systematisierende Zusammenstellung von Vokallauten jeder möglichen Art ihrer Produktion, verbunden mit einer sich parallel dazu fortlaufend zu entwickelnden neuen Methode der Beschreibung ihrer akustischen Merkmale.

Eine derartige Phänomenologie bezieht sich zunächst auf isoliert gebildete oder aus einem Lautkontext isolierte Vokallaute jeweils einer Einzelsprache mit quasi-konstantem Klang- oder Geräuschcharakter. Sie setzt voraus, dass diese Vokallaute einheitlich erkannt werden und bezieht entsprechende Identifikationstests mit ein.

### 13.7 Theoriebildung

Wir wiederholen und insistieren: Vokallaute, als isolierte Klänge oder Geräusche wahrgenommen, sind der Möglichkeit nach für sich verständlich, und diese Tatsache gehört zum Kern der Sprache selbst. Für sie stellt sich die psychophysische Frage, welches physikalisch beschreibbare Merkmal oder welches Ensemble physikalisch beschreibbarer Merkmale jeweils einem wahrnehmend erkannten Vokal entspricht, beziehungsweise welche Unterschiede dieses Merkmals oder Merkmalensembles den Unterschieden verschiedener Vokale entsprechen.

Eine Theoriebildung sieht sich dabei mit einem dreifachen Anspruch konfrontiert: (1) Sie muss eine Systematik einer orthogonalen Beschreibung vokalspezifischer akustischer Merkmale hervorbringen. Nur eine solche Beschreibung erlaubt es, von empirisch ermittelten Merkmalen ausgehend, Vokallaute synthetisch zu reproduzieren und in dieser Weise die Analysen auf ihren Aussagewert hin zu überprüfen. (2) Sie muss aus der Phänomenologie heraus Thesen ableiten, gemäss welchen Vokale unabhängig vom jeweiligen Fall eines einzelnen Lauts physikalisch repräsentiert werden. Diese Thesen müssen einerseits mit Anforderungen von Verifikation und Falsifikation verbunden werden und andererseits auf verschiedene Einzelsprachen übertragbar sein. (3) Sie muss empirische Befunde und abgeleitete Thesen zu erklären suchen.

## Nachwort

Die vorliegende Schrift wird viele Fragen aufwerfen, auf welche sie, wie einleitend erwähnt, nicht im Einzelnen eingeht, und es ist aufgrund unserer Erfahrung bisheriger Diskussionen zu erwarten, dass ein Teil der Darstellungen und Argumentationen gar grundsätzlich zurückgewiesen wird.

Auf Einzelaspekte, wie sie in der Literatur erörtert werden, wird hier nicht nur im Hinblick auf Einfachheit, Allgemeinheit und Deutlichkeit der Argumentation nicht vertiefend eingegangen, sondern auch deshalb, weil diese Einzelaspekte aus der Perspektive der vorliegenden Darstellungen in einem anderen Licht erscheinen und in einem anderen Kontext zu diskutieren sind, als dies häufig der Fall ist. Dazu die folgenden Beispiele und Hinweise.

Fehlende Ausprägungen spektraler Energiemaxima in allgemeiner Weise über Charakteristika der auditiven Wahrnehmung als «Formantverschmelzungen» («formant merging») erklären zu wollen, ohne der gesamten Systematik vokalspezifischer spektraler Charakteristika — insbesondere deren Abhängigkeit von der Grundtonhöhe und der möglichen «Ambiguität» von Hüllkurvenspitzen und von Formantmustern — Rechenschaft zu leisten, erachten wir als kritisch.

Entsprechendes gilt für «Normalisierungsversuche» von — der Annahme nach — grundsätzlich verschiedenen vokalspezifischen Formantmustern für Kinder, Frauen und Männer: Diese «Normalisierungsversuche» würden nur schon dann ganz anders angesetzt werden müssen, wenn die Vergleiche von Lauten und Formantmustern der verschiedenen allgemeinen Sprechergruppen, auf welche sich die «Normalisierungen» beziehen, nicht nur je verschiedene, sondern auch untereinander identische Grundtonhöhen mit einbeziehen würden.

Entsprechendes gilt auch für den Versuch, beobachtbare «Verschiebungen» von Formanten für Laute eines Vokals bei ausgeprägter Variation der Grundtonhöhe in einen grundsätzlichen Zusammenhang mit paralinguistischen Charakteristika zu stellen, insbesondere in einen Zusammenhang mit der Lautstärke («vocal effort»): Tief und

hoch gebildete Laute können gleichermaßen laut und leise gebildet werden, und Formantmuster von Lauten beinahe aller Vokale sind grundsätzlich abhängig von der Grundtonhöhe, was direkt eine «Ambiguität» von Formantmustern — isoliert und für sich betrachtet — ableiten lässt und darin in keinem Zusammenhang mit paralinguistischen Aspekten steht.

Entsprechendes gilt, um ein letztes Beispiel aufzuführen, auch für die Erörterung des «Formanttunings» im Rahmen der Untersuchung von Vokalisationen bestimmter Gesangsstile: Bisherige Studien des Gesangs lassen nicht ableiten, ob die dokumentierten Beobachtungen sich nur auf jeweils individuelle Eigenheiten einzelner Sängerinnen und Sänger beziehen, oder auf stilistische Charakteristika jeweils einer Gesangstechnik und eines Gesangsstils, oder ob sie Grundcharakteristika von Vokallauten selbst betreffen.

Wie denn die in der Literatur häufig vorzufindende Tendenz, Vokallaute auf tieferen und nur beschränkt ändernden Grundtonhöhen im Wesentlichen als Eigenart des Sprechens und Vokallaute auf mittleren und höheren beziehungsweise stark variierenden Grundtonhöhen als Eigenart des Singens aufzufassen und entsprechend zu untersuchen, grundsätzlich zurückzuweisen ist: Es besteht kein «normales» Sprechen, für welches im Unterschied zum Singen statistisch eine einzelne «durchschnittliche» Grundfrequenz ermittelt werden könnte, weder im Alltag, noch in unterhaltenden stimmlichen Äusserungen, noch in der Kunst und der Interpretation. Entsprechende Angaben in bestehenden Formantstatistiken sind nicht repräsentativ für das erfahrbare Sprechen und die dabei beobachtbaren akustischen Merkmale von Vokallauten, sondern nur repräsentativ für Laute, welche in einem kleineren Raum vor einem Mikrofon in entspannter Weise und quasi-monoton geäußert werden. (Auch noch bei einer solch eingeschränkten Formulierung fehlt zudem die Relativierung im Hinblick auf den jeweiligen Kontext von Einzelsprache und «Kultur».) Der Bereich der Grundtonhöhen für den Gesang ist nicht grundsätzlich verschieden von demjenigen des Sprechens, wie immer auch diese beiden Kategorien in wissenschaftlich sinnvoller Weise bestimmt und voneinander unterschieden werden. Jedes aufmerksame Hinhören im Alltag wird diese Feststellung evident werden lassen, jede ausgedehnte Erfahrung von Theater erlaubt es, sie auszudifferenzieren, und beides ist in einer Phänomenologie der Akustik des Vokals einzuho-

len. (In den Materialien finden sich dazu entsprechende Beispiele.) In dieser Weise erscheinen Einzelaspekte in einem anderen Zusammenhang, als sie in der Literatur meistens erörtert und ausgelegt werden.

Die vorliegende Kritik des theoretischen Grundsatzes muss, so haben wir zu Beginn des Nachworts angedeutet, eine Skepsis dieser Kritik selbst gegenüber hervorrufen. In vielen Fachdiskussionen, so haben wir ebenfalls erwähnt, ist uns denn diese Skepsis auch deutlich geworden, verbunden mit je einzelnen Argumentationsweisen. Wir haben versucht, diesen Argumenten in der vorliegenden Schrift Rechnung zu tragen. Im Rahmen dieses Nachworts seien zusätzlich die folgenden allgemeinen Bemerkungen beigelegt.

Welch ausserordentliche und häufig überraschende Rolle die Wahrnehmung bei der Erkennung von Sprachlauten auch spielt, so relativiert diese Rolle nicht die Tatsache, dass isoliert gebildete Vokallaute mit quasi-statischem Klang ausserhalb eines konkreten syntaktischen und semantischen Kontextes verständlich und ihre Partialtonspektren vokalspezifisch sind. Für solche Vokallaute ist eine Psychophysik, eine Theorie der Entsprechung von wahrgenommener Identität und physikalischem Merkmal oder Ensemble von Merkmalen also nicht nur möglich, sondern notwendig. Eine Psychophysik der Vokallaute gehört zu den Grundlagen der Untersuchung von Stimme und gesprochener Sprache.

Eine entsprechende Theorie über Formantmuster zu formulieren kann jedoch — dies soll die vorliegende Schrift deutlich machen — nicht gelingen. Also steht ein anderer, neu zu formulierender Ansatz an.

Wie aber kommt es, dass eine solche Argumentation bis anhin nicht mit ins Zentrum der Erörterung des Vokals und mit ihm der Stimme und der gesprochenen Sprache gestellt wird? Mehrere Gründe mögen eine Rolle spielen. Insbesondere aber fehlt — und dies erachten wir als vorrangig — eine robuste, empirische, systematische und repräsentative Dokumentation der im Textteil erläuterten Aspekte, auf welche eine Erörterung in verbindlicher Weise als empirische Referenz Bezug nehmen kann, und es fehlt die Formulierung einer alternativen Theorie.

So kommt es, dass bestehende Einzelwerte in Untersuchungen von Vokallauten zwar für die jeweiligen Bedingungen ihrer Erhebung zutreffen mögen, dass die Interpretation dieser Werte aber häufig im Hinblick auf eine allgemeine physikalische Repräsentation des Vokals vorgenommen wird, was empirisch nicht nachvollziehbar ist. Die Verallgemeinerung ist das kritische Moment.

Wir wiederholen: Es trifft zu, dass durchschnittliche Formantmuster, statistisch für jeweils eine der drei allgemeinen Sprechergruppen ermittelt und bezogen auf durchschnittliche Grundtonhöhen, wie sie sich bei entspanntem und quasi-monotonem Sprechen in einem geschlossenen kleineren Raum vor einem Mikrofon ergeben, in der Regel vokalspezifisch sind, aber es trifft nicht zu, dass dies auch für ausgeprägte Variationen der Grundtonhöhe gilt, wie sie als prosodische Charakteristika nur schon in der Alltagssprache erscheinen. — Es trifft zu, dass bei Vokallauten von tieferen Männerstimmen und einer Variationen der Grundtonhöhe von einer Oktave die Formantmuster für jeweils einen Vokal als unabhängig von der Grundtonhöhe erscheinen können, aber Umgekehrtes ist der Fall für Vokallaute der Mehrzahl der Frauen und beinahe aller Kinder. — Es trifft zu, dass eine bestimmte Grundtonhöhe und ein bestimmtes Muster von Filtern in einer Klangsynthese einen bestimmten Vokal erkennen lassen, aber es trifft in vielen Fällen nicht zu, dass derselbe Vokal auch dann erkannt wird, wenn das Muster der Filter konstant bleibt, die Grundtonhöhe aber substantziell geändert wird. Und Ähnliches mehr.

Weil verlässliche und für die wissenschaftliche Gemeinschaft vollständig nachvollziehbare empirische Referenzen mit systematischem Charakter fehlen, und weil in der allgemeinen und zusammenfassenden Literatur die akustischen Merkmale von Vokalen häufig ohne jede Relativierung und Problematisierung abgehandelt werden, sind die hier dargestellten Überlegungen, Erfahrungen und Beobachtungen zum Teil nicht bekannt, nicht nachvollzogen und in keinen gegenseitigen Zusammenhang gestellt. Dies erschwert zunächst die Erörterung innerhalb des Bereichs der Phonetik und der Psychophysik. Darüber hinaus ist aber auch zu bedenken, welche Bedeutung diese fehlende Relativierung für andere Bereiche der Wissenschaft einnimmt, wie — um nur zwei Beispiele zu geben — für die Neuropsychologie oder die Pathologie: Wie sollen deren experimentelle Anlagen in verlässlicher Weise gestaltet werden, wenn die Problematik

der Verallgemeinerung von Einzelmessungen nicht mit ins Zentrum der Voraussetzungen und Analysemethoden gestellt wird!

Hinzu kommt, dass einzelne der dargestellten Probleme in Fachkreisen zwar erkannt sind, einer grundsätzlich kritischen Reflexion wie der vorliegenden gegenüber aber häufig der Vorwurf erhoben wird, sie bleibe ohne Formulierung einer neuen Theorie. Eine solche Argumentation entspricht aber nicht unserer Auffassung und Einstellung. Die Kritik des bestehenden akustischen Grundsatzes — trifft sie zu — hat darin ihren eigenen und vollständigen Wert, ganz unabhängig davon, was diejenigen, welche sie formulieren, darüber hinaus zu leisten vermögen.

Wie denn auch das Vorhaben einer Phänomenologie und neuen Theoriebildung ausserordentlich aufwendig ist und entsprechende Ressourcen verlangt. Dies bedingt seinerseits, dass die Fachgemeinschaft die Bedeutung eines solchen Vorhabens anerkennt. Eine solche Anerkennung aber verlangt vorgängig nach einer nachvollziehbaren Kritik des bestehenden Grundsatzes, verbunden mit einer neuen Interpretation bisheriger empirischer Befunde.

Wir haben die vorliegende Schrift auch deshalb verfasst, weil wir nicht sicher sind, wie weit unsere eigenen Beiträge zu einer Phänomenologie und neuen Theoriebildung gehen werden. Erst kurz vor Abschluss des Manuskripts gelang es uns, die Ressourcen für eine Untersuchung zu sichern, in welcher unausgebildete Stimmen (Kinder, Frauen und Männer) und ausgebildete Stimmen (Theater und Gesang, Frauen und Männer) miteinander verglichen werden, um dabei die spektralen Charakteristika, welche die Vokallaute und die Stimmcharakteristik betreffen, zu beobachten und zu dokumentieren. Dies wird es uns erlauben, einen Beitrag zur anstehenden Phänomenologie leisten zu können, welcher Klärungen zum Verhältnis von Sprechen und Singen, von Vokalspektrum und Grundtonhöhe, von Vokalspektrum und Produktionsstil sowie von Vokalspektrum und Sprechergruppe beziehungsweise Vokaltraktgrösse verspricht.

Was die anstehende Theoriebildung betrifft, erwarten wir zwar, dass eine Phänomenologie der Vokallaute die Regeln des Verhaltens von Partialtonspektren jeweils eines Vokals formulieren lässt, dass aber eine solche Formulierung nur den Übergang zu einer eigentlichen

Theorie darstellt: Partialtonspektren sind als solche abhängig von der jeweiligen Grundtonhöhe, der Vokal aber wird unabhängig von letzterer erkannt, und also muss die physikalische Beschreibung des Vokals gleichermassen von der Grundtonhöhe unabhängige Grössen betreffen; ganz abgesehen von der Tatsache, dass Vokale auch geflüstert und dementsprechend als Geräusche erkannt werden können, und sie dann im Spektrum keine Partialtöne aufweisen.

Was die allgemeine Bedeutung der Thematik akustischer Merkmale von Vokallauten betrifft, sollte man diese nicht als blosser Sache einer einzelnen Disziplin einschätzen. Dazu die beiden abschliessenden Hinweise.

Derzeit erfährt die Stimme in Human- und Kunstwissenschaften eine besondere Aufmerksamkeit. Die entsprechende Reflexion nimmt dabei direkten Bezug auf Wissen und Erfahrungen der künstlerischen Kreation und Interpretation, und die sich heute ausdifferenzierende Stimmforschung erhebt mit Nachdruck den Anspruch auf eine interdisziplinäre Anlage. Die Forschungskultur zur Ästhetik der Stimme sollte sich dabei unserer Auffassung nach auch eine besondere Haltung zur Akustik der Stimme zu eigen machen: nicht nur auf die bestehenden Beschreibungen stimmlicher Äusserungen in der Phonetik zu verweisen, sondern sie ausdrücklich ins Zentrum der Erörterung zu stellen. Dass man sich nicht täusche: Stellt sich die Frage nach dem stimmhaften Sprachlaut, so stellt sich die Frage nach der Stimme selbst.

Die Stimmbänder erzeugen einen Klang, und die Hohlräume von Rachen, Mund und Nase könnten diesen Klang durch ihre Resonanzwirkung zu einem Formantmuster eines Lauts bilden, welches den Vokal physikalisch repräsentiert und also erkennen lässt. Empirisch untersucht, weichen die spektralen Charakteristika von Vokallauten aber von einer solchen Möglichkeit in systematischer Weise ab. Diese Feststellung lässt darauf schliessen, dass wir uns erst im Vorfeld einer Theorie der physikalischen Repräsentation des Vokals, *und damit seiner Form*, befinden.



## 05 Einleitung

### Teil 1

#### **Bestehender theoretischer Grundsatz, bestehende empirische Referenzen**

- 17 1 Bestehender theoretischer Grundsatz**
  - 1.1 Allgemeine akustische Charakteristik von Vokallauten
  - 1.2 Im engen Sinne sprachspezifische akustische Charakteristik von Vokallauten
  - 1.3 Relativierung in Bezug auf allgemeine Sprechergruppen
  - 1.4 Einschränkung auf stimmhafte orale Vokallaute
  - 1.5 Einschränkung auf isolierte Vokallaute
  - 1.6 Einschränkung auf Vokallaute mit quasi-konstanter Klangcharakteristik
  - 1.7 Einschränkung auf eine Sprachgemeinschaft
  - 1.8 Theoretischer Grundsatz zur physikalischen Repräsentation des Vokals
  - 1.9 Formalisierung des theoretischen Grundsatzes
  - 1.10 Illustration
- 28 2 Bestehende empirische Referenzen**
  - 2.1 Allgemeine Referenzen
  - 2.2 Referenz für die deutsche Sprache (Standard)
  - 2.3 Zusatz: Verweise auf andere statistische Angaben

## Teil 2

### Überlegungen

- 43 3 Vokale und Anzahl Formanten**
- 3.1 Inkonstanz der Anzahl vokalspezifischer spektraler Energiemaxima bei hinteren Vokalen und /a-ɑ/
  - 3.2 Inkonstanz der Entsprechung von vokalspezifischen spektralen Energiemaxima mit vokalspezifischen Formantmustern
  - 3.3 Inkonstanz der Anzahl vokalspezifischer Formanten bei vorderen und r-gefärbten Vokalen
  - 3.4 Zusatz: «Unechte» Formanten
  - 3.5 Zusatz: Inkonstanz der Anzahl vokalspezifischer Formanten in Synthesen
- 46 4 Vokale und Grundtonhöhe**
- 4.1 Grundtonhöhe, erster Formant und «Qualität» von einzelnen Vokalen
  - 4.2 Grundtonhöhe, spektrale Hüllkurve, Formantmuster und «Qualität» von Vokallauten
- 49 5 Formantmuster und allgemeine Sprechergruppen**
- 5.1 Grundtonhöhe, spektrale Hüllkurve, Formantmuster und «Qualität» von Vokalen von Kindern, Frauen und Männern
  - 5.2 Verschiedene Formantmuster für gleiche Vokale
  - 5.3 Gleiche Formantmuster für verschiedene Vokale
  - 5.4 Lücke in der Begründung
  - 5.5 Zusatz: Formantmuster stimmhafter und geflüsterter Vokallaute
- 56 6 Begriffe, Messmethode der Formantbestimmung**
- 6.1 Formant und Lautspektrum
  - 6.2 Allgemeine Sprechergruppe und Vokaltraktgröße
  - 6.3 Objektivität von Formantanalysen im Allgemeinen
  - 6.4 Abhängigkeit und Beschränkung von Formantanalysen aufgrund ihrer Beziehung zu Grundtonhöhe und Sprechergruppe beziehungsweise Vokaltraktgröße
  - 6.5 Zusatz: Anpassungen von Parametern in der Formantanalyse und Inkonsistenz der Bezugnahme auf eine Vokaltraktgröße
  - 6.6 Zusatz: Spektrum, Formantmuster und Resynthesen

### Teil 3

#### Erfahrungen und Beobachtungen

- 67**    **7**    **Unsystematische Entsprechung von Vokalen mit Mustern spektraler Energiemaxima und mit Formantmustern**
- 7.1 Inkonstanz der Anzahl vokalspezifischer spektraler Energiemaxima und mögliche Inkongruenz von Formantmustern
  - 7.2 Teilweise fehlende Ausprägung vokalspezifischer spektraler Energiemaxima
  - 7.3 Zusatz: Nachvollzug anhand von Resynthesen und Synthesen
- 71**    **8**    **Fehlende Entsprechung von Vokalen mit Mustern spektraler Energiemaxima oder mit Formantmustern**
- 8.1 Abhängigkeit der vokalspezifischen spektralen Energiemaxima und der tieferen Formanten  $\leq 1,5$  kHz von der Grundtonhöhe
  - 8.2 Erkennbare Vokallaute auf Grundtonhöhen oberhalb statistischer Werte der ersten Formantfrequenz
  - 8.3 «Inversionen» spektraler Maxima und Minima und «inverse» Formantmuster bei Lauten eines einzelnen Vokals
  - 8.4 Zusatz: Flüstervokale, Grundtonabhängigkeit vokalspezifischer spektraler Merkmale, «Inversionen»
  - 8.5 Zusatz: Nachvollzug anhand von Resynthesen und Synthesen
- 76**    **9**    **«Ambige» Entsprechung von Vokalen mit Mustern spektraler Energiemaxima oder mit Formantmustern oder mit vollständigen spektralen Hüllkurven**
- 9.1 «Ambige» Muster spektraler Energiemaxima und «ambige» Formantmuster
  - 9.2 «Ambige» spektrale Hüllkurven
  - 9.3 «Ambiguität» und Einzelvokale
  - 9.4 Zusatz: «Ambiguität» spektraler Charakteristika von Flüsterlauten
  - 9.5 Zusatz: Nachvollzug anhand von Resynthesen und Synthesen

- 79 10 Fehlende Entsprechung von Mustern spektraler Maxima oder von Formantmustern mit allgemeinen Sprechergruppen beziehungsweise Vokaltraktgrößen**
- 10.1 Gleiche Muster spektraler Maxima und gleiche Formantmuster  $\leq 1,5$  kHz für verschiedene allgemeine Sprechergruppen wie für verschiedene Vokaltraktgrößen
  - 10.2 «Dichotomie» des Vokalspektrums
  - 10.3 Zusatz: Geflüsterte Vokallaute im Vergleich verschiedener allgemeiner Sprechergruppen beziehungsweise verschiedener Vokaltraktgrößen
  - 10.4 Zusatz: Vokalimitationen durch Vögel
  - 10.5 Zusatz: Nachvollzug anhand von Resynthesen und Synthesen
- 84 11 Fehlende Entsprechung von methodischen Begrenzungen der Formantbestimmung mit Begrenzungen der Vokalerkennung**
- 11.1 Erkennbarkeit von Vokallauten auf Grundtonhöhen  $> 350$  Hz
  - 11.2 Fehlende Entsprechung von methodischen Problemen der Bestimmung von Formantmustern auf Grundtonhöhen  $\leq 350$  Hz mit einer Abnahme der Erkennbarkeit von Vokallauten
  - 11.3 Zusatz: Fehlende methodische Grundlage für die Bestimmung von Formantmustern bei Vokalimitationen durch Vögel

**Teil 4**  
**Falsifikation**

- 87 12 Empirische Falsifikation trotz methodischer Beschränkungen der Bestimmung von spektralen Hüllkurvenmaxima oder von Formantmustern**
- 12.1 Fehlende methodische Grundlage für eine Verifikation des theoretischen Grundsatzes
  - 12.2 Systematisches Abweichen empirischer Befunde von Voraussagen des theoretischen Grundsatzes
  - 12.3 Dem theoretischen Grundsatz direkt widersprechende empirische Befunde

**Teil 5**  
**Kommentar**

- 91 13 Präliminarien**
- 13.1 Hindernisse einer Anpassung oder Ausdifferenzierung des theoretischen Grundsatzes
  - 13.2 Der theoretische Grundsatz als Index
  - 13.3 Einschub: Voraussetzung einer einheitlichen Erkennung von Vokalen
  - 13.4 Vorfeld
  - 13.5 Zwei Ansätze
  - 13.6 Phänomenologie
  - 13.7 Theoriebildung

**98 Nachwort**

**Materialien**

<http://subtexte08.ipf.zhdk.ch>





## Impressum

© 2013 für diese Ausgabe: Zürcher Hochschule der Künste.  
Alle Rechte vorbehalten. Jede Form der Vervielfältigung,  
Mikroverfilmung oder Einspeicherung und Verarbeitung in  
elektronischen Medien nur mit schriftlicher Genehmigung  
des Institute for the Performing Arts and Film.

Das IPF bedankt sich für die freundliche Unterstützung der  
Publikation durch das Konsortium W6 zur Förderung des  
Wissens- und Technologietransfers.

Herausgeber: IPF, Institute for the Performing Arts and Film

Leitung: Anton Rey

Departement Darstellende Künste und Film, ZHdK

Direktion: Hartmut Wickert

Redaktion: Yvonne Schmidt

Lektorat: druckreif, Zürich

Grafik: Jeannine Herrmann

Druck und Bindung: OK Haller Druck AG, Zürich

Papier: Normaset Puro, naturaweiss, matt 120g/m<sup>2</sup>

Cyclus Offset, weiss, matt, 300g/m<sup>2</sup> (Umschlag)

Schriften: Utopia, Trade Gothic

1. Auflage, Mai 2013

Gessnerallee 11, 8001 Zürich

[ipf.zhdk.ch](http://ipf.zhdk.ch), [www.subtexte.ch](http://www.subtexte.ch)

[subtexte08.ipf.zhdk.ch](mailto:subtexte08.ipf.zhdk.ch) (Materialien)

[info.ipf@zhdk.ch](mailto:info.ipf@zhdk.ch)

Vertrieb: Museum für Gestaltung Zürich

Ausstellungsstrasse 60, Postfach

8031 Zürich

[verlag@museum-gestaltung.ch](mailto:verlag@museum-gestaltung.ch)

ISBN 978-3-906437-39-2

**Z**

hdk

Zürcher Hochschule der Künste  
Institute for the Performing Arts and Film

**w6**  
KONSORTIUM