

Simultane Inversfilterung und Schätzung des glottalen Flusses aus akustischen Stimmsignalen

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Matthias Fröhlich
aus Stuttgart

Göttingen 1999

D 7

Referent: Prof. Dr. M. R. Schroeder

Korreferent: Prof. Dr. D. Ronneberger

Tag der mündlichen Prüfung: 1. November 1999

Inhaltsverzeichnis

Zu dieser Arbeit	5
1 Grundlagen	7
1.1 Hintergründe	7
1.2 Stand der Forschung	15
2 Methoden	19
2.1 Das LF-Modell	19
2.2 Das selbstschwingende Zweimassen-Modell	22
2.3 Der Sprachsynthesator	26
2.4 Bestimmung der Periodenlängen	28
2.5 Verwendete Testsignale	31
2.6 Exkurs: Bestimmung des Referenzmodells bei Signalen 2MMS und 2MMF	35
2.7 Parameter zur Charakterisierung des glottalen Flusses	36
2.7.1 Der Öffnungsquotient (OQ)	37
2.7.2 Der Schließquotient (CQ)	38
2.7.3 Der Asymmetriequotient (SQ)	38
2.7.4 Der „parabolic spectral parameter“ (PSP)	39
2.8 Statistische Auswertungen	40
2.9 Die „Diskrete Nulpole-Modellierung“ (DAP)	41
2.9.1 Ursprüngliches Verfahren	41
2.9.2 Erweiterung zur Schätzung des glottalen Flusses	44
2.10 Die SIM-Methode	49
2.11 Multidimensionale Optimierung	51
2.12 Verwendete Fehlermaße und Optimierungsmethoden	55
3 Ergebnisse	59
3.1 Signaltyp FLT und SYN	59
3.1.1 Bestimmung der Schwelle σ in der Berechnung von OQ	59
3.1.2 Streuung der Schätzwerte	60
3.1.3 Genauigkeit der Schätzung	62
3.2 Signaltyp 2MMS	67
3.2.1 Bestimmung der Referenzparameter	67
3.2.2 Abhängigkeit von θ	68

3.3	Signaltyp 2MMF	73
3.4	Signaltyp ECHT	74
3.4.1	Unterschiede zwischen Phonationsarten	74
3.4.2	Übergänge zwischen Phonationsarten	75
3.4.3	Zusammensetzung des Datenraums	78
4	Diskussion und Ausblick	83
5	Zusammenfassung und Schlußfolgerungen	87
	Literatur	89
	Anhang:	
A	Erweiterung des Zweimassenmodells auf eine nicht-rechteckige Glottisfläche	99
A.1	Ursprüngliches Zweimassenmodell	99
A.2	Erweitertes Zweimassenmodell	101
B	Der DAP-Algorithmus	105
C	Analyseergebnisse für verschiedene Signalarten	107
D	Liste der häufig verwendeten Abkürzungen	125
	Danksagung	127

Zu dieser Arbeit

Ausgangspunkt dieser Arbeit war der von phoniatriischer Seite geäußerte Bedarf, zuverlässig bestimmbare Größen zur objektiven Beschreibung der Stimmgebung (Phonation) anhandgestellt zu bekommen. Um die Anwendbarkeit dieser Größen in der phoniatriischen Praxis zu gewährleisten, sollten sie eine direkte Interpretation im Sinne des Stimmgebungsmechanismus zulassen. Bei diesen Anforderungen bot sich die Analyse des akustischen Stimmsignals an, da es eine objektive, d.h. automatische und unüberwachte Bestimmung charakteristischer Größen ermöglicht, die sich prinzipiell im Sinne der zugrundeliegenden glottalen Vorgänge interpretieren lassen sollten.

In der vorliegenden Arbeit wird ein neues Verfahren zur Berechnung solcher glottalen Parameter aus dem akustischen Stimmsignal vorgestellt. Die grundlegende Idee ist dabei die gleichzeitige Anpassung eines Modells des glottalen Flusses an das akustische Stimmsignal während der Inversfilterung, die dazu dient, den Einfluß des Vokaltrakts im abgestrahlten Signal zu eliminieren und somit das Stimmsignal rechnerisch auf seine Ursache – die Vorgänge während der Phonation – zurückzuführen. Dieses Vorgehen ermöglicht einerseits eine „bessere“ Inversfilterung, andererseits die leichtere und konsistentere Berechnung glottaler Parameter, die sich nun störungsfrei aus dem angepaßten Modell bestimmen lassen.

Die Arbeit verfolgt daher mehrere Ziele: Zum einen wird die neue Methode eingeführt, mittels der sich phonationsbezogene Parameter aus dem Stimmsignal berechnen lassen, zum anderen wird die Anwendung dieses Verfahrens und der sich ergebenden Parameter anhand verschiedener Testsignale überprüft und bewertet.

Hieraus ergibt sich der Aufbau der Dissertation: In Kapitel 1 erfolgt die Einführung in die Thematik und eine Darstellung des momentanen Forschungsstandes bezüglich der relevanten Themen. Daraufhin werden in Kapitel 2 die einzelnen Komponenten des Verfahrens und der vorgenommenen Untersuchungen beleuchtet. Die Abschnitte 2.1–2.8 referieren dabei die Grundlagen verschiedener Einzelkomponenten, wonach in den Abschnitten 2.9–2.12 die neu entwickelte Methode dargestellt wird. In Kapitel 3 werden die Ergebnisse der verschiedenen Untersuchungen – nach Signaltyp geordnet – vorgestellt und diskutiert. Eine allgemeine Diskussion schließt sich in Kapitel 4 an. Nach der Zusammenfassung und den Schlußfolgerungen (Kapitel 5) folgen die Literaturangaben. Anhang A enthält die Beschreibung der Erweiterung des selbstschwingenden Stimmlippenmodells, die eine Voraussetzung für einen Teil der durchgeführten Tests darstellt. In Anhang B wird der DAP-Algorithmus, auf dem das entwickelte Verfahren beruht, mathematisch ausgebreitet. Anhang C beinhaltet ergänzende Tabellen zu Kapitel 3 „Ergebnisse“. Eine Übersicht über die verwendeten Abkürzungen findet sich zu Referenzzwecken in Anhang D.

Im gesamten Text werden Dezimalstellen von Zahlen mit „.“ anstatt – wie im Deutschen sonst üblich – mit „ , “ angezeigt, um eine konsistente Darstellung der Ergebnisse der meist angelsächsisch ausgerichteten Analyse- und Darstellungsprogramme zu erreichen (die im Deutschen mögliche Markierung der Tausender durch „.“ wird nicht verwendet). Englische Bezeichnungen werden, sofern keine allgemein übliche deutsche Entsprechung existiert, *kursiv* gesetzt.

1. Grundlagen

1.1 Hintergründe

Akustische Größen zur Charakterisierung einer Stimme finden sich in einem weiten Bereich des modernen täglichen Lebens. So können beispielsweise die unter kommunikativen Aspekten bedeutungstragenden Signalstrukturen extrahiert und durch geeignete Größen beschrieben werden, wodurch sich ihre Verwendung im Bereich der Spracherkennung begründet. Daneben kann der „Träger“ der Sprache, d.h. die Stimme, unabhängig von der Bedeutungsübertragung analysiert werden. Diese Zielrichtung wird im Bereich der Sprecherverifizierung oder Sprechererkennung verfolgt, die beide in zunehmendem Maß einen wichtigen Aspekt in der Sicherheitstechnik darstellen. Da sich der größte Teil der zwischenmenschlichen Kommunikation auf sprachlicher Ebene vollzieht, stellt die Stimme auch unter gesundheitlichen Aspekten ein wichtiges Untersuchungs- und Forschungsobjekt dar, wobei hier ebenfalls der Bedarf besteht, Stimmereigenschaften durch geeignete Kenngrößen charakterisieren zu können. Nicht zuletzt unter volkswirtschaftlichen Gesichtspunkten ist eine adäquate medizinische Versorgung und Überwachung von Stimmgebrauch und -abusus zur möglichst frühzeitigen Diagnose und Prophylaxe von Stimmschäden in zunehmendem Maß erforderlich (PEVOC-II, 1997).

Um diese verschiedenen Forschungs- und Anwendungsbereiche genauer zu umreißen, soll der Sprechvorgang detailliert analysiert werden (Abb. 1.1). Während der Sprachzeugung sind zwei prinzipielle Bereiche zu trennen: die Stimmgebung und die Artikulation. Die Artikulation erfolgt durch Zunge, Lippen, Kiefer und andere supra-laryngeale Strukturen und erzeugt die Kodierung der Information, die im linguistischen Sinn übertragen werden soll. Hiervon weitgehend unabhängig dient die Stimme als Träger dieser Information, wobei ihre Steuerung teilweise an die Artikulation gekoppelt ist (z.B. bei stimmlosen Phonemen). Auch mit verfälschtem Träger ist eine Übertragung der Information möglich (z.B. bei Flüstern oder Verwendung von Stimmprothesen oder Kehlkopfsummern), im Extremfall des Lippenlesens bei Gehörlosen sogar ganz ohne Träger.

Das aus Träger und kodierter Information bestehende akustische Signal wird in den Außenraum abgestrahlt und pflanzt sich dort als Schallwelle fort. Akustische Messungen erfolgen generell durch Aufzeichnung und Analyse der Signale in diesem Bereich, d.h. nach Abstrahlung in den Außenraum, entweder im akustischen Nah- oder Fernfeld. Aus der Warte des Zuhörers (*Rezipient*) gelangt das Schallsignal an sein Ohr und wird dort in neuronale Impulse umkodiert. Diese Impulse führen schließlich zur Wahrneh-

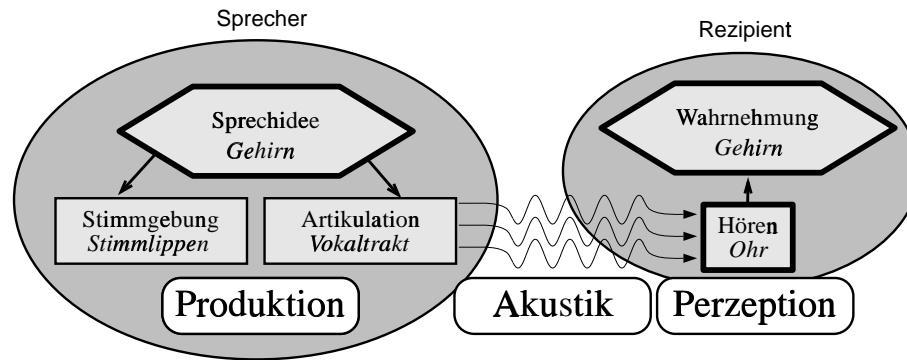


Abbildung 1.1: Sprache zur Kommunikation.

mung des Signals bei dem Rezipienten. Hierbei gehen psychoakustische und psychische Effekte ein, die teils generalisierbar, teils individuell geprägt sind, letzteres insbesondere bei komplexeren oder subtileren Nuancen der Informationsübertragung, die sich linguistisch dem Bereich der Pragmatik zuordnen lassen (z.B. Ironie). Dient die sprachliche Äußerung primär der Kommunikation, „filtert“ der Rezipient die Information heraus und läßt die Stimmgebung in erster Näherung unberücksichtigt. Durch Training kann jedoch die Aufmerksamkeit des Rezipienten verstärkt auf die Stimmgebung gelenkt werden, was insbesondere in der Gesangsausbildung oder – bei gesprochener Sprache – in der logopädischen Therapie als eine fundamentale Voraussetzung für die Eigenschaft, eine gute Therapeutin oder ein guter Therapeut zu sein, angesehen werden kann. Geschulte Personen sind in einem außerordentlich hohen Maß zur Differenzierung zwischen verschiedenen Stimmqualitäten fähig.

Auf der anderen Seite erschwert diese subjektive Komponente in der perzeptiven Bewertung von Stimmeigenschaften die Objektivierung der Ergebnisse (Kreiman und Gerratt, 1994, 1996, 1998). Um eine gute Reproduzierbarkeit der Ergebnisse zu erhalten, bietet sich die Analyse des akustischen Signals an („akustische Analyse“), da sie einfach durchzuführen ist und zu einem gegebenen Signal bereits durch die Definition der zu verwendenden Algorithmen spezifiziert ist.

Die verschiedenen Vorteile der akustischen Analyse haben in den vergangenen Dekaden eine starke Bewegung in der Entwicklung neuer Analyseverfahren hervorgerufen, sowohl unter wissenschaftlichen Fragestellungen als auch im Hinblick auf die klinische Versorgung und phoniatische Praxis. Die wichtigsten Aspekte mögen hierbei die bereits erwähnte Objektivität und Reproduzierbarkeit der Ergebnisse sein, die eine untersucherunabhängige Beurteilung der Stimme erlauben. Methodisch stellt die Analyse akustischer Aufnahmen eine nicht-invasive Technik dar, die weder zu physischen Eingriffen bei der untersuchten Person noch zu Beeinflussung der Stimmgebung führt. Sie ist zudem apparativ kostengünstig und benötigt kein spezialisiertes Bedienungspersonal, so daß sich hierdurch eine relativ einfache Möglichkeit ergibt, Stim-

men routinemäßig zu untersuchen. Diese theoretischen Vorteile werden z.B. im Göttinger Heiserkeits-Diagramm vereint (Michaelis und Strube, 1995; Michaelis et al., 1997, 1998a,b; Strube et al., 1996; Fröhlich et al., 1997, 1998a,b,c,d, 2000; Kruse et al., 1997, 1998). Unter dem bereits erwähnten prophylaktischen Aspekt der frühzeitigen Erkennung von Stimmstörungen bietet die akustische Stimmanalyse eine gute Ausgangsbasis für breitgefächerte Vorsorgeuntersuchungen (neudeutsch: „Screening“).

Jedoch stellt sich hier ein prinzipielles Problem, da sich aus phoniatischer Sicht die entscheidenden Informationen, die den Stimmgebungsprozeß betreffen, im akustischen Signal nur mittelbar äußern (Michaelis et al., 1998c,d). Daher kommt der Bestimmung direkter Größen der Stimmgebung (sogenannter „glottaler Parameter“) besondere Bedeutung zu. Zu ihrer Berechnung können unterschiedliche Herangehensweisen verfolgt werden, die verschiedene Vor- und Nachteile aufweisen. Die Laryngoskopie, bei der mittels eines optischen Geräts die Stimmlippenschwingung direkt betrachtet oder zeitlupenhaft sichtbar gemacht wird, stellt ein unverzichtbares Instrument für die phoniatische Untersuchung dar. Allerdings gestaltet sich eine weiterführende quantitative und automatische Auswertung der so erhaltenen Bildsequenzen aus verschiedenen Gründen schwierig. Bei der üblicherweise verwendeten Laryngo-Stroboskopie wird die Periodizität des Signals vorausgesetzt, die insbesondere bei mittel- und hochgradigen Stimmpathologien nicht notwendigerweise vorliegt. Die Bildanalyse von hochgeschwindigkeitsglottografischen Aufnahmen befindet sich noch in der Entwicklungs- bzw. Validierungsphase und ist aufgrund der vorherrschenden Rottöne, mangelnd scharf umrissener Bildelemente und Störungen durch Lichtreflexe und Schleimfäden mit deutlichen Unsicherheiten verbunden (Wittenberg et al., 1997; Lehmann et al., 1998). Die Methode der Kymogramme (Svec und Schutte, 1996) erlaubt keine direkte Aussage über die glottale Fläche und scheint aufgrund der eindimensionalen Datenerfassung anfällig gegenüber Verkippungen der Kamera bei quantitativen Auswertungen.

Invasive Methoden wie die Messung des glottalen Flusses durch Einführen einer Flußsonde bzw. zweier Mikrofone von außen durch den Larynx mögen zwar eine genaue Beschreibung der phonatorischen Vorgänge ermöglichen, sind jedoch aufgrund ihrer Invasivität nur in Ausnahmefällen für spezielle wissenschaftliche Untersuchungen vertretbar.

Das Elektroglottogramm (EGG) mißt den elektrischen Widerstand zwischen zwei Elektroden, die links und rechts des Larynx auf der Haut fixiert werden. Dieses Verfahren führt, außer einem eventuellen leichten Würgegefühl durch die Elektrodenbefestigung, zu keiner Beeinträchtigung der Stimmgebung. Da hierbei direkte Eigenschaften des laryngealen Schwingungssystems gemessen werden, bietet das EGG prinzipiell die Möglichkeit, direkte phonationsbezogene Parameter zu definieren. Die genauere Interpretation des EGG-Signals – abgesehen von der Bestimmung der lokalen Periodenlänge, die sich sehr präzise ablesen läßt – stellt jedoch ein solch gravierendes Problem dar (Baken, 1992), daß weiterführende Untersuchungen in dieser Richtung eher spärlich blieben (Titze et al., 1983; Titze, 1990; Holmberg et al., 1995a,b).

Eine Alternative zu den genannten Herangehensweisen bietet die „Rückrechnung“ der laryngealen Bedingungen aus dem abgestrahlten akustischen Signal. Bei diesem theoretisch motivierten Ansatz müßte der Einfluß des akustischen Signalleitungswegs

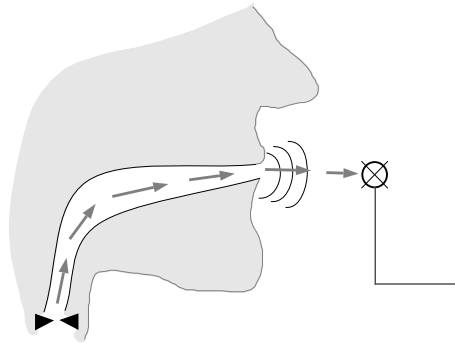


Abbildung 1.2: Skizze des akustischen Signalleitungswegs bestehend aus Schallerzeugung an der Glottis, Ausbreitung im Vokaltrakt, Abstrahlung an den Lippen, Fortpflanzung im Außenraum, Aufnahme durch das Mikrofon.

ab dem Punkt der Stimmgebung bekannt sein, um ihn dann „invers“ rechnerisch wieder zu eliminieren. Der Leitungsweg setzt sich dabei aus Vokaltrakt, Abstrahlung in den Außenraum, Fortpflanzung im Außenraum und Aufnahme durch das Mikrofon zusammen (Abb. 1.2).

Während die akustischen Eigenschaften der Schallausbreitung im Freifeld und die des Mikrofons als bekannt vorausgesetzt werden können, stellen die Komponenten „Vokaltrakt“ und „Abstrahlung“ insbesondere aufgrund der Variabilität des Vokaltrakts während einer sprachlichen Äußerung Unbekannte dar, die geschätzt werden müssen. Die Modellierung der Abstrahlung wurde in verschiedenen Näherungen durchgeführt (Wakita und Fant, 1978) und wirft keine besonderen Probleme auf. Die Schätzung der akustischen Eigenschaften des Vokaltrakts ist dagegen eine schwierige Aufgabe. Selbst unter der Bedingung eines stationären Vokaltrakts (d.h. keine Veränderung der Artikulatoren) sind zuverlässige Schätzungen schwer zu realisieren. Bildgebende Methoden liefern im Prinzip die Möglichkeit, aus der Geometrie des Vokaltrakts auf seine Resonanzeigenschaften zurückzuschließen. Methoden dieser Art sind jedoch in ihrer Durchführung aufwendig (Story et al., 1998) und setzen die Versuchsperson bei Verwendung von Röntgenstrahlen hohen Strahlendosen aus. Zudem beträgt die Auflösung der resultierenden Vokaltraktgeometrie 3–5mm und erfordert das Beibehalten der Artikulatorstellungen in der Größenordnung von einigen Minuten (Story et al., 1998), so daß die Resonanzeigenschaften nur mit gewisser Unsicherheit abgeschätzt werden können.

Die Alternative, d.h. die Schätzung der Eigenschaften des Stimmkanals ausschließlich aus akustischen Daten, besitzt eine lange Geschichte (vgl. Fant, 1993). Hierzu wird der Stimmgebungsprozeß in verschiedene Komponenten aufgeteilt, die sich getrennt beschreiben und unabhängig voneinander modellieren lassen. Die Anregung durch den glottalen Fluß (Abb. 1.3), die Resonanzen des Vokaltrakts und die Abstrahlung werden als getrennte Komponenten aufgefaßt, die zusammen ein lineares System bilden (Titze, 1994). Dieses sogenannte *Quelle-Filter-Prinzip* ist in Abbildung 1.4 dargestellt.

Da das System als linear angesetzt wird, kann das Ausgangssignal als Signal betrachtet werden, das sich aus dem Eingangssignal nach Passieren verschiedener linearer Filter ergeben hat. Nach dieser Modellvorstellung kann somit ein inverses Filter entworfen werden, mittels dessen das Eingangssignal verlustfrei aus dem Ausgangssignal rekonstruiert werden kann. Dieses Prinzip hat sich auch in der Sprachkodierung und -komprimierung als äußerst leistungsfähig erwiesen (Atal und Schroeder, 1970; Schroeder und Sloane, 1987; Schroeder, 1999). In der Praxis ergeben sich allerdings Einschränkungen bezüglich der theoretischen Annahmen aus Abweichungen vom Quelle-Filter-Prinzip (Fant, 1993), da z.B. der Vokaltrakt in Wirklichkeit einen nicht unerheblichen Einfluß auf die Glottisschwingung aufweist (Milenkovic, 1987; Kröger, 1991; Michaelis et al., 1998c,d).

Weitere Probleme ergeben sich bei der Methode zur Bestimmung der Koeffizienten des inversen Filters. Ein grundlegender Ansatz hierbei ist die lineare Prädiktion (LP), bei der lineare Redundanzen des Signals entfernt werden (Überblicke geben Makhoul, 1975; Markel und Gray, 1976; Schroeder, 1982, 1986). Kurzreichweitige Beziehungen zwischen Abtastwerten sind durch die Formanten, längerreichweitige durch die Grundfrequenz des Sprachsignals gegeben. Während die Prädiktion der letztgenannten, grundfrequenzabhängigen Signaleigenschaften einen wichtigen Aspekt für eine effiziente Signalkodierung darstellt (Schroeder und Atal, 1985; Schroeder, 1986), sollen im folgenden nur die kurzreichweitigen, auf die Resonanzen des Vokaltrakts zurückzuführenden Korrelationen zwischen benachbarten Abtastwerten betrachtet werden.

Bei der Verwendung von LP- oder LPC (*linear predictive coding*)-basierten Ansätzen können die Filterkoeffizienten streng genommen nur während der Verschußphase ge-

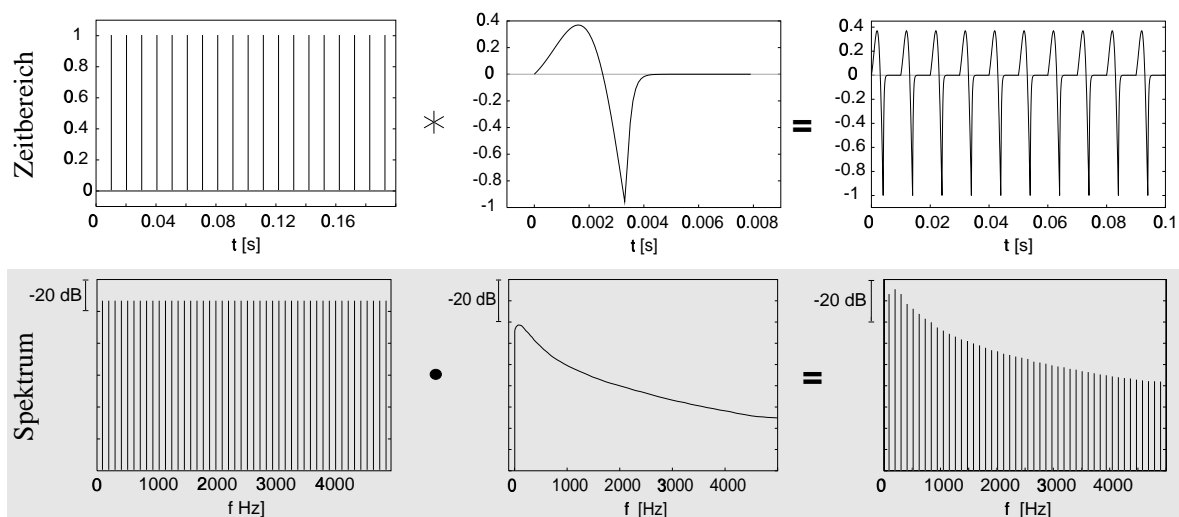


Abbildung 1.3: Die Erzeugung des Flußsignals (rechts) erfolgt im Zeitbereich (obere Reihe) durch Faltung einer δ -Pulsfolge (links) mit der Glottispulsform (mitte). Im Spektrum (untere Reihe) bedeutet dies eine multiplikative Zusammensetzung.

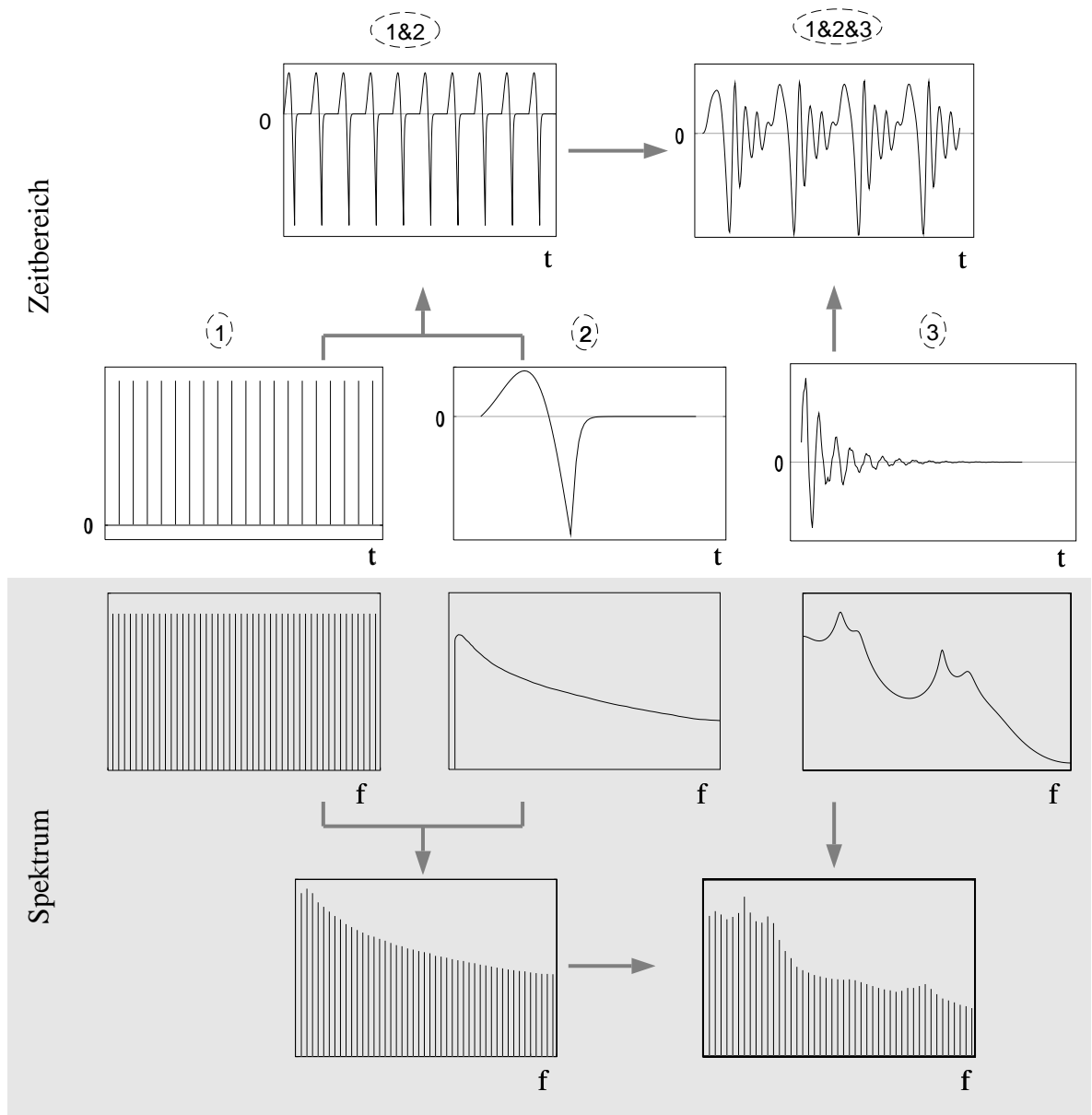


Abbildung 1.4: Das Quelle-Filter-Prinzip. Das Sprachsignal (1&2&3) entsteht durch Faltung des Glottissignals (1&2, siehe auch Abb. 1.3) mit der Impulsantwort (3) des Vokaltrakts. Die unteren, grau unterlegten Abbildungen zeigen spiegelsymmetrisch die Spektren der Signale.

nau bestimmt werden. Nur in diesem Fall darf der Vokaltrakt als angeregtes System mit schallhartem Abschluß betrachtet werden, das exponentiell gedämpft mit seiner Eigenfrequenz schwingt (Strube, 1974). Zur Anwendung der LPC-basierten Schätzung der Vokaltraktresonanzen muß daher eigentlich die genaue Kenntnis der Verschlußphase einerseits sowie ihre hinreichende Dauer andererseits vorausgesetzt werden, wobei letzterer Punkt insbesondere bei höheren Grundfrequenzen ein Problem darstellen kann. Lösungsansätze zur Bestimmung der Verschlußphase wurden sowohl rein auf der Grundlage des Zeitsignals vorgeschlagen (Strube, 1974; Ma et al., 1994), als auch unter Verwendung externer, synchronisierter Signale wie z.B. des EGGs (Childers und Ahn, 1995).

Als Alternative zur Bestimmung der Verschlußphase im Signal kann die theoretisch mögliche Exaktheit der Schätzung zugunsten einer robusteren Signalverarbeitung durch Integration über mehrere Glottiszyklen geopfert werden, da hierbei die Notwendigkeit der genauen Verschlußzeitbestimmung und die Anforderung an eine minimale Verschlußlänge entfallen (Alku, 1992). Des weiteren wurde gezeigt, daß aufgrund der Anpassung der Spektren an einer diskreten Anzahl spektraler Stützstellen die herkömmliche Methode der linearen Prädiktion durch das sogenannte *discrete all-pole modeling* (DAP) ersetzt werden sollte (El-Jaroudi und Makhoul, 1991). Andere Weiterentwicklungen betreffen die Verbesserung der Prädiktion durch stärkere Gewichtung des um genau einen Abtastwert zurückliegenden Signalwerts in der *separated linear prediction* (Varho und Alku, 1998). In der Sprachkomprimierung und -synthese stellte die Einführung der linearen Prädiktion unter Verwendung einer nichtlinearen Frequenzskala, die dem menschlichen Gehör besser angepaßt ist als die lineare Frequenzdarstellung, eine wichtige Neuerung dar (Strube, 1980; Karjalainen et al., 1998). Diese Idee der Anpassung der Frequenzskala an die Eigenschaften des menschlichen Gehörs wurde in der Sprachkodierung ebenfalls eingesetzt, um Quantisierungsrauschen in Bereiche zu transformieren, in denen es bei gegebenem Sprachsignal durch Maskierung weniger wahrgenommen wird (Atal und Schroeder, 1979).

Da selbst bei optimaler Bestimmung der Filterkoeffizienten die Inversfilterung auf dem Quelle-Filter-Ansatz basiert, kann sie nicht zur Berechnung des tatsächlichen glottalen Flusses führen, sondern liefert „einen Kompromiß“ (Fant, 1993; Childers und Ahn, 1995). Die genannten Probleme bei der Bestimmung der Filterkoeffizienten führen zu Schwierigkeiten bei der Validierung der Ergebnisse, die aus dem resultierenden inversgefilterten Signal (IF-Signal) gewonnen werden. Die automatische Inversfilterung auf der Grundlage von LPC oder DAP ergibt ein möglichst flaches Spektrum des Residuums, dem – bei korrekter Berücksichtigung der Phase – ein pulsartiges Signal im Zeitbereich entspricht. Aus diesem Signal lassen sich zwar Periodenlängen relativ zuverlässig ermitteln (Childers und Lee, 1991), doch sind weiterführende Interpretationen des Signals in bezug auf Eigenschaften der Stimmgebung nur eingeschränkt möglich (Eskenazi et al., 1990; Klasmeyer, 1998).

In der Literatur häufig verwendete „glottale“ Parameter wie Öffnungsquotient (*OQ*) oder Parameter zur Beschreibung der Asymmetrie des Pulses (*speed quotient*, *SQ*) lassen sich dagegen aus der semi-manuellen Inversfilterung bestimmen (Holmberg et al., 1994, 1995a,b; Hertegård und Gauffin, 1995; Sulter und Wit, 1996). Hierbei werden

durchstimmbare Antiresonanzfilter zur Auslöschung der ersten Formanten des Vokaltrakts interaktiv so eingestellt, daß das resultierende IF-Signal möglichst „gut aussieht“ (Schroeder, 1999, S. 94). Für dieses Signal können manuell oder durch empirische Festlegung von Schwellwerten verschiedene Phasen des Glottiszyklus bestimmt werden (Öffnungszeitpunkt, Verschußzeitpunkt, etc.), aus denen sich die Parameter OQ , SQ , usw. berechnen lassen. Die offensichtlichen Probleme bei diesem Vorgehen sind der zeitliche Aufwand der manuellen Bearbeitung jedes Glottiszyklus, der eine routinemäßige Anwendung ausschließt, und die Subjektivität der Ergebnisse, da in Ermangelung eines objektiven und definierten Kriteriums zur Einstellung der Filter ein gewisses Maß an Einarbeitung und Routine bei der durchführenden Person vorausgesetzt werden muß. Die Parametrisierungen, die sich bei verschiedenen Experten für dasselbe Signal ergaben, unterschieden sich in einer Studie um bis zu 10% (Fant, 1995). Diese Abhängigkeit der resultierenden Analysewerte von der die Analyse durchführenden Person sowie der mit einer manuellen Markierung verbundene Arbeitsaufwand sind Grund genug, nach einem automatisierten Verfahren zu suchen.

Die Schwierigkeiten in der Inversfilterung, die eine manuelle Bearbeitung erforderlich machen, bestehen somit einerseits in der automatischen Bestimmung der Koeffizienten des Filters, das den akustischen Effekt des Vokaltrakts optimal rückgängig macht, andererseits in der Bestimmung der Zeitmarken innerhalb eines Glottiszyklus, die der Berechnung aussagekräftiger glottaler Parameter zugrunde liegen. Price schrieb hierzu: „perhaps unscrambling an egg is easier“ (Price, 1989, S. 262). Eine Lösung dieses Problems ist das Ziel der vorliegenden Arbeit.

1.2 Stand der Forschung

Die Inversfilterung zur Bestimmung von Anregungspulsen geht auf Miller (1959) zurück, der den Einfluß des ersten Formanten auf ein Stimmsignal durch Anpassung elektrischer Schwingkreise (RCL-Glieder) auslöschte. Mathews et al. (1961) führte bereits eine digitale, grundperiodensynchrone Inversfilterung durch Anpassen eines ARMA-Modells (*auto regressive, moving average*) an das Leistungsdichtespektrum durch, wobei das AR-Modell die Resonanzeigenschaften des Vokaltrakts, das MA-Modell den glottalen Puls beschrieb. Jedoch waren viele manuelle Eingriffe nötig und die Autoren kamen zu dem Schluß, daß für diese Anwendung eine digitale Signalverarbeitung eigentlich zu aufwendig und zu kompliziert sei.

Nach solchen anfänglichen Schwierigkeiten in der Durchführung der Inversfilterung wurden in der Vergangenheit bereits mehrfach Versuche unternommen, die Aussagekraft von aus inversgefilterten Signalen (im folgenden kurz als *IF-Signale* bezeichnet) gewonnenen Parametern durch gleichzeitige oder nachfolgende Anpassung eines glottalen Modells zu erhöhen.

Eine recht frühe Publikation zur manuellen Inversfilterung und Anpassung eines Modells mit dem Ziel der Bestimmung von Pulsformen, die einen möglichst „guten“ Klang in Sprachsynthesatoren ermöglichen, stellt die Arbeit von Rosenberg (1970) dar. Nach interaktiver Anpassung von Antiresonanzfiltern zur Inversfilterung schlug er verschiedene Parametrisierungen der resultierenden Pulse vor, deren Eigenschaften bezüglich seiner Zielsetzung er durch systematische Hörtests der resynthetisierten Sprachproben überprüfte.

Ein ähnliches Ziel, nämlich Pulse zu bestimmen, die eine bessere Synthese insbesondere weiblicher Stimmen ermöglichen sollte, verfolgte Price (1989). Anstatt eines LPC-basierten Ansatzes schätzte er automatisch die Formanten in einem semi-manuellen, mehrstufigen Prozeß durch zunächst visuelle Bestimmung der Formanten aus mehreren Kurzzeitleistungsdichtespektren, die aus Signalabschnitten mit der höchsten Amplitude innerhalb einer Periode berechnet wurden. Eine anschließende Optimierung dieser Werte erfolgte durch sukzessive, selektive Auslöschung aller Formanten bis auf jeweils einen, der dann im Zeitbereich automatisch vermessen wurde. Daraufhin wurden interaktiv Zeitmarken gesetzt, die die Berechnung einiger abgeleiteter Parameter ermöglichten. Price stellte ein eigenes Glottismodell vor, das sich durch diese Parameter steuern ließ, doch wiesen die Ergebnisse innerhalb der männlichen und weiblichen Kategorien beträchtliche Streuungen auf und befanden sich teilweise im Widerspruch zu Arbeiten anderer Autoren. Die Steuerung des vorgeschlagenen Modells durch die berechneten Parameter wurde nicht mittels systematischer Perzeptionstests resynthetisierter Signale oder anderer, nachfolgender Untersuchungen überprüft.

Childers und Lee (1991) verwendeten eine zweistufige Inversfilterung, die im ersten Durchlauf mit fester Fensterlänge die Verschlußzeitpunkte als Spitzen (*spikes*) im Residuum lieferte, während der zweite, grundperiodensynchrone Durchlauf zur eigentlichen Schätzung der Antiresonanzen diente. Aus dem IF-Signal wurden verschiedene Parameter im Zeit- und Frequenzbereich teils manuell, teils automatisch extrahiert, anhand derer sich verschiedene Phonationsarten charakterisieren ließen. Ein um eine additive

Rauschanregung erweitertes Liljencrants-Fant (LF)-Modell wurde daraufhin in bezug auf die perzeptuelle Relevanz der einzelnen Modellparameter untersucht. Die Anpassung zur Beschreibung der verschiedenen Phonationsarten erfolgte auf diese Weise nur in generischem Sinn anhand der Ergebnisse der Perzeptionsstudie.

Childers und Ahn (1995) untersuchten drei Phonationsarten (modal, Strohmaß, behaucht) an drei Versuchspersonen mit einem zweistufigen Analyseverfahren. Die Inversfilterung wurde durch grundperiodensynchrone LPC-Analyse während der Verschlußphase realisiert, wobei die Bestimmung des Verschlußzeitpunktes mit Hilfe des Elektrolottograms (EGG) erfolgte. In einem manuellen Nachbearbeitungsschritt wurden einzelne Pole zurückgewiesen, wenn sie bestimmte heuristische Kriterien nicht erfüllten. Die Anpassung des LF-Modells zur Beschreibung des glottalen Flusses erfolgte im Zeitbereich mit der Begründung, daß „eine korrekte Anpassung im Zeitbereich automatisch zu einer korrekten spektralen Anpassung führt, was umgekehrt bei Anpassung des Leistungsdichtespektrums nicht zutrifft“ (Childers und Ahn, 1995, S. 515). Als Fehlerkriterium wurde der quadratische Abstand zwischen inversgefiltertem Signal und Referenzmodell minimiert. Die Bewertung der Ergebnisse erfolgte über die möglichst gute Vorhersage der perzeptiven Bewertungen, die für die untersuchten Stimmen vorlagen. Während aus prinzipieller Sicht Einwände gegen die Anpassung aller LF-Modellparameter im Zeitbereich erhoben werden können, stellt diese Methode dennoch eine Möglichkeit der manuellen Anpassung eines Glottismodells an das inversgefilterte Signal dar.

Fant (1993) schlug eine selektive (manuelle) Inversfilterung vor, bei der – ähnlich wie in Price (1989) – alle Formanten bis auf einen ausgelöscht werden. Durch Ausmessen der Eigenschaften des verbleibenden Formanten konnte die schwierige Schätzung des Parameters t_a im LF-Modell auf ca. 20% genau durchgeführt werden. Daneben gibt dieser Artikel einen ausführlichen Überblick über die Geschichte der Inversfilterung an der Königlich Technischen Hochschule Stockholm.

Alku und Vilkmann (1996) untersuchten verschiedene abgeleitete Parameter des inversgefilterten Signals für normale, gepreßte und behauchte Stimmen ohne Anpassung eines glottalen Modells. Basierend auf dem DAP-Verfahren wurde die Inversfilterung bei der relativ geringen Abtastrate von 4kHz durchgeführt. Die Prädiktionsordnung wurde in Abhängigkeit von der Phonationsart zwischen 8 und 12 gewählt, wonach die Parameterbestimmung durch interaktive Markensetzung im IF-Signal erfolgte. Zur Vermeidung von Markierungsfehlern und -ungenauigkeiten wurden die Parameter auf Grundlage der lokalen, über 4 Perioden berechneten Mittelwerte bestimmt.

Eine manuelle Anpassung an ein ebenfalls manuell inversgefiltertes Signal wurde von Gobl und Chasaide durchgeführt (Gobl und Chasaide, 1992). Hintergrund waren die Erfassung und Modellierung von Emotionen in fortlaufender Sprache, weshalb sich der untersuchte Sprachkorpus aus Äußerungen mit verschiedenen Phonationsarten zusammensetzte. Aufgrund des hohen Arbeitsaufwands wurden nur wenige Daten analysiert, so daß sich ihre Ergebnisse nicht notwendigerweise verallgemeinern lassen.

Neben diesen Arbeiten, denen allen eine manuelle Anpassung eines Glottismodells bzw. eine interaktive Bestimmung der Koeffizienten des inversen Filters zugrunde liegt, finden sich in der Literatur auch einige Versuche einer automatischen Modellanpassung, generell einhergehend mit einer automatischen Inversfilterung. Die gegenüberstellende

Untersuchung der Inversfilterung bei manueller Einstellung der Resonanzfilter und automatischer, LPC-basierter Inversfilterung von Södersten et al. (1999) ergab eine hohe Übereinstimmung zwischen den (in beiden Fällen manuell bestimmten) Parametern. Diese wurden aus den resultierenden IF-Signalen berechnet, wobei in Einzelfällen auch bei der automatischen Inversfilterung manuelle Eingriffe zur Stabilisierung der Ergebnisse vorgenommen wurden.

Eine automatische Methode zur Anpassung des LF-Modells an ein IF-Signal wurde von Strik et al. (1993) vorgestellt. Das IF-Signal wurde hierbei über Anpassung von Antiresonanzen nach Lage und Bandbreite mittels Dynamischer Programmierung (*dynamic programming*) ermittelt. Bis auf eine Ausnahme wurden alle LF-Parameter über Minimierung des mittleren Fehlerquadrats zwischen IF-Signal und angepaßtem Modell unüberwacht geschätzt. Die Anpassung des letzten, primär den spektralen Abfall betreffenden Parameters erfolgte im Spektralbereich, wobei nur der IF-Signalabschnitt während der Verschußphase zur Berechnung herangezogen wurde. Nachfolgende Optimierungsversuche ergaben, daß die Güte der Schätzung im wesentlichen durch die Initialisierung der Optimierungsmethoden festgelegt war und sich durch die multidimensionale Optimierung selbst kaum Verbesserungen erzielen ließen.

Milenkovic (1986) beschrieb ebenfalls eine automatische, simultane Inversfilterung und Anpassung eines glottalen Flußmodells. Die Berechnung des Fehlermaßes basierte auf dem Vorhersagefehler (Residuum), wobei das Ergebnis der Inversfilterung über die Differenz zwischen Signal und resynthetisiertem Signal (d.h. das Filter aus der Inversfilterung durch das Glottismodell angeregt) überprüft wurde. Das Glottismodell wurde hier allgemein als Superposition verschiedener Basisfunktionen, von denen eine die Öffnungsphase, eine andere die Verschußphase beschrieb, angesetzt und als „Kalmanformulierung des ARMA-Modells“ direkt in die Inversfilterungsalgorithmen inkorporiert. Die beiden Basisfunktionen des Glottismodells wurden als tiefpaßgefilterte Polynome vierter Ordnung angesetzt. Die Tests erfolgten sowohl mit synthetisierten Pulsfolgen als auch mit Stimmufnahmen bei verschiedenen Phonationsarten. Ziel der Untersuchung war aber nicht eine möglichst gute Differenzierung zwischen den verschiedenen Sprachproben, sondern eine Einschätzung bezüglich der Robustheit und Konsistenz des neuen Verfahrens in der Analyse echter Stimmen. Zur Verbesserung der Robustheit wurde die Initialisierung der Anpassung in einem Analysefenster mit den Ergebnissen des vorherigen Fensters vorgenommen. Der Rechenbedarf betrug damals 10 Sekunden pro analysierter Periode. Die gesamte Methode läßt sich schwer in einen Kontext einordnen, da alle Bestandteile neu konzipiert wurden und weiterführende Arbeiten über die Anwendung der Methode nicht bekannt sind.

Ein automatisches Anpassungsverfahren des Rosenberg-Klatt-Glottismodells an ein inversgefiltertes Signal wurde von Funaki et al. (1996) vorgestellt. Das System aus Vokaltrakt und Anregungsmodell wurde als „ARMA-Modell mit exogenem Input“ (ARMAX-Modell) angesetzt und über ein kombiniertes Verfahren aus *simulated annealing* (SA) und genetischen Algorithmen (GA) gelöst. Da leider keine genaueren Angaben zu Rechenbedarf (der bei SA- und GA-Algorithmen im allgemeinen vergleichsweise hoch ist) oder Anwendungsergebnisse gemacht wurden, kann die Leistungsfähigkeit des Verfahrens nicht beurteilt werden.

Neben diesen direkt zum Thema der vorliegenden Arbeit in Beziehung stehenden Publikationen sind noch einige weitere Veröffentlichungen zu erwähnen. So stellt die Arbeit von Titze (1984) eine theoretischere Arbeit zu dem Thema dar, wie sich mechanische Eigenschaften des Stimmlippensystems in aussagekräftigen Parametern widerspiegeln sollten, auf deren Grundlage sich verschiedene Phonationsarten unterscheiden ließen. Titze et al. (1983) hatten bereits im Jahr zuvor die Anpassung eines eigenen Glottismodells an das elektro- und photoglottographische Signal im Sinne des kleinsten Fehlerquadrats im Zeitbereich vorgestellt.

Eine wichtige Neuerung für die Durchführung der Inversfilterung stellte die Inversfilterung auf der Basis des Fluß- anstatt des Drucksignals dar, die durch Einführung einer speziellen Maske („Rothenbergmaske“) möglich wurde (Rothenberg, 1973). Die Maske besitzt zwar eine deutliche Tiefpaßcharakteristik und erschwert durch die Einführung neuer Resonanzen die Auswertung des aufgezeichneten Signals, bietet aber dafür den prinzipiellen Vorteil der absoluten Kalibrierbarkeit des inversgefilterten Signals. In etlichen neueren Studien wurde daher die flußbasierte Inversfilterung verwendet (Holmberg et al., 1994, 1995a,b; Hertegård und Gauffin, 1995; Sulter und Wit, 1996; Kobler et al., 1998).

Im Zusammenhang mit dem in dieser Arbeit beschriebenen Ansatz zur automatischen Anpassung eines Glottismodells ist ebenfalls die Arbeit von Skoglund (1998) von Interesse. Hierbei wurde zur Sprachkodierung eine *codebook*-orientierte Vektorquantisierung verschiedener Prototyppulse vorgeschlagen. Zur Generierung der Prototyppulse wurde eine Inversfilterung durchgeführt, bei der das glottale Spektrum durch eine iterative Präemphase in die Inversfilterung einging. Obwohl dem in dieser Arbeit vorgestellten Verfahren in der Idee nicht unähnlich, lieferte die Methode von Skoglund jedoch ein Signal, das „lediglich eine sehr grobe Abschätzung des glottalen Flusses“ darstellt (Skoglund, 1998, S. 128).

2. Methoden

2.1 Das LF-Modell

Zur Parametrisierung des glottalen Flusses existieren eine Vielzahl verschiedener Modelle (Rosenberg, 1970; Fant, 1979; Titze et al., 1983; Titze, 1989; Ananthapadmanabha, 1984; Milenkovic, 1986; Klatt und Klatt, 1990). Ein besonders wichtiges Modell stellt hierbei das Liljencrants-Fant (LF)-Modell (Fant et al., 1986) dar, das als eines der meistverwendeten Modelle bereits unter vielen verschiedenen Fragestellungen untersucht wurde (Gobl, 1988; Gobl und Chasaide, 1992; Childers und Lee, 1991; Childers und Ahn, 1995; Strik et al., 1993; Strik, 1994, 1996; Fant, 1993, 1995; Ma et al., 1994; Fröhlich und Strube, 1996; Karlsson und Liljencrants, 1996; Qi und Hillman, 1997; Veldhuis, 1998). In diesem Modell erlaubt die Beschreibung der glottalen Verschlussphase, Stimmen bzw. Phonationsarten zu charakterisieren, bei denen die Stimmlippen nicht simultan auf ihrer gesamten Länge schließen, sondern die einen graduellen oder sogar unvollständigen Verschluss aufweisen. Damit bietet sich die Verwendung des Modells auch zur Beschreibung bestimmter Klassen pathologischer Stimmen an.

Im LF-Modell wird die Ableitung $E(t) = \frac{dU}{dt}$ des glottalen Flusses $U(t)$ durch die Parameter t_p, t_e, t_a, E_e beschrieben (Abb. 2.1), wobei zur vollständigen Modellierung des Flusses noch die Periodenlänge $T_0(n)$ als freier Parameter – indiziert durch die Periodennummer n – hinzukommt. Die Ableitung $E(t)$ des glottalen Flusses wird in zwei Abschnitten modelliert: Zunächst wird $E(t)$ durch eine exponentiell anwachsende Sinusfunktion beschrieben („Offen-Phase“, *open phase*), während im zweiten Abschnitt

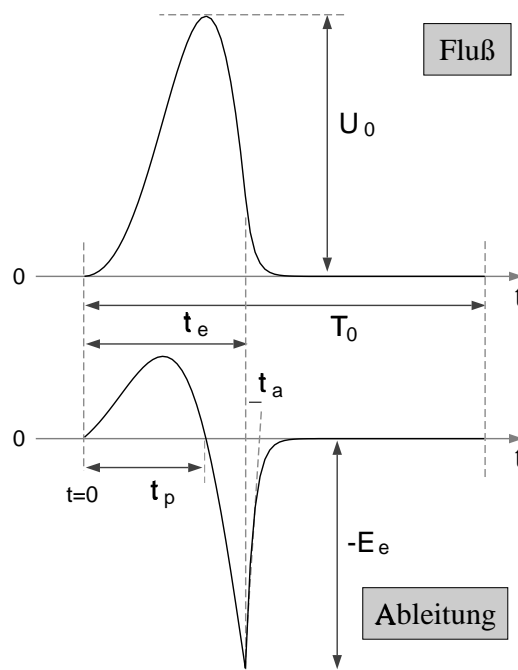


Abbildung 2.1: LF-Marken in einer Periode T_0 .

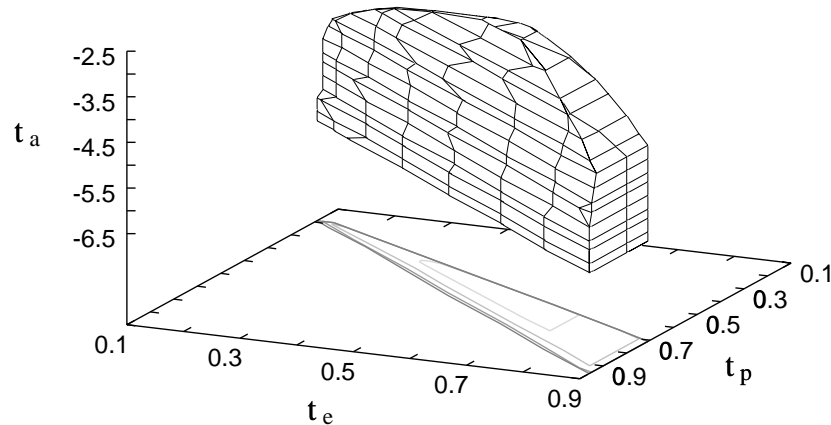


Abbildung 2.2: Der zulässige LF-Parameterraum für die Variablen t_p, t_e, t_a .

für $t > t_e$ („Verschlußphase“, *closed phase*)¹ $E(t)$ exponentiell auf 0 zurückgeht:

$$E(t) = \frac{dU(t)}{dt} = \begin{cases} E_0 e^{\alpha t} \sin \omega_g t, & 0 \leq t \leq t_e \\ -\frac{E_e}{\varepsilon t_a} [e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)}], & t_e < t < t_c. \end{cases} \quad (2.1)$$

Die Zeitmarken t_c, t_p, t_e, t_a werden bezüglich der Periodenlänge T_0 normiert und sind daher dimensionslos und auf den Wertebereich $[0, 1]$ beschränkt. Der Parameter t_c wird üblicherweise – so auch in dieser Arbeit – konstant als $t_c \equiv 1$ gesetzt.

Der Zusammenhang zwischen den verschiedenen Sätzen an unabhängigen Parametern, z.B. zwischen t_p, t_e, t_a, E_e und den abhängigen Größen α, E_0, ε in Gl. (2.1), ist nicht für alle Variablen geschlossen darstellbar und muß daher iterativ berechnet werden. Aufgrund der Gestalt von Gl. (2.1) können die Parameter t_p, t_e und t_a ihre Werte nicht unabhängig voneinander annehmen (Abb. 2.2). Bedingt durch die zeitliche Auflösung der Verschlußphase bei gegebener Abtastrate steigt zudem die untere Grenze der realisierbaren t_a geringfügig mit wachsender Grundfrequenz an (Abb. 2.3).

Eine von Fant (1995) vorgeschlagene Reduktion der Modellkomplexität bietet der Parameter R_d , der als eindimensionale Parametrisierung auf einfache Weise die Steuerung des Modells zwischen „hypofunktioneller“ (schwacher) und „hyperfunktioneller“ (gepreßter) Phonation ermöglicht. Ausgehend von statistischen Untersuchungen von Analyseergebnissen verschiedener Phonationsarten wurden Beziehungen angegeben, an-

¹Begrifflich zu unterscheiden ist die „Schließphase“ (*closing phase*), die die Zeitspanne zwischen t_p und t_e beschreibt.

hand derer sich die LF-Modellparameter aus R_d vorhersagen lassen:

$$t_a = \frac{-1 + 4.8R_d}{100} \quad (2.2)$$

$$t_p = \left(\frac{0.11R_d}{0.5 + 1.2R_{kp}} - t_a \right) \frac{2}{R_{kp}} \quad (2.3)$$

$$t_e = (R_{kp} + 1)t_p \quad (2.4)$$

mit

$$R_{kp} = \frac{22.4 + 11.8R_d}{100}. \quad (2.5)$$

R_d selbst wird aus dem Verhältnis des maximalen Flusses U_0 der Periode zu E_e bestimmt (f_0 : Grundfrequenz):

$$R_d = \frac{U_0}{E_e} \frac{f_0}{110\text{Hz}}. \quad (2.6)$$

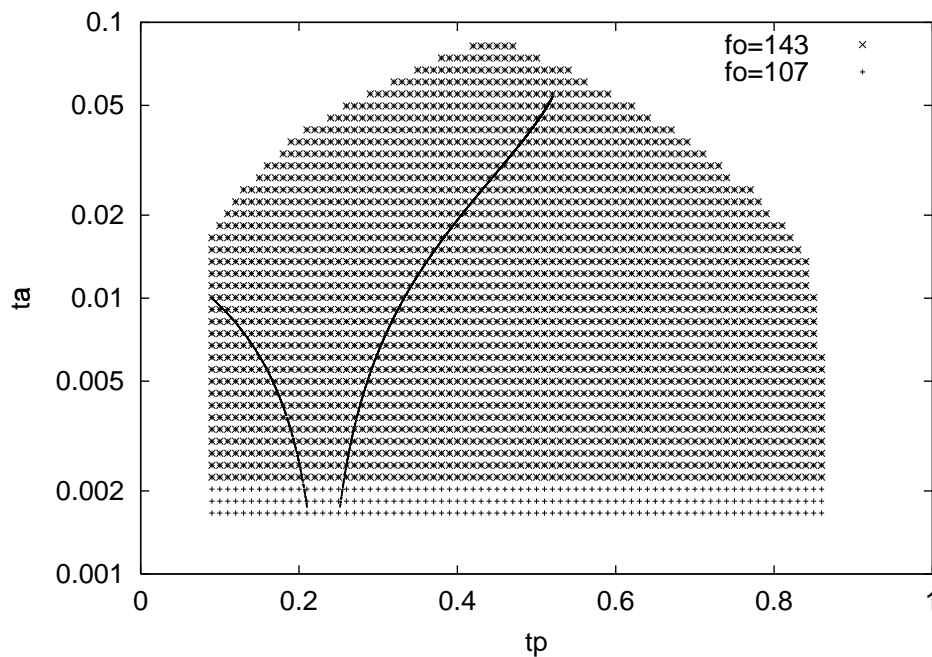


Abbildung 2.3: Abhängigkeit des zulässigen t_a -Bereichs von der Grundfrequenz. Dargestellt ist der $t_a t_p$ -Raum für $f_0 = 107\text{Hz}$ und $f_0 = 143\text{Hz}$. Die „v“-förmige Linie zeigt die durch R_d darstellbaren Parameterkombinationen ($f_0 = 107\text{Hz}$).

2.2 Das selbstschwingende Zweimassen-Modell

Soll der glottale Fluß nicht durch ein parametrisches Modell beschrieben werden, kann er über eine explizite Modellierung des laryngealen Schwingungssystems generiert werden. Die schwingungsmechanischen Eigenschaften des Stimmlippensystems lassen sich dabei in unterschiedlichen Vereinfachungen nachbilden.

Im simpelsten Fall („Einmassenmodell“) wird jede Stimmlippe als Federpendel mit denselben Eigenschaften in einer symmetrisch gegenüberliegenden Anordnung angesetzt (Flanagan und Landgraf, 1968). Diese Näherung läßt jedoch aufgrund der geringen Komplexität der Schwingungsmöglichkeiten keine weitreichende Modellierung der Stimmlippenschwingung zu.

Bei der Kopplung zweier Federpendel, die in Richtung des Luftstroms übereinander liegen, ist bereits eine viel bessere Modellierung möglich (Abb. 2.4). Da die Anzahl der freien Parameter zwar schon erheblich ist, jedoch noch übersichtlich bleibt (vgl. Tabelle 2.1), hat sich das „Zweimassenmodell“ nach Ishizaka und Flanagan (1972) als eines der weitverbreitetsten selbstschwingenden Stimmlippenmodelle etabliert.

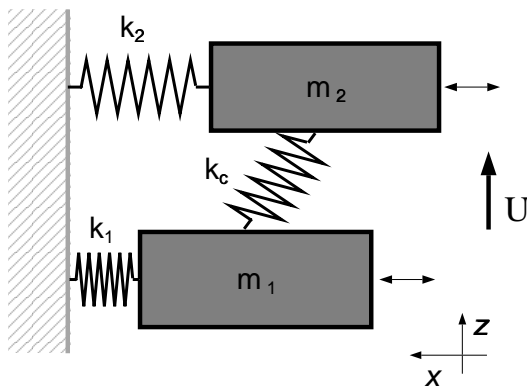


Abbildung 2.4: Das „selbstschwingende Zweimassenmodell“.

Zur verbesserten Modellierung spezieller Aspekte der Stimmgebung wurden einige Weiterentwicklungen vorgeschlagen. Durch asymmetrische Formulierung, d.h. unterschiedliche Wahl der mechanischen Konstanten für das linke und rechte Pendelsystem, konnten spezielle Phänomene insbesondere in der Analyse pathologischer Stimmen untersucht und modelliert werden (Ishizaka und Isshiki, 1976; Mergell und Herzel, 1997; Fröhlich et al., 1999). Die Ablösung des Luftstroms von den Wänden an einem variablen Ablösungspunkt, d.h. eine genauere Berücksichtigung

der strömungsmechanischen Eigenschaften, wurde von Pelorson et al. (1994) beschrieben. Eine ähnlich genaue Modellierung der strömungsmechanischen Eigenschaften bei gleichzeitiger Reduzierung der Anzahl der Modellparameter erfolgte in einem von Lous et al. (1998) vorgestellten Zweimassenmodell.

Ein Nachteil des Zweimassenmodells liegt in der Schwierigkeit bei der Zuordnung der Modellparameter bzw. Pendelmassen zur Histologie der Stimmlippen (Story und Titze, 1995). Insbesondere die *mucosal wave*, d.h. die vertikale Schleimhautwelle, die für die Energieübertragung zwischen Gewebe und Luftstrom verantwortlich ist, kann mit dem ursprünglichen Zweimassenmodell höchstens phänomenologisch nachgebildet werden. Daher wurde durch Hinzunahme einer weiteren Masse das „Dreimassenmodell“ eingeführt (Story und Titze, 1995), dessen Pendelmassen sich im Hinblick auf die Histologie der Stimmlippe interpretieren lassen, wodurch eine einfachere (patho-)physiologische Interpretation der Simulationsergebnisse möglich wird. Höherdimensionale

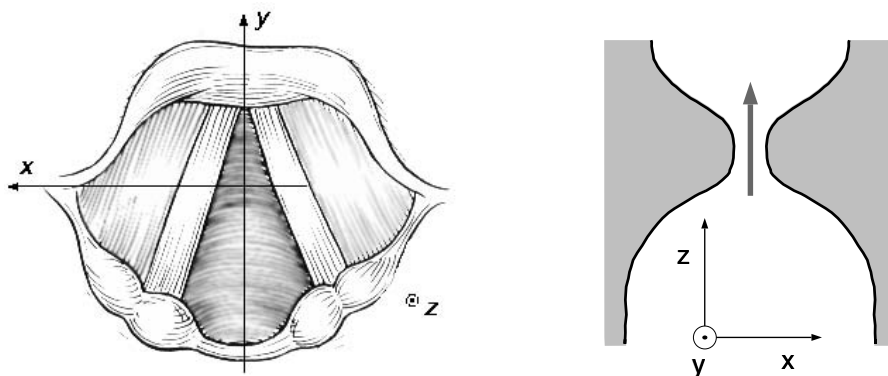


Abbildung 2.5: Orientierung des verwendeten Koordinatensystems. Links: Transversalschnitt (Sicht durch den Mund auf die Glottis), rechts: Frontalschnitt.

Beschreibungen, z.B. mittels finiter Elemente (Alipour und Titze, 1996) oder durch Kontinuumsmodelle (Berry und Titze, 1996) sind zur Beantwortung spezieller, prinzipieller Fragestellungen hilfreich, doch ist der Rechenaufwand enorm und die Validierung aufgrund der hohen Komplexität der Modelle, die mit der großen Anzahl an Freiheitsgraden einhergeht, schwierig durchzuführen.

Im ursprünglichen Zweimassenmodell sind die Pendelkörper durch quaderförmige Massen beschrieben, die allein in x -Richtung (Abb. 2.5) beweglich sind. Dies bedeutet zwangsläufig, daß die Glottis, d.h. die Öffnungsfläche zwischen den beiden Stimmlippen, im Modell einen rechteckigen Querschnitt besitzt, wobei die Glottislänge in y -Richtung einen konstanten Wert aufweist. Die Modellierung des glottalen Verschlusses, also die Unterbrechung des Luftstroms, erfolgt aus diesem Grund immer gleichzeitig auf der gesamten Länge der Massen, so daß im Prinzip der Luftstrom übergangslos von einem Abtastwert auf den nächsten abbricht. Damit besteht keine Möglichkeit, den Rückgang der Flußableitung zu Beginn der Verschlussphase zu beschreiben, die im LF-Modell durch t_a parametrisiert ist und die bei einem „weichen“ Schluß (z.B. bei hypofunktionaler Phonation) zu beobachten ist.

Um verschiedene Phonationsarten, insbesondere hypofunktionelle Phonation modellieren zu können, wurde das Modell erweitert, indem ausgehend von einer symmetrischen Anordnung der linken und rechten Stimmlippe ein Anstellwinkel θ eingeführt wurde (Abb. 2.6). θ beschreibt den Winkel zwischen yz -Ebene und der Pendelfläche, die dem Luftstrom ausgesetzt ist, so daß die geometrische Form des Pendels keinen Quader mehr darstellt, sondern ein Parallelepiped. Die Glottisfläche geht hiermit von einer rechteckigen in eine trapezförmige bzw. (bei teilweisem Verschluss) dreieckige Fläche über. Schließen jetzt die gegenüberliegenden linken und rechten Massen, so erfolgt die Unterbrechung des Luftstroms nicht simultan auf der ganzen Länge der Stimmlippe, sondern bei hohen y -Koordinaten zuerst. Die erhöhten Federkonstanten, die im ursprünglichen Modell bei Kontakt der Stimmlippen zur Erfassung der „Abrollbewegung“ aneinander eingeführt worden waren, beginnen nun, die starren Körper der beiden ge-

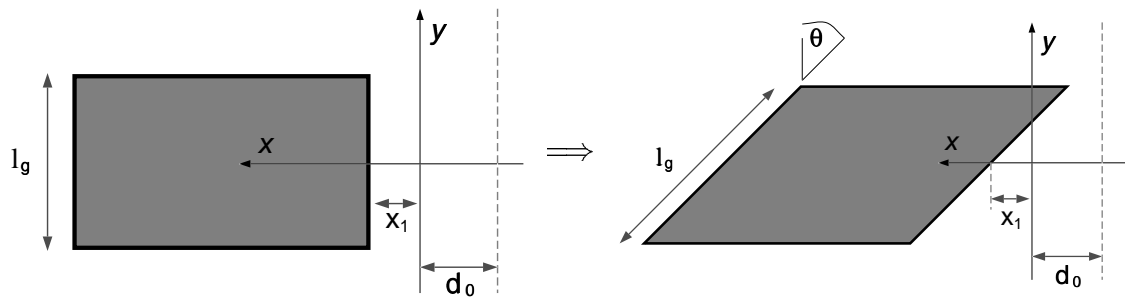


Abbildung 2.6: Übergang von rechteckiger Glottisfläche zu trapezförmiger/dreieckiger Fläche durch Einführung des Winkels θ . Dargestellt ist die „linke“ Masse, so daß die Schwingungsrichtung x nach links gerichtet ist.

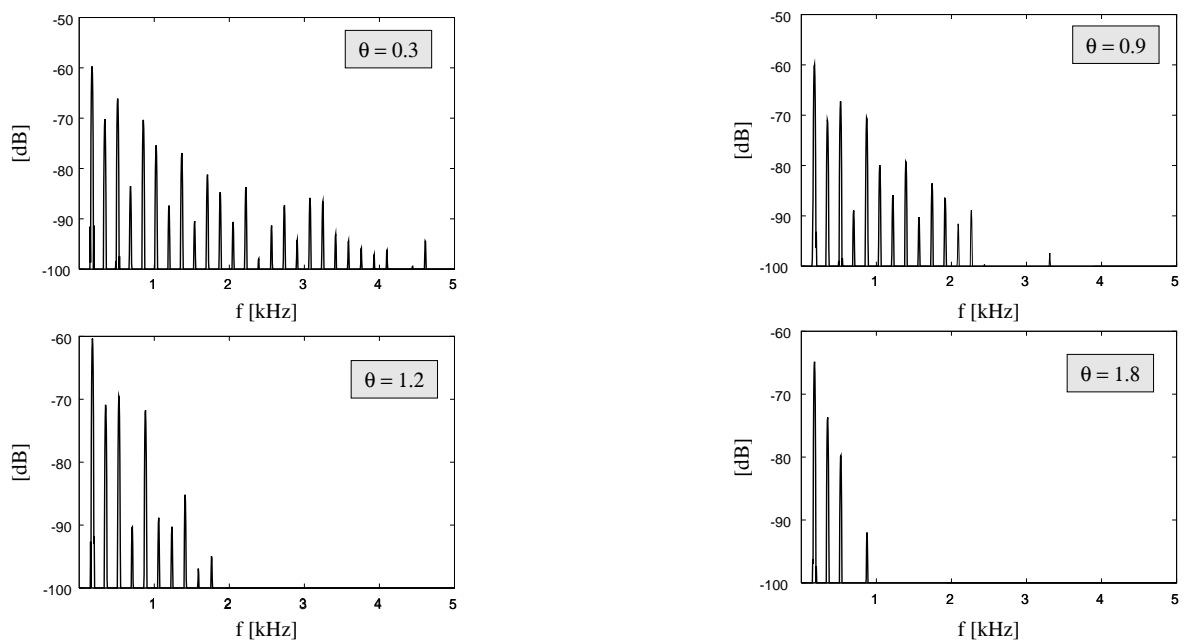


Abbildung 2.7: Leistungsdichtespektren des glottalen Flusses für den synthetisierten Vokal /a:/ in Abhängigkeit von θ (oben links: $\theta = 0.3^\circ$, oben rechts: $\theta = 0.9^\circ$, unten links: $\theta = 1.2^\circ$, unten rechts: $\theta = 1.8^\circ$).

genüberliegenden Pendel abzubremesen, wobei der Luftstrom jedoch noch nicht unterbrochen ist. Da die ursprünglichen Materialkonstanten beibehalten wurden, bei denen die Federkonstanten bei Verschluß um den Faktor 3 bis 5 erhöht werden (siehe Tabelle 2.1), genügen bereits geringe Werte von θ , um die glottale Fläche überhaupt nicht mehr auf Null zurückgehen zu lassen. Damit ist θ formal dem Adduktionswinkel zwischen den beiden Stimmlippen ähnlich, jedoch auf keinen Fall als solcher interpretierbar. Der erwünschte Effekt für die Modellierung verschiedener Phonationsarten, nämlich ein gradueller Verschluß mit resultierender spektraler Abschwächung bei langsamer Unterbrechung des Luftstroms, wird in Abbildung 2.7 deutlich. Die mathematische Formulierung der Erweiterung des Zweimassenmodells auf trapezförmige Glottisfläche erfolgt in Anhang A.

Tabelle 2.1: Verwendete Materialkonstanten für das selbstschwingende Zweimassenmodell.

Variable	Beschreibung	Wert	Einheit
a_{01}	Ruhe-Glottisöffnungsfläche des 1. Pendels	0.05	cm ²
a_{02}	Ruhe-Glottisöffnungsfläche des 2. Pendels	0.05	cm ²
m_1	Masse des 1. Pendels	0.125	g
m_2	Masse des 2. Pendels	0.025	g
d_1	Dicke des 1. Pendels	0.25	cm
d_2	Dicke des 2. Pendels	0.05	cm
l_g	Länge der Pendel	1.8	cm
k_1	lineare Federkonstante	80	
k_2	lineare Federkonstante	8	
k_c	lineare Kopplungsfederkonstante	50	
h_1	lineare Federkonstante	240	
h_2	lineare Federkonstante	24	
η_{h1}	nichtlineare Federkonstante	500	
η_{h2}	nichtlineare Federkonstante	500	
η_{k1}	nichtlineare Federkonstante	100	
η_{k2}	nichtlineare Federkonstante	100	
P	subglottaler Gleichdruck	8000	dyn/cm ²
ρ	Dichte der Luft	$1.14 \cdot 10^{-3}$	g/cm ³

2.3 Der Sprachsynthesator

Der zur Generierung einiger der untersuchten Testsignale verwendete Sprachsynthesator geht in seiner Formulierung auf Meyer (1987) und Meyer et al. (1989) zurück und gehört zur Klasse der artikulatorischen Synthesatoren im Zeitbereich. Hierbei wird die eindimensionale Schallausbreitung in einem abschnittsweise homogenen, schallharten Rohr berechnet, das sich aus 10 gleichlangen Segmenten der Länge 1.75cm zusammensetzt. An den Segmentgrenzen werden die Druckwellen teilweise reflektiert, wobei sich die Reflexionsfaktoren zwischen dem i -ten und $(i + 1)$ -ten Segment aus den Querschnittsflächen A_i, A_{i+1} gemäß

$$r_{i,i+1} = \frac{A_i - A_{i+1}}{A_i + A_{i+1}} \quad (2.7)$$

berechnen. Hierdurch bildet sich im Vokaltrakt ein charakteristisches Wellenfeld in Abhängigkeit von der Querschnittsflächenkonfiguration aus.² Während der Abschluß des Vokaltrakts glottisseitig als schallhart angenommen wird, erfolgt die Modellierung der Abstrahlung an den Lippen über eine trichterförmige Aufweitung der Querschnittsflächen in fünf Segmenten, die die Abnahme der charakteristischen Impedanz beim Übergang in das Freifeld nachbildet. Realisiert wird die Berechnung der Schallausbreitung durch Wellendigitalfilter (Fettweis und Meerkötter, 1975; Fettweis, 1986).

Zur Signalerzeugung wird das Vokaltraktmodell durch ein Modell des glottalen Flusses angeregt. Bei einer parametrischen Beschreibung des Flusses, z.B. durch das LF-Modell, wird das anregende Signal als hinlaufende Welle am glottisseitigen Abschluß des Vokaltrakts eingespeist. Bei Verwendung eines selbstschwingenden Stimmlippenmodells wird ein konstanter subglottaler Druck vorausgesetzt. Die Glottisöffnung wird als zeitlich veränderliche Impedanz mit einem resistiven und einem induktiven Anteil beschrieben, die den subglottalen Druck moduliert und deren Wert sich zu einem gegebenen Zeitpunkt aus der momentanen glottalen Öffnungsfläche berechnet (Ishizaka und Flanagan, 1972). Die auslenkende Kraft zur Anregung der Pendelmassen des Stimmlippenmodells ist dabei durch den Bernoullidruck

$$P_B = \frac{\rho}{2} \frac{|U_g|^2}{A_g^2} \quad (2.8)$$

gegeben (U_g : glottaler Fluß, A_g : glottale Fläche, ρ : Dichte der Luft, siehe Anhang A). Da U_g von der Eingangsimpedanz des darüberliegenden Vokaltrakts, d.h. von der geometrischen Konfiguration der Vokaltraktsegmente, abhängt, besteht eine Rückkopplung des Vokaltrakts auf die Schwingung des glottalen Modells. Diese Abweichung vom linearen Quelle-Filter-Prinzip ist bei der Sprachsynthese ein erwünschter Effekt, da sie zu einem natürlicheren Stimmeindruck führt. Die Form des glottalen Flusses weicht jedoch hierdurch von der parametrischen Beschreibung durch das LF-Modell deutlich ab (Abb. 2.8).

²Im Englischen wird diese Art von Synthesemodell *reflection type analog* genannt.

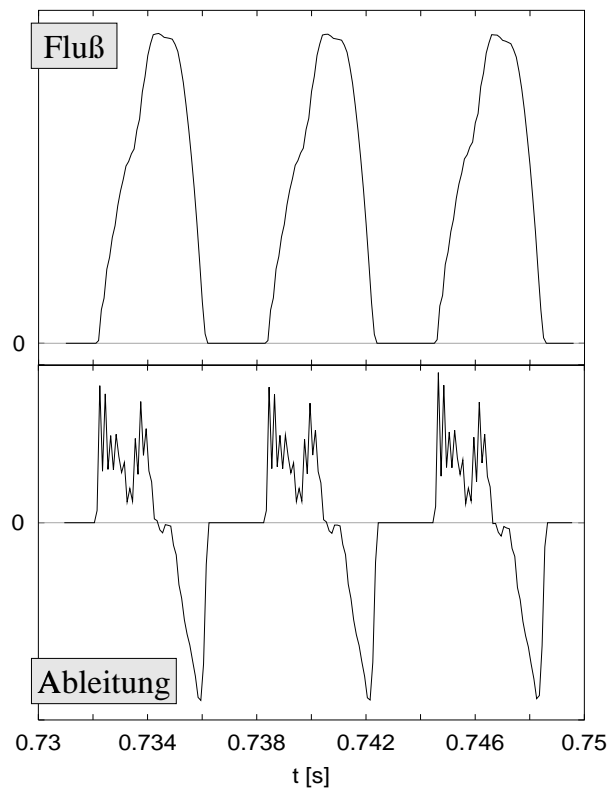


Abbildung 2.8: Fluß (oben) und Ableitung des Flusses (unten) bei Anregung des Vokaltraktmodells mit dem selbstschwingenden Zweimassenmodell.

2.4 Bestimmung der Periodenlängen

Viele akustische Stimmanalyseverfahren basieren auf einer akkuraten Periodenlängenbestimmung des Eingangssignals. Auch bei dem in dieser Arbeit vorgestellten Verfahren geht an mehreren Stellen die genaue Kenntnis der Periodenlänge ein (z.B. S. 47, 49). Eine Vielzahl an Methoden zur Bestimmung der Periodendauer sind in der Literatur beschrieben, die je nach Fragestellung und Rahmenbedingungen unterschiedliche Vorteile aufweisen (Schroeder, 1968, 1981; Hess, 1983; Milenkovic, 1987; Hess und Indefrey, 1987; Medan et al., 1991; Titze und Liang, 1992; Parsa und Jamieson, 1999).

Ein gängiges Verfahren beruht auf der *peak picking*-Methode, bei der Extremwerte in Signalabschnitten der erwarteten Periodenlänge automatisch detektiert und zur Markierungen der Periodengrenzen herangezogen werden (Abb. 2.9). Dieses Verfahren funktioniert zufriedenstellend bei normalen Stimmen und nicht zu hoher zeitlicher Auflösung und findet daher auch in gebräuchlichen kommerziellen Stimmanalysesystemen Verwendung (Kay, 1992). Probleme ergeben sich bei dieser Methode, wenn die Stimmanregung teilweise durch Rauschen erfolgt, d.h. insbesondere bei behauchten Stimmen. Dies liegt daran, daß durch die Methode ein einzelnes Ereignis (*event*) in einer Periode bestimmt und der Periodenlängendefinition zugrundegelegt wird, sich dieses Ereignis jedoch durch additives Rauschen in seiner zeitlichen Lage um einige Abtastwerte verschieben kann. Für pathologische Stimmen sind daher das *peak picking*- oder andere ereignisbasierte Verfahren (z.B. durch Bestimmung der Nulldurchgangsrate) nur unter Vorbehalt anzuwenden.

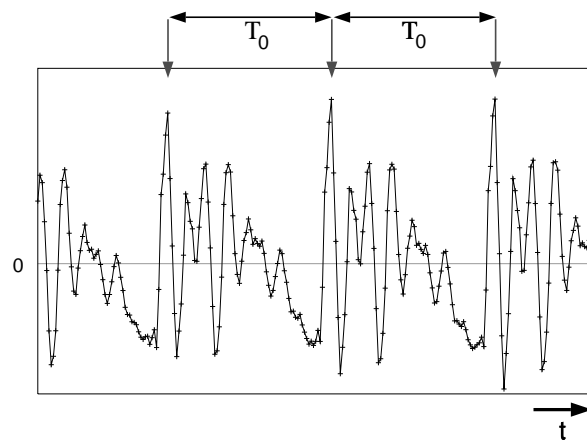


Abbildung 2.9: Periodenlängenextraktion nach dem *peak picking*-Verfahren. Der Abstand zwischen zwei Extremwerten definiert die momentane Periodenlänge T_0 .

Eine weit höhere Robustheit bezüglich additivem Rauschen bieten sogenannte integrale Ansätze wie z.B. das *waveform matching*-Verfahren (Titze und Liang, 1992; Parsa und Jamieson, 1999). Bei diesem Ansatz, der bereits von Milenkovic (1987) vorgeschlagen wurde (jedoch noch nicht unter diesem Namen), dient das Maximum der

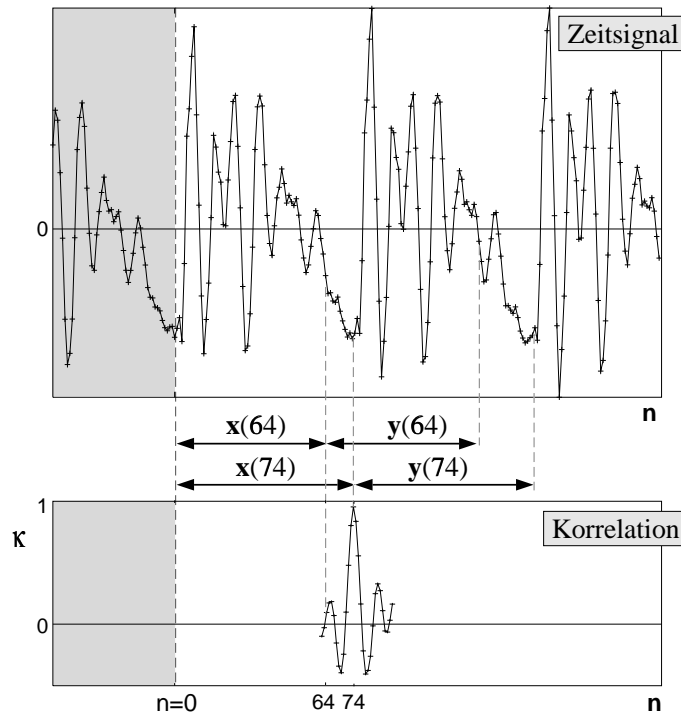


Abbildung 2.10: Periodenlängenextraktion nach dem *waveform matching*-Verfahren. Die Lage des Maximums der Kreuzkorrelation κ definiert die Periodenlänge (hier: $n = 74$).

Kreuzkorrelation κ zwischen aufeinanderfolgenden Signalstücken variabler Länge zur Definition der Periode. Die beiden miteinander zu korrelierenden Signalabschnitte \mathbf{x}, \mathbf{y} der Länge n werden bei festgehaltenem Startpunkt n_0 in einem vorgegebenem Intervall variiert (Abb. 2.10):

$$\kappa(n) = \frac{\mathbf{x}(n) \cdot \mathbf{y}(n)}{|\mathbf{x}(n)| |\mathbf{y}(n)|} \quad (2.9)$$

mit

$$\begin{aligned} \mathbf{x}(n) &= (s(n_0), s(n_0 + 1), \dots, s(n_0 + n - 1))^T \\ \mathbf{y}(n) &= (s(n_0 + n), s(n_0 + n + 1), \dots, s(n_0 + 2n - 1))^T. \end{aligned}$$

Die Periodenlänge n_{wm} (in Abtastwerten) wird als diejenige Zeitverzögerung n definiert, bei der $\kappa(n)$ maximal wird:

$$n_{\text{wm}} = \left\{ n \mid \max_{n_{\text{min}} \leq n \leq n_{\text{max}}} (\kappa(n)) \right\}. \quad (2.10)$$

Zur Umrechnung von Abtastwerten auf dimensionsbehaftete Zeiten wird n_{wm} auf die Abtastfrequenz f_s bezogen, so daß sich $T_0 = n_{\text{wm}}/f_s$ ergibt. Der Zeitpunkt $n_0 + n_{\text{wm}}$ wird als Startpunkt zur Bestimmung der darauffolgenden Periode gesetzt:

$$n_0 \longrightarrow n_0 + n_{\text{wm}} . \quad (2.11)$$

Zur vollautomatischen Anwendung des Verfahrens muß der Suchbereich $[n_{\text{min}}, n_{\text{max}}]$ bzw. $[T_{\text{min}}, T_{\text{max}}]$ sinnvoll gewählt werden. Bei der automatischen Analyse der Stimmen in einem unbekanntem Stimmkorpus kann dies anhand der Zeitverzögerung T_{akf} des ersten Autokorrelationsmaximums für Fenster der Länge 200-500ms geschehen. Die Suchgrenzen werden dann als $T_{\text{min}} = 0.5 \cdot T_{\text{akf}}$ und $T_{\text{max}} = 1.5 \cdot T_{\text{akf}}$ gesetzt (Michaelis et al., 1998a; Fröhlich et al., 2000). Für die in dieser Studie analysierten Signale liegt der mögliche Frequenzbereich zwischen 80Hz und 200Hz. Daher wurde der Suchbereich $[T_{\text{min}}, T_{\text{max}}]$ fest auf [5ms, 13ms] gesetzt.

Die Beschreibung der Periodenlängen in der Auflösung von ganzzahligen Abtastwerten ist für viele Anwendungen unzureichend. Daher wurde in Anlehnung an Medan et al. (1991) die Auflösung durch parabolische Interpolation zwischen den Korrelationswerten κ an den Stellen $n_{\text{wm}} - 1, n_{\text{wm}}, n_{\text{wm}} + 1$ auf nicht-ganzzahlige Abtastwerte erhöht. Aufgrund dieser Interpolation ist die Summe von k aufeinanderfolgenden Periodenlängen im allgemeinen nicht identisch zur Differenz des Endwerts der k -ten Periode und dem Startwert der ersten Periode, sondern kann kleine Abweichungen davon aufweisen.

2.5 Verwendete Testsignale

Bei Einführung eines neuen Verfahrens müssen seine Eigenschaften zur Einschätzung der Anwendbarkeit im Hinblick auf die Konsistenz der numerischen Ergebnisse („Reliabilität“) und der Interpretationen („Validität“) untersucht werden. Bei einem neuen Schätzverfahren zur Ermittlung glottaler Parameter gestattet die Analyse einfacher Signale die prinzipielle Beurteilung der Methode, doch können aufgrund der getroffenen Vereinfachungen nur eingeschränkt Rückschlüsse auf die Anwendung für die Analyse echter Stimmen erfolgen. Kompliziertere Testsignale bilden das echte Stimmsignal unter realistischeren Annahmen nach, doch bleiben die Probleme hinsichtlich der Verallgemeinerbarkeit bestehen. Echte Stimmsignale bieten die Möglichkeit, Ergebnisse *post hoc* zu überprüfen, bieten aber den Nachteil, daß spezielle Eigenschaften der Analysemethode aufgrund der Komplexität des Phonationsvorgangs nur schwer isoliert untersucht werden können. Einzelne Aspekte der Analysemethode im Hinblick auf ihre Interpretationsmöglichkeiten lassen sich jedoch durch die Analyse verschiedener Phonationsarten näher beleuchten.

Ausgehend von diesen prinzipiellen Überlegungen wurden verschiedene Testsignale definiert und untersucht (eine Übersicht ist in Tabelle 2.2 dargestellt). Der einfachste Signaltyp realisiert das Quelle-Filter-Prinzip (vgl. Abb. 1.4): Eine Sequenz von LF-Pulsen wird mit einem Nulpole-Modell 11. Ordnung, das die Formanten des Vokals /a:/ nachbildet, gefiltert. Dieser Signaltyp wird als FLT (Abkürzung für „Filter“) bezeichnet (Abb. 2.11). Dabei wird eine konstante Grundfrequenz vorausgesetzt, die zur Vermeidung von systematischen Fehlern in der Berechnung der LF-Pulse in einem „möglichst irrationalen“ Verhältnis³ zur generell verwendeten Abtastfrequenz $f_s = 10\text{kHz}$ stehen sollte und als $f_0 = 107\text{Hz}$ gewählt wurde. Die Analyse dieser Art von Testsignalen ermöglicht die Bewertung grundlegender Eigenschaften der Methode wie Streuung der Ergebnisse und Konvergenzverhalten. Auf diese Weise kann eine Referenz (*baseline*)

³Die soll bedeuten, daß Grundfrequenz und Abtastfrequenz ein möglichst großes „kleinstes gemeinsames Vielfaches“ besitzen.

Tabelle 2.2: Übersicht über die untersuchten Signalarten.

FLT	LF-Pulse mit Nulpole-Filter gefiltert
SYN	Vokaltraktmodell (Synthesator) mit LF-Pulsen angeregt
2MMS	Vokaltraktmodell (Synthesator) mit Fluß aus selbstschwingendem Stimmlippenmodell angeregt
2MMF	Fluß aus selbstschwingendem Stimmlippenmodell mit Nulpole-Filter gefiltert
ECHT	Verschiedene Phonationsarten bei echten Stimmen

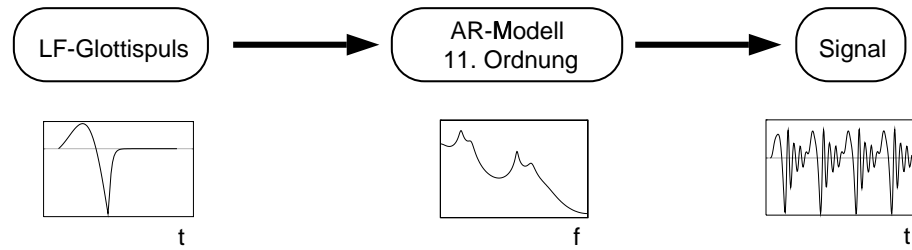


Abbildung 2.11: Signaltyp FLT.

bestimmt werden, bezüglich der die Ergebnisse bei komplexeren Testsignalen gedeutet werden können.

Der nächstkompliziertere Signaltyp ergibt sich durch Anregung des Sprachsynthetisators (Abschnitt 2.3) mit einem durch das LF-Modell parametrisierten Fluß (Abb. 2.12). Dieser als SYN (Abkürzung für „Synthetisator“) bezeichnete Signaltyp entspricht immer noch dem Quelle-Filter-Prinzip im Sinne der linearen Separierbarkeit zwischen Anregung und nachfolgender Filterung des Signals, bedeutet aber das Verlassen der Nulpole-Näherung für die Übertragungsfunktion des Vokaltrakts und der Differenzierung zur näherungsweise Beschreibung der Lippenabstrahlung. Die Übertragungsfunktion des supraglottalen Systems ist nicht mehr durch ein einfaches rekursives Filter realisierbar, sondern wird durch die explizite Simulation der Wellenausbreitung im Vokaltrakt, die insbesondere eine Modellierung der Dämpfung beinhaltet, ersetzt. Diese Art von Testsignalen wurde ebenfalls von Strik et al. (1993) in der Anpassung des LF-Modells an das IF-Signal eingesetzt.

Als nächste Näherung wird die lineare Separierbarkeit zwischen Quelle und supraglottalem System fallengelassen. Dies geschieht bei zwei Signalarten, bei denen der glottale Fluß mit Hilfe des selbstschwingenden Zweimassenmodells (Abschnitt 2.2) generiert wird. Das Schwingungsmuster der Stimmlippen und damit auch die Gestalt

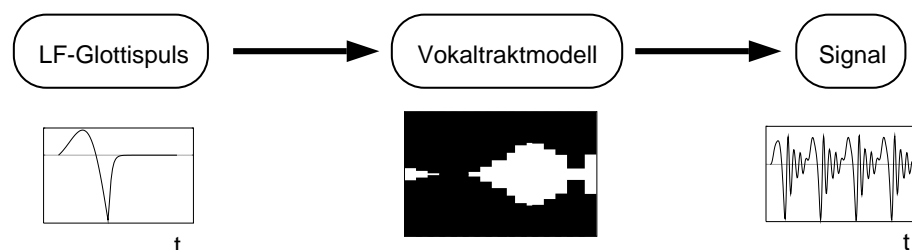


Abbildung 2.12: Signaltyp SYN.

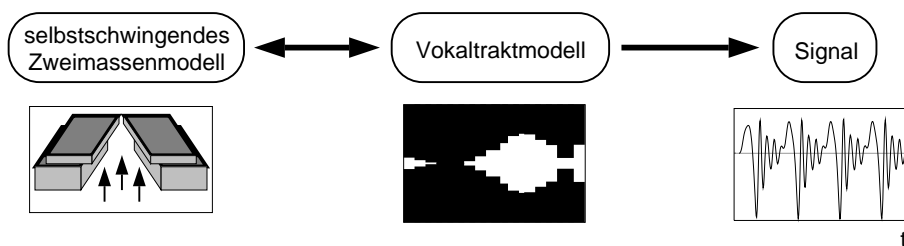


Abbildung 2.13: Signaltyp 2MMS.

des glottalen Flusses wird durch das Druckprofil innerhalb des supraglottalen Systems beeinflusst, d.h. der glottale Fluß hängt nun von der akustischen Impedanz des supraglottalen Systems (von der Glottis aus „gesehen“) ab. Bei Signaltyp 2MMS (Abkürzung für „Zweimassenmodell, Synthetisator“) dient das Zweimassenmodell mit trapezförmiger Glottisfläche zur Anregung des Vokaltraktmodells (Abb. 2.13). Generell werden bei diesem Signaltyp nur Signalteile analysiert, nachdem sich eine stabile Schwingung eingestellt hat – in anderen Worten, die stationäre Schwingung ohne Einschwingvorgänge. Diese Art von Testsignalen wurde ebenfalls in anderen Arbeiten zur Anpassung eines Glottismodells an das IF-Signal untersucht (Milenkovic, 1986; Strik et al., 1993).

Da die Übertragungsfunktion des Synthetisators, wie bereits bei Signaltyp SYN erläutert, nicht durch ein AR-Modell beschrieben werden kann, wird für den Signaltyp

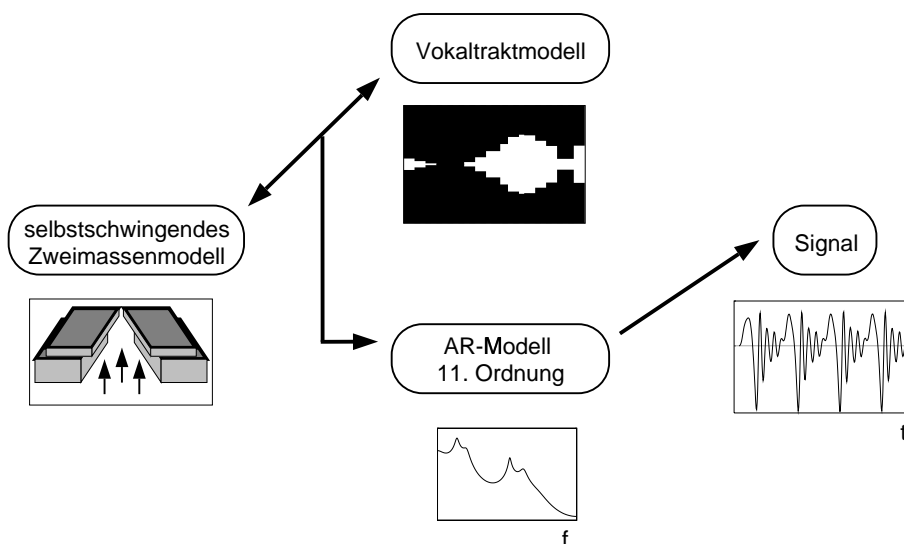


Abbildung 2.14: Signaltyp 2MMF.

2MMF (Abkürzung für „Zweimassenmodell, Filter“) das Zweimassenmodell zunächst dem Synthetisator wie bei Signaltyp 2MMS vorgeschaltet und der sich einstellende glottale Fluß für den Vokal /a:/ ermittelt. Dieses Flußsignal wird danach mit einem Nulpole-Filter 11. Ordnung, das die Formantstruktur dieses Vokals approximiert, gefiltert (Abb. 2.14). Auf diese Weise entsteht ein Anregungssignal, das bei der Erzeugung keine lineare Separierbarkeit zwischen Anregung und Filterung aufweist, jedoch ein Testsignal ergibt, bei dem prinzipiell die Resonanzen des Nulpole-Filters exakt durch das inverse Filter eliminiert werden können.

Der komplexeste Signaltyp wird mit ECHT bezeichnet. Hierbei phonierten drei verschiedene männliche Sprecher mit gesunden Stimmen die Vokale /ε:/ und /a:/ je zweimal in den Phonationsarten „normal“, „gepreßt (hyperfunktionell)“, „schwach (hypofunktionell)“ und „laut“ (Abb. 2.15). Von einem Sprecher (Sprecher 1) wurden insgesamt drei Aufnahmen im Abstand von einigen Wochen durchgeführt, so daß insgesamt fünf verschiedene Aufnahmesätze vorlagen. Sprecher 1 führte ebenfalls an zwei Aufnahmesitzungen einen kontinuierlichen Übergang von hypofunktioneller zu gepreßter Stimmgebung und wieder zurück je zweimal für beide Vokale durch. Die Aufnahmen auf DAT (*digital audio tape*) erfolgten bei der Abtastfrequenz $f_s = 48\text{kHz}$ mit einem Recorder Pioneer D-07 in einem reflexionsfreien Raum unter Verwendung eines Sennheiser-Mikrofons (MKH 105) und eines speziell dafür eingerichteten Vorverstärkers. Der Abstand zwischen Mikrophon und Testsprecher betrug ca. 2m. Zur weiteren Analyse wurden die Signale auf die generell für alle Analysen verwendete Abtastrate $f_s = 10\text{kHz}$ gebracht.

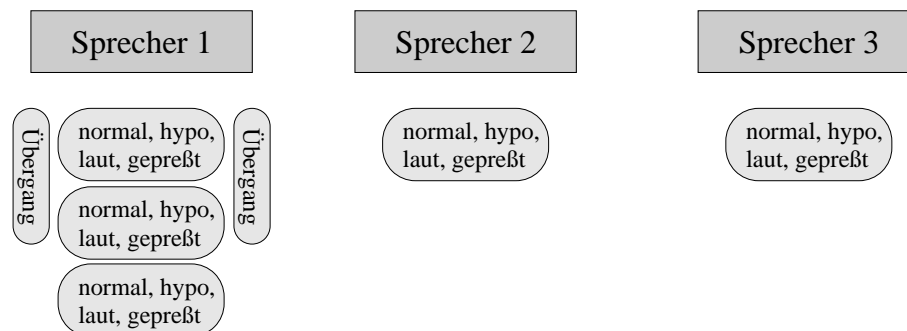


Abbildung 2.15: Signaltyp ECHT.

2.6 Exkurs: Bestimmung des Referenzmodells bei Signalen 2MMS und 2MMF

Bei der Verwendung des Zweimassenmodells zur Generierung des anregenden Flusses (Signaltyp 2MMS und 2MMF) existiert keine definierte Referenzkonfiguration des LF-Modells, anhand derer die Resultate der Schätzung überprüft werden könnten. Da der anregende glottale Fluß jedoch bekannt ist, kann an ihn direkt ein LF-Modell angepaßt werden, was allerdings einen geeigneten Anpassungsalgorithmus voraussetzt. Versuche, die Anpassung analog zu den Methoden der Inversfilterung (Abschnitt 2.10) unter Verwendung des direkt aus Eingangssignal (Flußsignal) und angesetzttem LF-Modell berechneten Itakura-Saito-Fehlers vorzunehmen, führten jedoch aufgrund der relativ unnatürlichen Gestalt des Flusses (vgl. Abb. 2.8) zu keinem zufriedenstellenden Ergebnis.

Das glottale Flußsignal weist jedoch Eigenschaften auf, die eine Anpassung im Zeitbereich erlauben: Die Schließphase (zwischen t_p und t_e , vgl. Abb. 2.1) sowie die „Rückkehrphase“ (nach t_e) zeigten sich weitgehend ungestört vom Einfluß höherer Formanten. Daher wurde das LF-Modell in Anlehnung an ein von Fröhlich und Strube (1996) beschriebenes Verfahren im Zeitbereich an das Flußsignal angepaßt. Der Algorithmus ist in Abbildung 2.16 beschrieben.

Zeitbereichsanpassung

1. Suche Minimum des Signals und bilde Betrag: E_e .
2. Zeitpunkt des Minimums (noch ohne Bezug zu $t = 0$): t_e .
3. Finde Nulldurchgang vor t_e (falls nötig mit *wrap around* der Periode, lineare Interpolation zwischen den eingrenzenden, benachbarten Abtastwerten): t_p .
4. Finde ausgehend vom linksseitigen Abtastwert vor t_p den Abtastwert, bei dem das Signal unter $10^{-5}E_e$ gefallen ist (falls nötig mit *wrap around* der Periode): $t = 0$.
5. Bestimme den Signalabschnitt, der bei t_e beginnt und an dem nächsten Punkt endet, der einen Betrag kleiner als $10^{-5}E_e$ aufweist. Passe eine Gerade $y = a + bx$ an die logarithmierten Beträge in diesem Abschnitt an. Bestimme t_a aus seiner Definition als $t_a = -E_e (b e^a)$.

Abbildung 2.16: Algorithmus der Anpassung des LF-Modells im Zeitbereich.

2.7 Parameter zur Charakterisierung des glottalen Flusses

Die vier Parameter des LF-Modells t_p, t_e, t_a, E_e (siehe Abschnitt 2.1) bieten die Möglichkeit, Eigenschaften des glottalen Flusses zu quantifizieren, und stellen damit eine Grundlage zur Differenzierung zwischen verschiedenen Phonationsarten dar. Es ist jedoch bekannt, daß die dem LF-Modell zugrundeliegende Dimensionalität geringer als vier ist (Veldhuis, 1998) und daß sogar ein einziger Parameter R_d (siehe Gl. (2.2)-(2.4)) innerhalb gewisser Grenzen zur Steuerung des Modells herangezogen werden kann (Fant, 1995).

Neben den direkten Modellparametern haben sich einige Größen zur Charakterisierung des Phonationsvorgangs in der Literatur als aussagekräftig herausgestellt und wurden daher auch in dieser Arbeit als sogenannte „abgeleitete Parameter“ untersucht. Die meisten dieser Größen setzen charakteristische Zeitdauern innerhalb des glottalen Zyklus zueinander in Beziehung. Auf diese Weise werden die Parameter „Öffnungsquotient“, „Asymmetriequotient“ und „Schließquotient“ berechnet (englisch: *open quotient*, *speed quotient*, *closing quotient*). Ein weiterer Parameter, der die tieffrequenten Eigenschaften der glottalen Pulsform quantifiziert, ist der *parabolic spectral parameter*. Diese Parameter werden in den folgenden Abschnitten genauer erläutert. Eine Zusammenstellung aller untersuchten „abgeleiteten Parameter“ findet sich in Tabelle 2.3.

Tabelle 2.3: Überblick über alle verwendeten „abgeleiteten“ Parameter und ihre Berechnung aus den ursprünglichen LF-Parametern (vgl. Abb. 2.1 und Abb. 2.18).

Öffnungsquotient	$OQ_1 = t_e - t_s$
Öffnungsquotient	$OQ_2 = t_e + t_a - t_s$
Schließquotient	$CQ_1 = t_e - t_p$
Schließquotient	$CQ_2 = t_e + t_a - t_p$
Asymmetriequotient	$SQ_1 = \frac{OQ_1 - CQ_1}{CQ_1}$
Asymmetriequotient	$SQ_2 = \frac{OQ_2 - CQ_2}{CQ_2}$
<i>parabolic spectral parameter</i>	$PSP =$ Parametrisierung der tieffrequenten spektralen Kontur nach Alku et al. (1997)

2.7.1 Der Öffnungsquotient (OQ)

Das Verhältnis der Dauer der offenen Phase, während der der glottale Fluß nicht verschwindet, zur Periodenlänge ist eine in der Literatur oft anzutreffende Größe (Titze et al., 1983; Price, 1989; Arends et al., 1990; Klatt und Klatt, 1990; Childers und Lee, 1991; Childers und Ahn, 1995; Gobl und Chasaide, 1992; Fant, 1993; Holmberg et al., 1995b; Hertegård und Gauffin, 1995; Alku und Vilkmán, 1995, 1996; Sulter und Wit, 1996; Veldhuis, 1998). In der vorliegenden Arbeit soll diese Größe in Übereinstimmung mit den meisten der zitierten Publikationen als „Öffnungsquotient“ (*open quotient, OQ*) bezeichnet werden.

Wird der *OQ* aus dem Residuum der Inversfilterung berechnet, stellt die Bestimmung des Verschlusszeitpunktes generell kein Problem dar. Dagegen läßt sich der Beginn der Öffnungsphase im IF-Signal oft nicht exakt ablesen (Abb. 2.17), da die Restwelligkeit des Signals während der Verschlussphase von der Güte der Inversfilterung und damit vom Zusammenspiel der Phase des Signals und der Übertragungsfunktion des inversen Filters abhängt. Die *OQ*-Bestimmung erscheint bei diesem Vorgehen somit zumindest fragwürdig. Um eine robustere Schätzung zu erhalten, wurde in manchen Arbeiten das elektroglossographische Signal (EGG) als zusätzliche Informationsquelle herangezogen (Milenkovic, 1986).

Dagegen ist der Öffnungsquotient im angepaßten LF-Modell im Prinzip direkt als $OQ = t_e$ gegeben. Diese unmittelbare Entsprechung wurde in der Literatur durchaus als

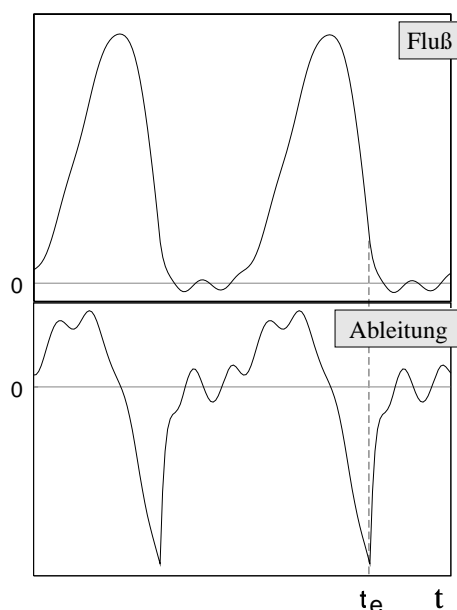


Abbildung 2.17: Fluß und Ableitung des Flusses nach Inversfilterung. Der Öffnungszeitpunkt ist aus den Signalen nicht zuverlässig bestimmbar.

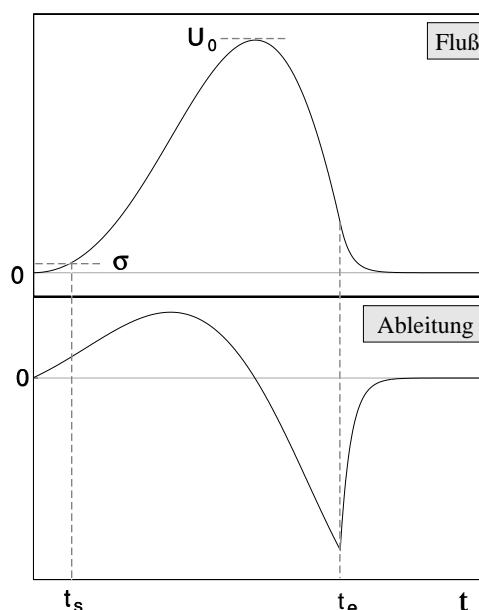


Abbildung 2.18: Fluß und Ableitung des Flusses beschrieben durch das LF-Modell. Der Öffnungszeitpunkt t_s wird durch eine Schwelle σ festgelegt.

Parameter verwendet (Gobl und Chasaide, 1992). Für große t_e ist jedoch $E(t)$ nach Gl. (2.1) zu Beginn der Periode vernachlässigbar klein, so daß ein Schwellwert σ angesetzt wird, ab dem die Glottis als offen zu betrachten ist (Abb. 2.18). Der Öffnungsquotient OQ ergibt sich dann als

$$OQ_1(\sigma) = t_e - t_s(\sigma). \quad (2.12)$$

In Anlehnung an die Literatur (Fant, 1995; Childers und Ahn, 1995) wurde alternativ der Öffnungsquotient von demselben Startpunkt aus bis zum (näherungsweise) vollständigen Verschluß als alternative Definition verwendet:

$$OQ_2(\sigma) = t_e - t_s(\sigma) + t_a. \quad (2.13)$$

Aus theoretischer Warte kann keiner der beiden Definitionen (2.12), (2.13) der Vorzug gegeben werden. Die Auswahl ist stattdessen von der Betrachtungsweise abhängig: Aus physiologischer Sicht mag der Zeitraum bis zum ersten Kontakt der Stimmlippen (erfaßt durch OQ_1) sinnvoll erscheinen, aus strömungsmechanischer Sicht der Zeitraum, in dem der Luftstrom nicht verschwindet (näherungsweise beschrieben durch OQ_2).

2.7.2 Der Schließquotient (CQ)

Ein weiterer Parameter, der sich direkt aus den LF-Parametern ableiten läßt, ist der „Schließquotient“ (*closing quotient*, CQ). Er gibt die relative Dauer der Schließphase bezogen auf die Gesamtlänge der Periode an:⁴

$$CQ_1 = t_e - t_p. \quad (2.14)$$

Wie im Fall des Öffnungsquotients ist es möglich, die Schließzeit der Glottis mit in die Formulierung einzubeziehen. Die alternative Definition lautet dann:

$$CQ_2 = t_e + t_a - t_p. \quad (2.15)$$

2.7.3 Der Asymmetriequotient (SQ)

Der „Asymmetriequotient“ *speed quotient* (SQ) stellt ein häufig verwendetes Maß zur Erfassung der Asymmetrie des glottalen Flusses während der Öffnungs- und Schließphase dar (Titze et al., 1983; Arends et al., 1990; Childers und Lee, 1991; Childers und Ahn, 1995; Gobl und Chasaide, 1992; Hertegård und Gauffin, 1995; Alku und Vilkmán, 1995, 1996; Sulter und Wit, 1996; Veldhuis, 1998). Er berechnet sich aus dem Verhältnis zwischen ihrer Dauern:

$$SQ = \frac{T_{\text{öffnend}}}{T_{\text{schließend}}}. \quad (2.16)$$

⁴Der Schließquotient CQ sollte nicht mit dem „Verschlußquotient“ $1 - OQ$, auch Adduktionsquotient (*adduction quotient*) genannt (Holmberg et al., 1995b), verwechselt werden.

Mit den Definitionen der Öffnungs- und Schließquotienten in (2.12)-(2.15) ergeben sich die beiden Definitionen:

$$SQ_1 = \frac{OQ_1 - CQ_1}{CQ_1} \quad (2.17)$$

$$SQ_2 = \frac{OQ_2 - CQ_2}{CQ_2} \quad (2.18)$$

2.7.4 Der „parabolic spectral parameter“ (PSP)

Ein recht neuer Parameter zur Charakterisierung des glottalen Flusses ist der *parabolic spectral parameter (PSP)* (Alku et al., 1997). Diese Größe parametrisiert die tieffrequente spektrale Kontur des Glottispulses durch Anpassen einer Parabel im Spektrum. Realisiert wird dies durch periodensynchrone Berechnung des Kurzzeitleistungsdichtespektrums, an dessen tieffrequenten Teil eine Parabel $y = ax^2 + b$ im Sinne des kleinsten Fehlerquadrats angepaßt wird. Die Grenze, bis zu welcher Frequenz die Anpassung erfolgt, wird in einem iterativen Prozeß bestimmt. Der Koeffizient a des quadratischen Terms beschreibt den tieffrequenten spektralen Abfall. Um die inhärente Grundfrequenzabhängigkeit zu eliminieren, wird a auf denjenigen Wert a_{ref} bezogen, der sich bei demselben Vorgehen für ein konstantes Signal gleicher Länge (d.h. der Dauer der betrachteten Grundperiode) ergibt. Der auf diese Weise normierte (und dadurch wieder positive) Wert a/a_{ref} definiert den Parameter *PSP*. In ihren Untersuchungen fanden Alku und Koautoren *PSP* anderen glottalen Beschreibungsgrößen in der Unterscheidung zwischen verschiedenen Phonationsarten überlegen. Die Betonung auf der tieffrequenten Beschreibung des Glottispulses empfiehlt *PSP* als Ergänzung zu den anderen „abgeleiteten“ Parametern, die generell eher die hochfrequenten spektralen Charakteristiken erfassen.

Die in Alku et al. (1997) genannte Grundfrequenzunabhängigkeit konnte jedoch in eigenen Versuchen nicht bestätigt werden. Vielmehr wurde eine deutliche Abhängigkeit des *PSP* von der Abtastrate des Signals bzw. vom Verhältnis der Periodenlänge zur verwendeten FFT-Länge (*fast Fourier transform*) beobachtet, das den Abbruch des iterativen Anpassungsprozesses der Parabel bestimmt. Für die Untersuchungen dieser Arbeit stellte dies jedoch für viele Signalarten keine gravierende Beeinträchtigung dar, da die Grundfrequenz für Signale mit parametrischer Anregung konstant bei 107Hz gehalten wurde (Signaltyp FLT, SYN) und bei Anregung mit dem selbstschwingenden Stimmlippenmodell (Signaltyp 2MMF, 2MMS) nur eine geringe Schwankung der Grundfrequenz auftrat (170-179Hz). Lediglich bei Signaltyp ECHT ist aufgrund der Unterschiede in der Grundfrequenz zwischen den einzelnen Stimmprouben (80-200Hz) eine Beeinflussung der Ergebnisse durch diesen Effekt nicht auszuschließen.

2.8 Statistische Auswertungen

Alle durchgeführten statistischen Auswertungen beruhen auf Standardverfahren, die in der Literatur ausführlich beschrieben sind. Sofern nicht anders angegeben, dient als Referenz für die im folgenden aufgeführten Tests das Buch „Numerical Recipes in C“ (Press et al., 1988).

Zum Überprüfen von Unterschieden zwischen Verteilungsmittelwerten wurde der t -Test für unbekanntes Varianzen angewendet. Er erlaubt unabhängig von der Verteilung der Datenpunkte eine Aussage über die Signifikanz der Mittelwertsdifferenzen.

Korrelationen wurden als Rangkorrelationen durch den *Spearman rank order correlation test* ermittelt und bezüglich ihrer statistischen Signifikanz überprüft. Dieser Test stellt keine Voraussetzungen an die Datenverteilungen, die insbesondere nicht normalverteilt sein müssen. Da solch eine Voraussetzung bei den vorliegenden Daten nicht garantiert werden konnte, wurde die Rangkorrelation der direkten Korrelation vorgezogen.

Um eine Aussage über die Wichtigkeit der verschiedenen Parameter in der Charakterisierung der verschiedenen Phonationsarten zu erhalten, wurde die Fishersche Diskriminanzanalyse mit sukzessiver Reduktion der Merkmale (Hartung und Elpelt, 1995) durchgeführt. Diese setzt allerdings eine geeignete Eingangsdatenmenge voraus, da die Koordinaten einerseits nicht linear abhängig sein dürfen, andererseits das Verfahren nicht zwischen gegebenen Parametern im Sinne einer Interpolation oder Bildung von Linearkombinationen verallgemeinern kann.

Um den Parameterraum durch Eliminieren linearer Abhängigkeiten in der Dimension zu reduzieren, wurde eine Hauptkomponentenanalyse durch *singular value decomposition* (SVD) durchgeführt. Hierbei wird durch die Berechnung neuer Hauptachsen („Eigenvektoren“) zusammen mit ihrer Bedeutung in der Beschreibung der Varianz der Verteilung („Eigenwerte“) der Datenraum bezüglich neuer Koordinaten beschrieben. Indem eine Schwelle für die Eigenwerte angesetzt wird, unterhalb der die zugehörigen Basisvektoren als vernachlässigbar in der Beschreibung der Varianz angesehen werden, erfolgt die Beschreibung der Daten nur noch bezüglich der verbleibenden wichtigen Hauptachsen (d.h. bezüglich der Eigenvektoren mit hohen Eigenwerten).

Ergebnisse der statistischen Auswertungen werden durch Verwendung verschiedener Symbole zur Markierung der Signifikanzniveaus \mathbf{p} in den Tabellen angezeigt. Tabelle 2.4 gibt einen Überblick über die verwendeten Symbole.

Tabelle 2.4: Symbole zur Kennzeichnung der unterschiedlichen Signifikanzniveaus \mathbf{p} .

kein Symbol	hochsignifikant	($\mathbf{p} < 0.001$)
+	signifikant	($0.001 \leq \mathbf{p} < 0.01$)
*	schwach signifikant	($0.1 \leq \mathbf{p} < 0.05$)
†	insignifikant	($\mathbf{p} \geq 0.05$)

2.9 Die „Diskrete Nurpole-Modellierung“ (DAP)

Die diskrete Nurpole-Modellierung (*discrete all-pole modeling, DAP*) wurde von El-Jaroudi und Makhoul vorgestellt (El-Jaroudi und Makhoul, 1991). Sie stellt eine Weiterentwicklung der linearen Prädiktion (LP) auf Spektren dar, bei denen die Anpassung des Modellspektrums an das Signalspektrum an diskreten spektralen Stützstellen erfolgt.

2.9.1 Ursprüngliches Verfahren

Die lineare Prädiktion (LP) dient dazu, die linear redundanten Eigenschaften eines Signals – anschaulich gesprochen, die spektrale Grobstruktur – durch wenige Parameter zu erfassen. Für ein Prädiktionsfilter p -ter Ordnung wird der Signalwert s zum (diskreten) Zeitpunkt n aus den vorherigen p Abtastwerten des Signals vorhergesagt (das Symbol $\hat{\cdot}$ bezeichnet im folgenden prädizierte Größen):

$$\hat{s}(n) = \sum_{i=1}^p a_i s(n-i). \quad (2.19)$$

Durch Minimierung der Summe über die quadrierten Vorhersagefehler

$$E_N = \sum_{n=1}^N (s(n) - \hat{s}(n))^2 \quad (2.20)$$

in einem vorgegebenen Zeitintervall $n = 1, \dots, N$ lassen sich die Koeffizienten a_i bestimmen:

$$\frac{\partial E_N}{\partial a_i} = 0, \quad 1 \leq i \leq p. \quad (2.21)$$

Die Koeffizienten a_i beschreiben ein Nurpole-Filter, dessen Leistungsdichtespektrum unter Verwendung der zugehörigen Übertragungsfunktion $A(\omega)$ lautet:

$$\hat{P}(\omega) = |H(\omega)|^2 = \frac{1}{|A(\omega)|^2} = \frac{1}{|\sum_{k=0}^p a_k e^{-i\omega k}|^2}, \quad (2.22)$$

wobei a_0 nicht auf den Wert 1 beschränkt ist. Bei festem a_0 ist die Minimierungsbedingung (2.21) äquivalent zur Minimierung des Fehlers (El-Jaroudi und Makhoul, 1991)

$$\epsilon_{\text{LP}} = \frac{1}{N} \sum_{m=1}^N \frac{P(\omega_m)}{\hat{P}(\omega_m)}. \quad (2.23)$$

$\hat{P}(\omega_m)$ ist hierbei das Leistungsdichtespektrum des angepaßten Nurpole-Filters nach

(2.22) an N diskreten Stützstellen, $P(\omega_m)$ dasjenige des vorgegebenen Signals

$$P(\omega) = \left| \sum_{n=0}^N s(n) e^{-in\omega} \right|^2 \quad (2.24)$$

an denselben Stützstellen ω_m . Eine weitere äquivalente Formulierung lautet, daß die Autokorrelationsfunktionen (AKF) des linearen Prädiktionsfilters und die des Signals in den ersten $p+1$ Werten übereinstimmen (Markel und Gray, 1976; Schroeder, 1982):

$$\hat{R}_{\text{LP}}(i) = R(i), \quad 0 \leq i \leq p, \quad (2.25)$$

wobei

$$\hat{R}_{\text{LP}}(i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{P}(\omega) e^{i\omega i} d\omega \quad (2.26)$$

$$R(i) = \frac{1}{N} \sum_{m=1}^N P(\omega_m) e^{i\omega_m i}. \quad (2.27)$$

Wird das Spektrum nach (2.27) also an diskreten Punkten ω_m abgetastet, ergibt sich mit dem allgemein gültigen Fouriertransformationspaar zwischen Leistungsdichtespektrum P und Autokorrelationsfunktion R_{orig}

$$R_{\text{orig}}(i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P(\omega) e^{i\omega i} d\omega \quad (2.28)$$

$$P(\omega) = \sum_{l=-\infty}^{\infty} R_{\text{orig}}(l) e^{-i\omega l}, \quad (2.29)$$

die Beziehung zu der AKF bezüglich des kontinuierlichen Spektrums R_{orig} :

$$R(i) = \sum_{l=-\infty}^{\infty} R_{\text{orig}}(i - lN), \quad \forall i. \quad (2.30)$$

Dies bedeutet, daß durch die spektrale Diskretisierung ein „Aliasing“⁵ der AKF erfolgt, dessen Ausmaß von der Anzahl der betrachteten spektralen Stützstellen N abhängt.

Aus Gleichungen (2.25) und (2.30) folgt, daß die geschätzte AKF \hat{R}_{LP} von der Original-AKF R_{orig} abweicht, und dies um so stärker, je geringer die Anzahl der spektralen Anpassungspunkte N ist. Ist die Grundfrequenz eines vorliegenden, periodischen

⁵Der englische Begriff „*aliasing*“ beschreibt die Verfälschung des Spektrums bei einem periodisch abgetasteten Signal, bei dem das Abtasttheorem nicht eingehalten wurde, d.h. die Bandbegrenzung auf die halbe Abtastfrequenz nicht erfolgte. Dieser wichtige Begriff in der digitalen Signalverarbeitung besitzt keine deutsche Übersetzung oder Entsprechung und wird daher eingedeutscht verwendet.

Signals bekannt, werden diese Stützstellen für die Anpassung gerne als Vielfache der Grundfrequenz gewählt, da dann die spektrale Hüllkurve der Harmonischen direkt approximiert wird. In diesem Fall nimmt allerdings die Abweichung zwischen den Autokorrelationsfunktionen mit steigender Grundfrequenz zu (d.h. weniger Stützstellen bei gegebener Abtastrate).

Um dem „Aliasing“ durch die Diskretisierung des Spektrums Rechnung zu tragen, wird die AKF des angepaßten Modells nicht aus dem kontinuierlichen Spektrum gemäß (2.26) berechnet, sondern an denselben diskreten Stützstellen ω_m , die zur Berechnung des Signalspektrums dienen:

$$\hat{R}(i) = \frac{1}{N} \sum_{m=1}^N \hat{P}(\omega_m) e^{i\omega_m i} \quad (2.31)$$

$$\stackrel{(2.22)}{=} \frac{1}{N} \sum_{m=1}^N \frac{e^{i\omega_m i}}{\left| \sum_{k=0}^p a_k e^{-i\omega_m k} \right|^2}. \quad (2.32)$$

Hierdurch wird sowohl die AKF des Eingangssignals $R(i)$ als auch die des prädierten Signals $\hat{R}(i)$ auf dieselbe Weise „aliased“, so daß die durch (2.30) ausgedrückte prinzipielle Abweichung zwischen den Autokorrelationsfunktionen aufgehoben wird. Die Bedingungen zur Bestimmung der Prädiktionskoeffizienten (2.25) lauten dann:

$$\hat{R}(i) = R(i), \quad 0 \leq i \leq p, \quad (2.33)$$

wobei sich die beiden Autokorrelationsfunktionen durch (2.27) und (2.32) berechnen.

Diese Berücksichtigung des Aliasing-Effekts sowie die Verwendung eines angemesseneren Fehlermaßes als des durch (2.23) gegebenen Ausdrucks bilden die Kernstücke des DAP-Algorithmus. Als Fehlermaß wird das auf diskrete Signale angepaßte Itakura-Saito-Fehlermaß verwendet:

$$\epsilon_{\text{IS}} = \frac{1}{N} \sum_{m=1}^N \left(\frac{P(\omega_m)}{\hat{P}(\omega_m)} - \ln \frac{P(\omega_m)}{\hat{P}(\omega_m)} - 1 \right). \quad (2.34)$$

Dieses Fehlermaß ist immer nicht-negativ und verschwindet genau dann, wenn $P(\omega_m) = \hat{P}(\omega_m)$ für alle ω_m . Eine äquivalente Minimaleigenschaft bei übereinstimmenden Spektren ist für das LP-Fehlermaß in (2.23) nicht gegeben. Die Minimierung bezüglich des Itakura-Saito-Fehlermaßes ist gleichbedeutend mit einer Maximierung der spektralen Flachheit des residualen Fehlers (El-Jaroudi und Makhoul, 1991). Diese Eigenschaft ist wichtig im Hinblick auf die im folgenden Abschnitt beschriebene Modifikation des DAP-Algorithmus zur Berücksichtigung des Glottisspektrums. Der vollständige Algorithmus zur Berechnung der Filterkoeffizienten a_i ist in Abbildung 2.19 zusammengefaßt. Die mathematische Formulierung des Algorithmus findet sich in Anhang B.

DAP Algorithmus

1. Berechne $R(i) = \frac{1}{N} \sum_{m=1}^N P(\omega_m) e^{i\omega_m i}$.
2. Berechne Initialisierungswerte der a_i durch lineare Prädiktion nach der Kovarianzmethode (Markel und Gray, 1976).
3. Berechne $A(\omega_m)$ für $1 \leq m \leq N$ nach (2.22).
4. Berechne die zeitinverse Impulsantwort $\hat{h}(-i)$ des Prädiktionsfilters für die ersten $p + 1$ Glieder (Gl. (B.5)).
5. Berechne die neuen Filterkoeffizienten a_i aus der Beziehung zwischen den ersten $p + 1$ AKF-Termen, den Filterkoeffizienten und der zeitinversen Impulsantwort (Gl. (B.7)).
6. Berechne ϵ_{IS} nach (2.34).
7. Wenn $\epsilon_{\text{IS}}(j - 1) - \epsilon_{\text{IS}}(j) > \xi$ (ξ : Schwellwert, j : Iterationsschritt): Erhöhe j um 1, gehe zu 3.
8. Normalisiere die Koeffizienten so, daß $\frac{1}{N} \sum_{m=1}^N \frac{P(\omega_m)}{\hat{P}(\omega_m)} = 1$.
9. Ende.

Abbildung 2.19: Ursprünglicher Algorithmus nach El-Jaroudi und Makhoul (1991).

2.9.2 Erweiterung zur Schätzung des glottalen Flusses

Im eingangs beschriebenen Quelle-Filter-Modell (vgl. Abb. 1.4) ergibt sich das glottale Signal aus der Faltung der δ -Pulsfolge mit der glottalen Pulsform (vgl. Abb. 1.3), die idealisiert durch ein parametrisches Glottismodell dargestellt werden kann. Das Itakura-Saito-Fehlermaß zur Bestimmung der Filterkoeffizienten in Gl. (2.34) impliziert, daß das Residuumsignal, d.h. der nicht präzifizierbare Signalanteil, ein möglichst flaches Spektrum besitzt (El-Jaroudi und Makhoul, 1991). Unter Annahme der korrekten Phase sollte demnach das Residuumsignal die ursprüngliche δ -Pulsfolge darstellen, so daß es nicht dazu dienen kann, Eigenschaften des glottalen Flusses oder seiner Ableitung direkt abzulesen.

Die Faltung der δ -Pulsfolge mit der glottalen Pulsform geht im Spektrum in eine Multiplikation über (vgl. Abb. 1.3) bzw. in eine Addition der beiden logarithmierten Leistungsspektren. Dies eröffnet einen einfachen Weg, die glottale Pulsform bei der

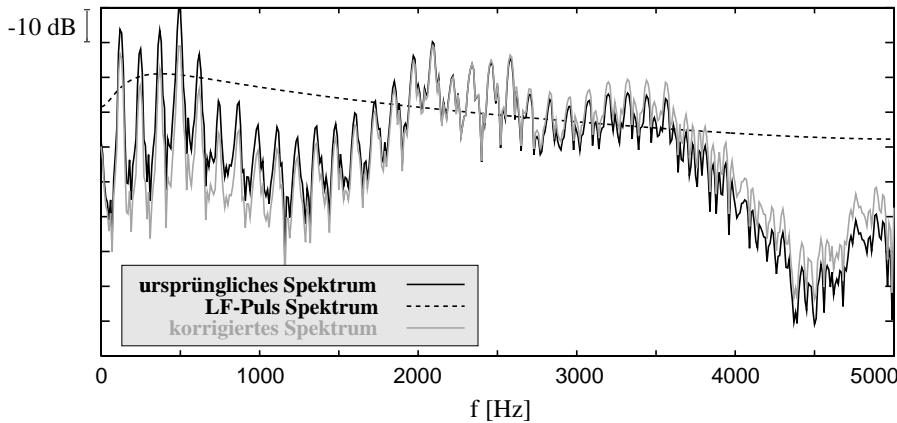


Abbildung 2.20: Signal-, LF-Puls- und LF-Puls-korrigiertes Spektrum.

Inversfilterung zu berücksichtigen: Ihr spektraler Einfluß P_{gl} wird durch Subtraktion des logarithmierten Leistungsspektrums aus dem Signalspektrum rechnerisch eliminiert. Konkret bedeutet dies, daß (2.27) durch

$$R(i) = \frac{1}{N} \sum_{m=1}^N \frac{P(\omega_m)}{P_{gl}(\omega_m)} e^{i\omega_m i} \quad (2.35)$$

ersetzt wird, bevor (2.33) ausgeführt wird. Um numerische Instabilitäten zu vermeiden, wird statt des Quotienten der Spektren in (2.35) die Differenz der logarithmierten Spektren verwendet. Die Spektren $\log(P)$, $\log(P_{ref})$ und $\log(P) - \log(P_{gl})$ sind in Abbildung 2.20 beispielhaft aufgezeigt.

Der gesamte modifizierte DAP-Algorithmus zur Berücksichtigung des glottalen Spektrums bei gegebenen Parametern des glottalen Modells ist in Abbildung 2.21 angegeben und in Abbildung 2.22 schematisch dargestellt. Abweichend von der ursprünglichen Formulierung stellte sich eine relative Definition des Abbruchkriteriums als geeigneter heraus als das absolute Kriterium im originalen Algorithmus. Die Verringerung des Fehlers während einer vollständigen Iteration ist in Abbildung 2.23 dargestellt.

In der modifizierten Form des DAP-Algorithmus wird die Inversfilterung bezüglich des „glottis-bereinigten“ Spektrums durchgeführt und sollte daher gemäß des Quelle-Filter-Ansatzes ein Signal mit weißem Spektrum erzeugen. Die Phase wird bei diesem Vorgehen nicht explizit berücksichtigt, so daß das resultierende IF-Signal im Zeitbereich nicht zwangsläufig die Eingangs- δ -Pulsfolge darstellen muß. Aufgrund der eingeschränkten Variationsmöglichkeiten der glottalen Pulsform, die durch das verwendete glottale Modell bestimmt ist, stellt dies jedoch kein gravierendes Problem dar, so daß andere Autoren selbst bei Nichtberücksichtigung der glottalen Pulsform bei herkömmlichen IF-Methoden sogar direkt Parameter aus dem Residuum im Zeitbereich ableiteten (Eskenazi et al., 1990). Nichtsdestotrotz wird in der vorliegenden Arbeit die zeitliche

Modifizierter DAP Algorithmus

1. Berechne $P_{\text{ref}}(\omega)$ des LF-Signals zu einer gegebenen Parameterkonfiguration.
2. Berechne $R(i)$ nach (2.35).
3. Berechne Initialisierungswerte der a_i durch lineare Prädiktion nach der Kovarianzmethode (Markel und Gray, 1976).
4. Berechne $A(\omega_m)$ für $1 \leq m \leq N$ nach (2.22).
5. Berechne die zeitinverse Impulsantwort $\hat{h}(-i)$ des Prädiktionsfilters für die ersten $p + 1$ Glieder (Gl. (B.5)).
6. Berechne die neuen Filterkoeffizienten a_i aus der Beziehung zwischen den ersten $p + 1$ AKF-Termen, den Filterkoeffizienten und der zeitinversen Impulsantwort (Gl. (B.7)).
7. Berechne ϵ_{IS} nach (2.34).
8. Wenn $\frac{\epsilon_{\text{IS}}(j-1) - \epsilon_{\text{IS}}(j)}{\epsilon_{\text{IS}}(j)} > \xi$ (ξ : Schwellwert, j : Iterationsschritt): erhöhe j um 1, gehe zu 3.
9. Normalisiere die Koeffizienten so, daß $\frac{1}{N} \sum_{m=1}^N \frac{P(\omega_m)}{\hat{P}(\omega_m)} = 1$.
10. Ende.

Abbildung 2.21: Modifizierter Algorithmus zur Schätzung des glottalen Flusses. Die Änderungen bezüglich Abbildung 2.19 sind durch Rahmen hervorgehoben.

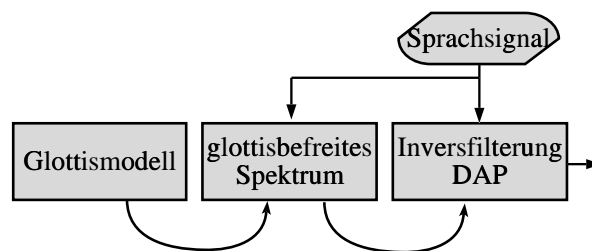


Abbildung 2.22: Inversfilterung nach dem modifizierten DAP-Algorithmus (Abb. 2.21) unter Berücksichtigung des glottalen Spektrums.

Struktur des Residuumsignals nicht weitergehend analysiert.

In das Itakura-Saito-Fehlermaß gehen sowohl Summen über die Quotienten spektraler Abtastwerte als auch Summen über die Logarithmen dieser Quotienten ein. Theoretisch läßt die Formulierung des modifizierten DAP-Algorithmus ebenso wie der ursprüngliche Algorithmus eine Anpassung des gesamten Spektrums an allen Abtastwerten zu, so daß das angepaßte Modell wie im Falle der Autokorrelationsmethode der „klassischen“ linearen Prädiktion als Anpassung eines AR-Modells an das gesamte Spektrum anzusehen ist (Markel und Gray, 1976). Rechnerisch besitzt solch ein Vorgehen den Vorteil, daß sich Spektren bzw. Autokorrelationsfunktionen effizient durch schnelle Fouriertransformation (FFT) berechnen lassen.

Es zeigte sich allerdings, daß bei diesem Vorgehen die zwischen den Harmonischen liegenden Spektralbereiche geringer Energie allein durch ihre deutlich stärkere Repräsentation (mehr Abtastwerte) das Ergebnis maßgeblich beeinflussten. Hierbei stellte sich ebenfalls eine Abhängigkeit von der Implementation der FFT-Routinen heraus, die zu numerischen Unterschieden in den hinteren, insignifikanten Dezimalstellen der Fließkomma-Repräsentation der Ergebnisse führten. Diese Unterschiede wurden aufgrund des logarithmischen Terms in Gl. (2.34) in solch einem Maß verstärkt, daß die Inversfilterroutine bei identischen Eingangssignalen für verschiedene FFT-Implementationen teilweise unterschiedlich viele Iterationsschritte benötigte, wodurch verschiedene Ergebnisse resultierten. Da auch aus theoretischer Sicht die Anpassung des Nurpole-Spektrums lediglich an spektral wichtige Punkte, z.B. die Harmonischen, sinnvoll erscheint, wurde die Möglichkeit der Methode ausgenutzt, die Stützstellen in der Anpassung der Spektren – beschrieben durch Gl. (2.34) – nicht-äquidistant wählen zu können.

Die konkret verwendeten ω_m wurden wie folgt berechnet: Die Median-Periodenlänge $T_{0,\text{median}}$ des aktuell betrachteten Signalabschnitts (10-Periodensequenz, siehe S. 50) der Länge N_{sig} definiert die Grund(kreis)frequenz $\tilde{\omega}_{0,\text{ref}} = 2\pi/(f_s \cdot T_{0,\text{median}})$. Für $m \geq 1$

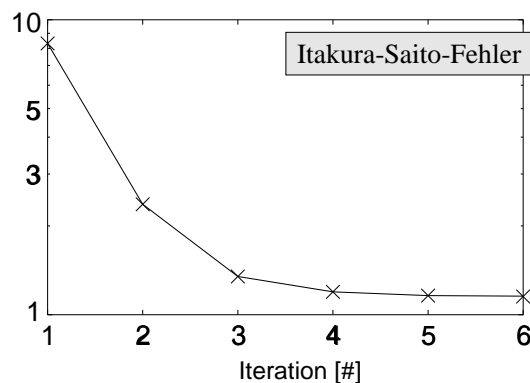


Abbildung 2.23: Entwicklung des Itakura-Saito-Fehlers während der Iteration.

wird die diskrete Frequenz ω_m als derjenige Abtastwert gewählt, der der echten m -ten Harmonischen $\tilde{\omega}_m = m \cdot \tilde{\omega}_{0,\text{ref}}$ am nächsten liegt:

$$\omega_m = \{\omega_i | \min(m \cdot \tilde{\omega}_{0,\text{ref}} - \omega_i); 1 \leq i \leq N_{\text{sig}}\} \quad (2.36)$$

Die Werte von ω_m sind auf diese Weise nicht notwendigerweise äquidistant verteilt, stellen aber die beste Nachbildung der ursprünglichen spektralen Hüllkurve dar, ohne auf ganzzahlige Vielfache der diskreten Kreisgrundfrequenz ω_0 festgelegt zu sein, die mit zunehmendem m um bis zu $\Delta\omega_m = m \cdot \frac{\pi}{N}$ von der tatsächlichen Harmonischen $\tilde{\omega}_m$ abweichen kann.⁶ Die Transformationen zwischen Zeit- und Spektralbereich können damit nicht mittels FFT ausgeführt werden, sondern müssen „zu Fuß“ durch explizite Berechnung der Summen erfolgen. Dies stellt jedoch keine gravierende Erhöhung des Rechenbedarfs dar, da einerseits die Anzahl m der Harmonischen bei den untersuchten Grundfrequenzen um 100Hz in der Größenordnung von 50 liegt, andererseits die Anzahl der zu berechnenden, zeitverzögerten Terme der Impulsantwort (Schritt 5 in Abbildung 2.21) durch die Prädiktionsordnung vorgegeben ist. Die trigonometrischen Terme $\cos(i\omega_m)$ und $\sin(i\omega_m)$ mit $0 \leq i \leq p, 1 \leq m \leq N$ ändern sich während der Iteration nicht und können daher in einer Tabelle abgelegt werden.

⁶Die Alternative der direkten Verwendung von $\tilde{\omega}_m = m \cdot \tilde{\omega}_{0,\text{ref}}$, d.h. eine neue Abtastung des Spektrums bei Vielfachen der echten Grundfrequenz, wurde aus Gründen der Recheneffizienz nicht vorgenommen: Angesichts der maximalen Abweichung von einem halben Abtastwert bei dem angewendeten Verfahren war hierdurch keine signifikante Verbesserung der Ergebnisse zu erwarten, wohingegen der Rechenaufwand beträchtlich gestiegen wäre.

2.10 Die SIM-Methode

Der im vorangegangenen Abschnitt beschriebene modifizierte DAP-Algorithmus ermöglicht zwar die explizite Berücksichtigung der glottalen Pulsform, setzt dabei jedoch Kenntnis der Parameter, die diese Form spezifizieren, voraus. Für die praktische Durchführung stellt sich nun ein klassisches „Henne – Ei Problem“: Zur spektralen Subtraktion des Glottispulses nach Gl. (2.35) muß die Inversfilterung bereits erfolgreich durchgeführt worden sein, um die korrekten Glottisparameter zu kennen, die jedoch erst die optimale Inversfilterung ermöglichen.

Zur Lösung dieses Problems kommt ein iteratives Verfahren in Frage. In jedem Iterationsschritt wird das verwendete LF-Modell zur Parametrisierung der Ableitung des glottalen Flusses mit neuen Parametereinstellungen angesetzt, dessen spektrale Eigenschaften dann in der Inversfilterung durch den modifizierten DAP-Algorithmus gemäß Gl. (2.35) berücksichtigt werden.

Konkret bedeutet dies, daß für jeden Iterationsschritt die glottale Pulsform durch einen vierdimensionalen Vektor $\mathbf{p} = (t_p, t_e, t_a, E_e)$ festgelegt ist. Mit diesen Parametern wird ein Prototyp-Puls synthetisiert und sein Spektrum P_{g1} gemäß Gl. (2.24) berechnet. Das komplette Verfahren zur simultanen Inversfilterung und Anpassung des glottalen Modells, kurz „SIM-Methode“, unter Verwendung des modifizierten DAP-Algorithmus aus Abbildung 2.22 ist in Abbildung 2.24 dargestellt.

Zur Steuerung des multidimensionalen Optimierers wird der Fehler zu gegebenem Vektor \mathbf{p} für ein Fenster von 200ms Länge wie folgt berechnet:

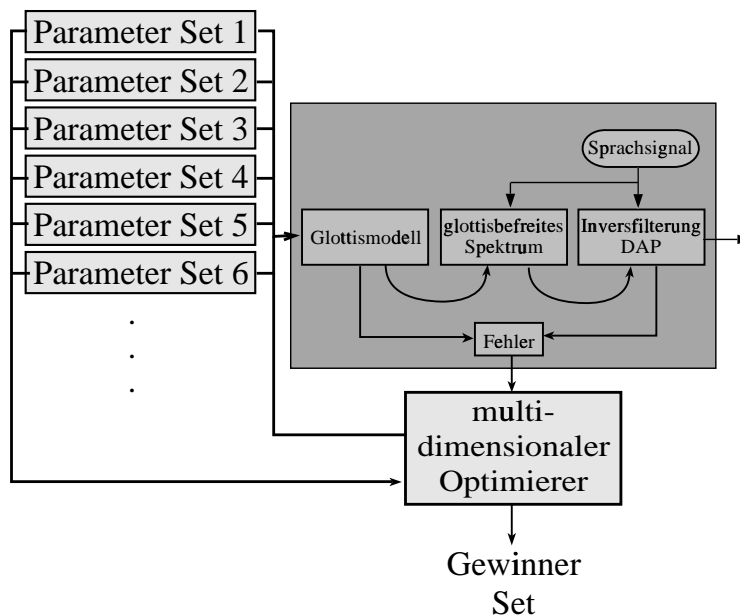


Abbildung 2.24: Simultane Inversfilterung und Anpassung des glottalen Modells (SIM-Methode).

1. Berechnung der mittleren Periodenlänge als Median der Periodenlängen T_i im vorliegenden Fenster (200ms).
2. Jede mögliche Sequenz von 10 aufeinanderfolgenden Perioden (d.h. Vorschub: eine Periodenlänge) wird mit dem modifizierten DAP-Algorithmus nach Abbildung 2.21 inversgefiltert und der Itakura-Saito-Fehler bei Abbruch der Iteration zwischengespeichert.
3. Die resultierenden Filter für jede 10-Periodensequenz werden auf ihre Stabilität überprüft:
 - (a) Wenn mindestens eine 10-Periodensequenz ein stabiles Filter geliefert hat: Itakura-Saito-Fehler bei DAP-Iterationsabbruch von derjenigen Sequenz, die ein stabiles Filter mit dem kleinsten Fehler ergeben hat, wird zurückgegeben.
 - (b) Wenn kein Filter stabil ist: Fehler derjenigen Sequenz wird zurückgegeben, die den kleinsten IF-Fehler bei DAP-Abbruch geliefert hat, und eine Warnung wird ausgegeben.

Im Fall 3b, daß bei allen möglichen 10-Periodensequenzen innerhalb des Fensters instabile Filter auftraten, wird der kleinste Fehler zurückgegeben und die Inversfilterung mit den Koeffizienten des zugehörigen Nulpole-Filters vorgenommen. Für die rechnerische Elimination des Vokaltrakt-AR-Modells, d.h. für die Inversfilterung, beschreiben die Koeffizienten ein FIR (*finite impulse response*)-Filter, so daß hierbei aufgrund der Nicht-Rekursivität keine Stabilitätsprobleme auftreten können. Zudem zeigte sich in Untersuchungen an synthetischen Signalen, daß heuristisch motivierte Verfahren wie Spiegelung des Pols mit Radius größer als 1 am Einheitskreis (Press et al., 1988), der ein stabiles IIR (*infinite impulse response*)-Filter zur Beschreibung des Vokaltrakt-AR-Modells bewirkt, zu schlechteren Ergebnissen führte als die Inversfilterung ohne Modifikation der Pole (Abb. 2.25). Daher ist keine gravierende Verfälschung der Ergebnisse aufgrund der direkten Verwendung der Filterkoeffizienten zu erwarten, auch wenn für diese Signalabschnitte die Interpretation der Filterkoeffizienten im Sinne der AR-Beschreibung theoretisch unbefriedigend wird.

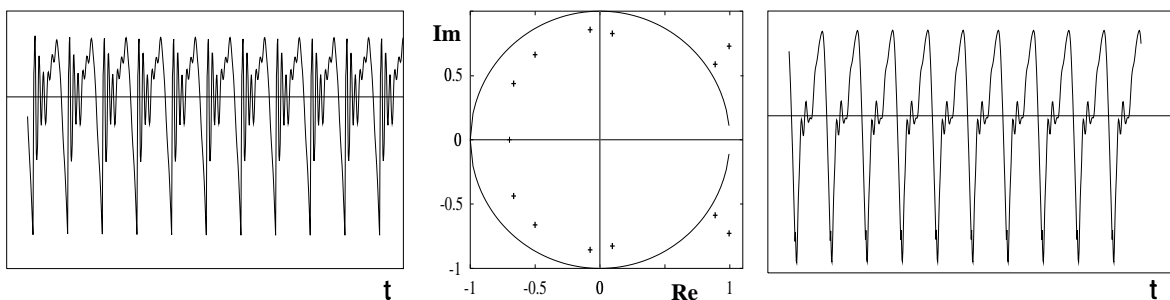


Abbildung 2.25: Inversfilterung mit Filterkoeffizienten, die ein instabiles AR-Modell beschreiben. Links: Eingangssignal (Signaltyp SYN), mitte: Filterkoeffizienten als Pole bzw. Nullstellen im komplexen Einheitskreis, rechts: Resultat der Inversfilterung mit dem durch die komplexen Nullstellen im mittleren Bild gegebenen FIR-Filter.

2.11 Multidimensionale Optimierung

Die SIM-Methode nach Abbildung 2.24 greift auf multidimensionale Optimierungsverfahren zur Schätzung der optimalen LF-Parameterkonstellation zurück. Hierzu können gebräuchliche Algorithmen eingesetzt werden, sofern in sie nicht die analytischen Ableitungen des Fehlersignals nach den einzelnen Koordinaten eingehen, die im vorliegenden Fall nicht in geschlossener Form angegeben werden können. Zwei in der Literatur beschriebene Methoden wurden untersucht: die n -dimensionale Simplex-Methode und die Methode der konjugierten Richtungssätze (Press et al., 1988). Die Simplexmethode diente bereits anderen Autoren zur Anpassung des LF-Modells an das IF-Signal (Strik, 1998).

Das n -dimensionale Simplex-Minimierungsverfahren (*downhill simplex method*) geht auf Nelder und Mead zurück (Nelder und Mead, 1965). Ein Simplex ist ein $n + 1$ -Eck im n -dimensionalen Raum und stellt damit die Verallgemeinerung eines Dreiecks in der Ebene bzw. eines Tetraeders im dreidimensionalen Raum auf n Dimensionen dar. Beginnend von einer Startkonfiguration, die durch $n + 1$ Punkte (Vektoren) im n -dimensionalen Parameterraum gegeben ist, wird der Punkt mit maximalem Fehler an dem durch die übrigen Punkte definierten Unterraum gespiegelt. Diese Vorschrift wird bis zur Erfüllung des Abbruchkriteriums wiederholt, wobei sich das Simplex alternativ zur Spiegelung auch zusammenziehen kann, entweder durch Bewegen des Punktes mit dem größten Fehler auf die anderen Punkte zu, oder durch Bewegen aller Punkte außer demjenigen mit dem kleinsten Fehler in dessen Richtung. Ist das Simplex durch Wiederholen dieser Schritte auf ein Volumen geschrumpft, das kleiner als ein vorgegebener Schwellwert ist, oder tritt keine nennenswerte Volumenreduktion mehr auf, wird die Iteration abgebrochen und der Punkt mit dem kleinsten Fehler als Minimum ausgegeben. In Anlehnung an Press et al. (1988) wird die implementierte Version des Verfahrens auch veranschaulichend als „*amoeba routine*“ bezeichnet.

Bei der Methode der konjugierten Richtungssätze (*direction set method* oder Powell-Methode) werden zur Initialisierung ein n -dimensionaler Punkt sowie n orthogonale Richtungsvektoren im n -dimensionalen Parameterraum vorgegeben, die normalerweise in Richtung der Basisvektoren weisen. Die Längen der Richtungsvektoren sollten dabei den zu schätzenden Werten angepaßt sein, d.h. in der Größenordnung der erwarteten Werte für jede einzelne Dimension liegen. Nun wird in jedem Iterationsschritt das Minimum der Funktion entlang einer der Richtungen bestimmt. Von diesem Punkt aus wird die neue Richtung senkrecht zu den bisherigen Richtungen gewählt, so daß ein neuer n -dimensionaler Satz von konjugierten Richtungen entsteht. Für die Minimierung einer quadratischen Form müssen maximal n Minimierungsrichtungen durchlaufen werden, um das korrekte Minimum zu erhalten. Für nicht-quadratische Formen wird das Verfahren iterativ fortgesetzt, wobei das Abbruchkriterium über die relative Verringerung des Fehlers während eines Durchlaufs über alle n konjugierten Richtungen definiert ist. Die genaue Beschreibung der verwendeten Implementation (inklusive einiger numerischer Tricks) findet sich in Press et al. (1988).

Ein Problem bei der multidimensionalen Extremwertsuche stellt die starke Abhängigkeit der Ergebnisse von der Initialisierung des Optimierungsalgorithmus dar

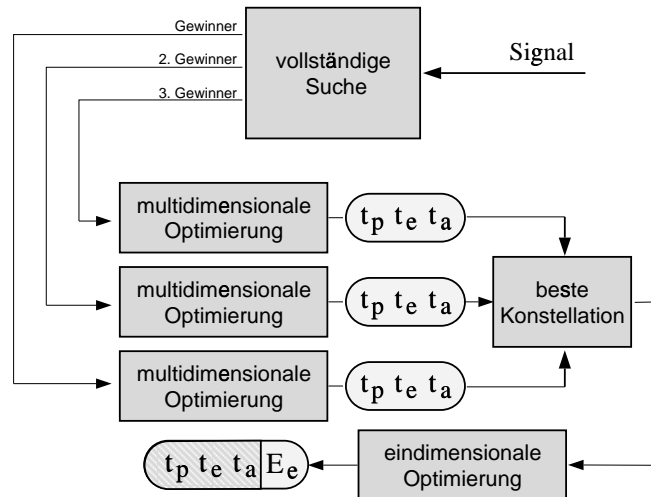


Abbildung 2.26: Mehrfache Durchführung der multidimensionalen Optimierung auf der Grundlage einer vollständigen Suche.

(Press et al., 1988). Andere Autoren haben dies konkret bei der Anpassung eines Glottismodells an das IF-Signal beschrieben (Strik et al., 1993; Milenkovic, 1986). Alternativ zu den beiden angeführten Methoden zur multidimensionalen Extremwertbestimmung, die von einem bzw. einer kleinen Anzahl an Initialisierungspunkten ausgehen, existieren Methoden, bei denen gleichzeitig möglichst der gesamte Raum in Betracht gezogen wird, wodurch sich das Initialisierungsproblem nicht in dieser Weise stellt. Zu dieser Art von Optimierungsstrategien gehören genetische Algorithmen (Beasley et al., 1993a,b), *simulated annealing*-Methoden (Kirkpatrick et al., 1983) oder „Sintflut-Algorithmen“ (Dueck und Scheuer, 1990; Dueck, 1993). Die Anwendung solcher Algorithmen zur Lösung des gegebenen Problems wurde nach einigen Versuchen aufgrund des hohen Rechenaufwands nicht weiter verfolgt, obwohl Ansätze in der Literatur beschrieben sind, solche Verfahren zur Anpassung von Glottismodellen an inversgefilterte Signale zu verwenden (Funaki et al., 1996).

Bei Versuchen mit verschiedenen Fehlermaßen stellte sich heraus, daß sich E_e unabhängig von den anderen LF-Parametern in einem nachfolgenden, eindimensionalen Optimierungsschritt durch die Brent-Methode (Press et al., 1988) schätzen läßt. Daher wird die multidimensionale Optimierung lediglich bezüglich des dreidimensionalen Raums, der durch die Parameter t_p, t_e, t_a aufgespannt wird, durchgeführt. Geeignete Startwerte für die Simplex- oder Powell-Methode lassen sich durch vollständige Suche (*exhaustive search*) ermitteln, da dieser dreidimensionale Raum ein endliches Volumen umschreibt (vgl. Abb. 2.2). Der Parameter des exponentiellen Rückgangs t_a wird hierbei zweckmäßigerweise nach logarithmischer Transformation $lt_a = \ln(t_a)$ verwendet, so daß der Optimierungsraum durch Vektoren der Form $\mathbf{q} = (t_p, t_e, \ln(t_a))$ beschrieben wird. Als obere Schranken der einzelnen Parameter lassen sich für t_p das Intervall $[0.1, 0.9]$, für t_e $[0.1, 0.9]$ und für lt_a $[-11, -1]$ (entsprechend $[0.00001, 0.36]$ für t_a) angeben.

Bei der Wahl der Schrittweite, in der die Rasterung bei der vollständigen Suche

erfolgt, ist der kubisch wachsende Rechenbedarf gegenüber der Feinheit des Rasters abzuwägen. Es zeigte sich bei Untersuchungen bezüglich verschiedener Rasterdimensionierungen, daß auch bei relativ engem Raster das Minimum der Rasterpunkte nicht notwendigerweise in der Nähe des globalen Minimums der Fehlerfunktion zu finden war, sondern sich ein Punkt in der Nähe des echten Minimums nur an zweiter Stelle, selten an dritter Stelle der nach steigendem Fehler geordneten Rasterpunkte befand. Daher wurden alle mehrdimensionalen Optimierungen bei Initialisierung auf der Basis einer vollständigen Suche dreifach durchgeführt, wobei als Referenzpunkt zuerst der Rasterpunkt mit kleinstem Fehler, dann derjenige mit zweitkleinstem Fehler und schließlich derjenige mit drittkleinstem Fehler gewählt wurde (Abb. 2.26).⁷ Hierbei wurde festgestellt, daß ein zu feines Raster nicht nur den Rechenaufwand erheblich erhöhte, sondern zudem die automatische Identifizierung guter Referenzpunkte für den zweiten und dritten Durchlauf aufgrund der vielen ähnlichen Fehlerwerte benachbarter Rasterpunkte erschwerte. Zur Vermeidung der mehrfachen Initialisierung in den drei Durchläufen mit räumlich nahe beieinanderliegenden Referenzwerten, die sich bei zu enger Rasterung aus der vollständigen Suche als die drei besten Punkte ergeben, wurde zusätzlich das heuristische Kriterium eingeführt, daß der Referenzpunkt zur Initialisierung in der zweiten und dritten Optimierung einen mindestens 5% höheren Fehler aufweisen muß als der Startwert des vorangegangenen Optimierungslaufs.

Die verwendeten Rasterpunkte sind in Abb. 2.27 dargestellt. Sie ergaben sich aus Untersuchungen mit verschiedenen Rasterdimensionierungen und folgen aus einer dreistufigen Abrasterung des LF-Raums: Im ersten Durchlauf lauten die Startwerte $\mathbf{q}_1^0 = (0.04, 0.08, -1.2)$ bei einer Schrittweite $d\mathbf{q}_1 = (0.1, 0.14, -2)$, im zweiten Durchlauf $\mathbf{q}_2^0 = \mathbf{q}_1^0 + 0.5 d\mathbf{q}_1$ bei unveränderter Schrittweite $d\mathbf{q}_2 = d\mathbf{q}_1$ und im dritten Durchlauf $\mathbf{q}_3^0 = \mathbf{q}_2^0 - 0.5 d\mathbf{q}_3$ mit $d\mathbf{q}_3 = (0.07, 0.98, -1.4)$.

⁷Ausgehend von dem jeweiligen Referenzpunkt wurden die weiteren Initialisierungspunkte für das Simplexverfahren durch Variation um 10% jeweils von $t_e, t_e - t_p$ und $\ln(t_a)$ gebildet (zehnprozentige Erhöhung des Wertes, wenn der Referenzwert in der unteren Hälfte des Wertebereichs lag, zehnprozentige Erniedrigung sonst).

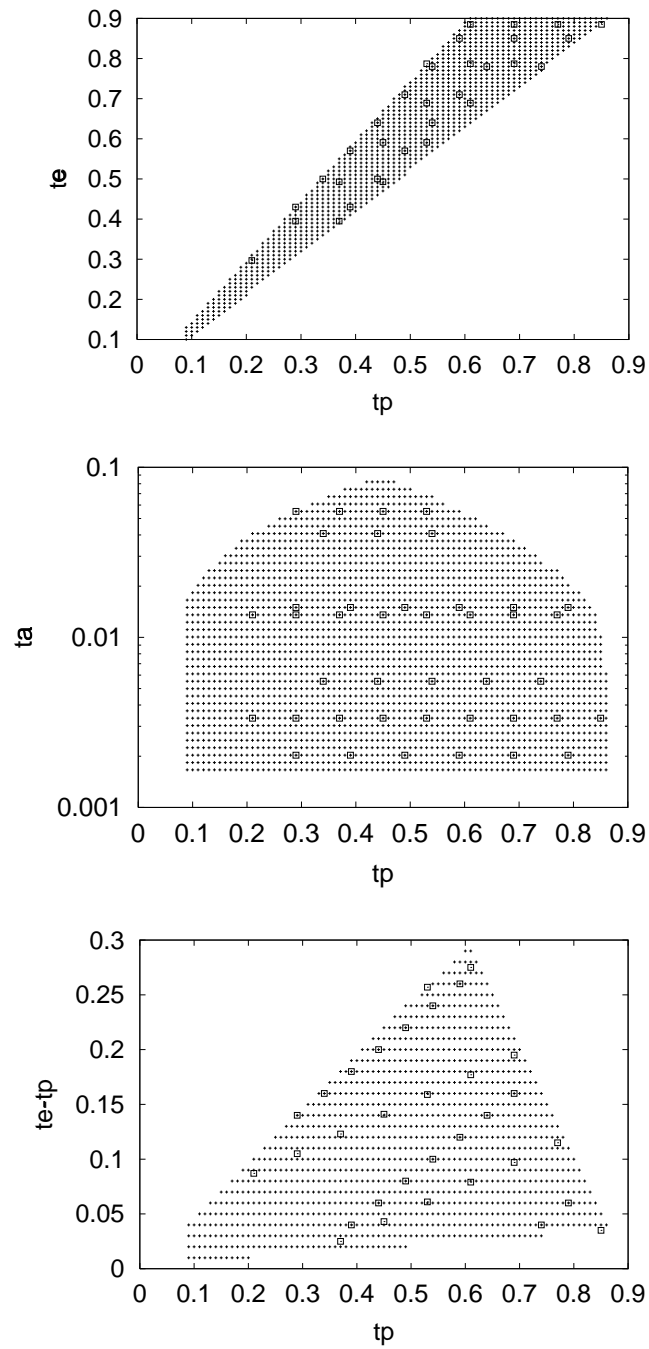


Abbildung 2.27: Gewählte Rasterpunkte (\square) bei der vollständigen Suche im Verhältnis zu den möglichen Werten ($+$) (Grundfrequenz $f_0 = 107\text{Hz}$).

2.12 Verwendete Fehlermaße und Optimierungsmethoden

Da der Erfolg einer multidimensionalen Optimierung kritisch von der Fehlerlandschaft abhängt, kommt der Wahl des Fehlermaßes besondere Bedeutung zu (Press et al., 1988). Es wurden daher umfangreiche Versuchsreihen in der Analyse verschiedener Signale mit der SIM-Methode unter Verwendung verschiedenster Arten von Fehlerdefinitionen, die teils im Zeitbereich, teils im Spektralbereich formuliert waren, durchgeführt. Hierbei stellte sich der Endwert des Itakura-Saito-Fehlers ϵ_{IS} aus dem modifizierten DAP-Algorithmus als geeignetstes Fehlermaß heraus:

$$\epsilon_{\text{IS}} = \frac{1}{N} \sum_{m=1}^N \left(\frac{P(\omega_m)}{\hat{P}(\omega_m)} - \ln \frac{P(\omega_m)}{\hat{P}(\omega_m)} - 1 \right) \quad (2.34)$$

(P : Leistungsdichtespektrum des Signals, \hat{P} : Leistungsdichtespektrum des angepaßten Nulpole-Modells).

Daneben ergab die Verwendung eines Fehlermaßes, das den tieffrequenten Anteil der spektralen Anpassung besonders gewichtet, ebenfalls in vielen Fällen gute Ergebnisse. Bei diesem Fehler ϵ_{lowf} wird die Entsprechung der Leistungsdichtespektren in den ersten 10 Harmonischen nach Konvergenz des modifizierten DAP-Algorithmus quantifiziert:

$$\epsilon_{\text{lowf}} = \sum_{m=1}^{10} \left(\frac{P(\omega_m)}{\hat{P}(\omega_m)} - 1 \right)^2. \quad (2.37)$$

Der Vorteil dieses Fehlermaßes insbesondere in der Nachoptimierung ist in den speziellen Eigenschaften des LF-Modells begründet. Die Phase des Signals wird bei der Inversfilterung nicht explizit berücksichtigt. Veränderungen in der Phase ergeben sich besonders bei Variation der genauen Lage der Parameter t_p und t_e , die ebenfalls primär für die tieffrequenten Eigenschaften der glottalen Pulsform verantwortlich sind.⁸ Die Fehlerlandschaft besteht daher bei konstantem t_a aus einer Rinne, deren Boden längs der konstanten Differenz $t_e - t_p$ verläuft (Abb. 2.28). Da sich Änderungen von t_p zu gegebener Differenz $t_e - t_p$ im wesentlichen im tieffrequenten Anteil des Spektrums äußern, läßt eine Nachoptimierung, die diesen Teil des Spektrums besonders gewichtet, zumindest bei synthetischen Signalen des Typs FLT oder SYN, die dem zugrundeliegenden Ansatz der SIM-Methode in hohem Maß entsprechen, eine Verbesserung der Schätzergebnisse für t_p und t_e erwarten.

Zur eindimensionalen Schätzung von E_e bei festen Werten von t_p, t_e, t_a (vgl. Abb. 2.26) wurde ein weiteres Fehlermaß ϵ_{dist} verwendet, das die beste Übereinstimmung im Zeitbereich im Sinne des kleinsten Fehlerquadrats zwischen synthetisiertem Signal und inversgefiltertem Signalabschnitt für die beste 10-Periodensequenz $\{T_i\}_{i=1, \dots, 10}$ (entsprechend $\{N_i\}_{i=1, \dots, 10}$ in Abtastwerten) im betrachteten 200ms-Fenster quantifiziert.

⁸Hochfrequente Abweichungen sind im wesentlichen durch t_a bestimmt (Fant, 1995), wobei die Differenz $t_e - t_p$ ebenfalls eine Rolle spielt.

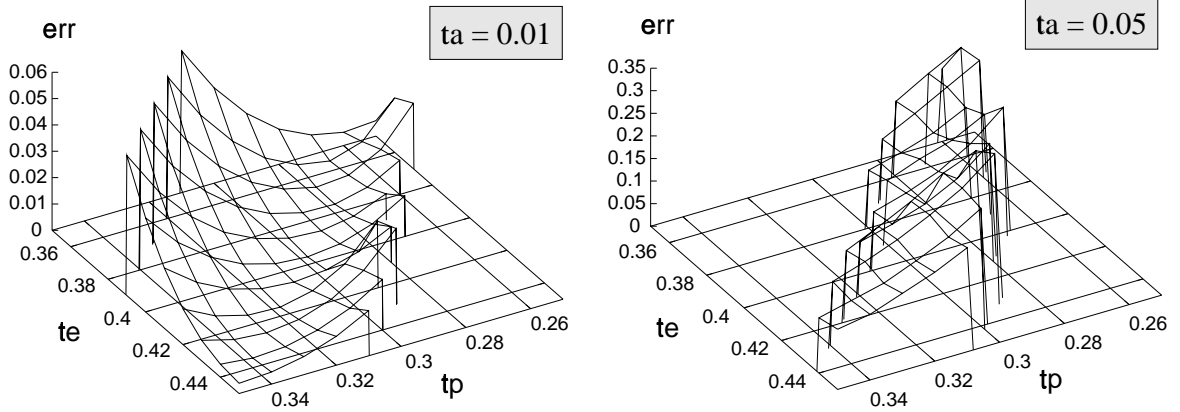


Abbildung 2.28: Fehlerlandschaft für die LF-Konfiguration $t_p = 0.3, t_e = 0.4, t_a = 0.01$ bei verschiedener Wahl von t_a im angesetzten Glottismodell für die Inversfilterung nach dem modifizierten DAP-Algorithmus (Abb. 2.21). Links: $t_a = 0.01$ (korrekter Wert), rechts: $t_a = 0.05$ (man beachte die unterschiedlichen Skalierungen der Fehlerachse err).

Die Zeitmarken, die eine Periode definieren, werden durch den *waveform matching*-Algorithmus (siehe Abschnitt 2.4) automatisch gesetzt und sind nicht zu einem bestimmten Zeitpunkt innerhalb des Glottiszyklus synchronisiert. Daher muß zunächst der Fehler σ für jede mögliche Verzögerung der Referenzkurve innerhalb der ersten Periodenlänge N_1 berechnet werden. Hierbei wird der Teil der Referenzkurve, der aufgrund der Verzögerung nicht mehr dem Zeitintervall des IF-Signals entspricht, von vorne wieder angesetzt („*wrap around*“-Vorgehen):⁹

$$\sigma(j) = \sum_{i=1}^N (x_{\text{ref}}((i+j) \bmod N_{\text{sig}}) - x_{\text{IF}}(i))^2 \quad (0 \leq j \leq N_1) \quad (2.38)$$

mit

$$N_{\text{sig}} = \sum_{i=1}^{10} N_i.$$

Das Fehlermaß ϵ_{dist} ist als Minimum von (2.38) über alle möglichen Verschiebungen innerhalb einer Periode definiert:

$$\epsilon_{\text{dist}} = \min_{0 \leq j \leq N_1} (\sigma(j)). \quad (2.39)$$

⁹Bei starkem Jitter, d.h. Abweichungen der individuellen Periodenlänge vom lokalen Mittelwert, können hierdurch theoretisch Probleme durch Unstetigkeiten im Signal auftreten, da der fehlende Teil der zehnten Periode aufgrund der verschiedenen Periodenlängen nicht mit der 1. Periode korrespondieren muß. Dies stellt jedoch in Anbetracht der Länge des Signalstücks (bei Grundfrequenz $f_0 = 107\text{Hz}$: $N_{\text{sig}} = 934$ Abtastwerte) keine signifikante Beeinträchtigung in der Berechnung von ϵ_{dist} dar.

Als Anpassung des Referenzsignals an das IF-Signal im Sinne des kleinsten quadratischen Abstands im Zeitbereich entspricht ϵ_{dist} dem Fehler, der in anderen Arbeiten als ausschließliches Kriterium für die Anpassung eines Glottismodells an ein inversgefiltertes Signal verwendet wurde (Milenkovic, 1986; Strik et al., 1993).

Für die praktische Durchführung wurden die eben beschriebenen Fehlermaße zur Definition unterschiedlicher Vorgehensweisen zur multidimensionalen Optimierung („Optimierungsmethoden“) herangezogen (vgl. Abb. 2.24). Eine Zusammenstellung dieser Methoden gibt Tabelle 2.5. Die Anwendung des Simplexverfahrens unter Verwendung des Fehlermaßes ϵ_{IS} bei Initialisierung mit einem festen Satz von Punkten wird als Optimierung A bezeichnet („*amoeba*“ in Anlehnung an Press et al. (1988)). Dieser Satz von Punkten umfaßt ein relativ großes Volumen des zulässigen Parameterraums ($\mathbf{q}'_1 = (0.52, 0.12, \ln(0.03))$, $\mathbf{q}'_2 = (0.72, 0.12, \ln(0.03))$, $\mathbf{q}'_3 = (0.52, 0.07, \ln(0.03))$, $\mathbf{q}'_4 = (0.52, 0.12, \ln(0.003))$) mit $\mathbf{q}' = (t_e, t_e - t_p, \ln(t_a))$). Die dreimalig Anwendung der Simplexmethode mit jeweiliger Initialisierung durch Variation des erst-, zweit- und drittbesten Punktes der vollständigen Suche (vgl. Abb. 2.26) und Nachoptimierung in jedem der drei Optimierungsläufe durch die Powell-Methode bezüglich des Fehlers ϵ_{lowf} wird als MA bezeichnet („*multi-amoeba*“). Werden die Nachoptimierungsschritte weggelassen, wird die Methode als MA_NN bezeichnet („*multi-amoeba* nicht nachoptimiert“). Bei Verwendung des Fehlermaßes ϵ_{lowf} bereits im Simplexverfahren anstelle von ϵ_{IS} heißt die Methode MA_LF („*multi-amoeba low-frequency*“). Hierbei erfolgt ebenfalls keine Nachoptimierung. Schließlich wurde anstatt der unabhängigen dreidimensionalen Optimierung im \mathbf{q} -Raum die Beziehung zwischen den LF-Parametern durch R_d aus Gleichungen (2.2)-(2.4) festgesetzt und eine eindimensionale Optimierung nach der Brent-Methode durchgeführt. Dieses Vorgehen wird als RD bezeichnet und wurde nur bei der Analyse echter Stimmen (Signaltyp ECHT) angewendet.

Weitere Kombinationen verschiedener Methoden, z.B. die direkte Anwendung des Powell-Verfahrens bei verschiedenen Initialisierungen oder die Nachoptimierung unter Verwendung unterschiedlicher Fehlermaße wiesen keine zufriedenstellende Konvergenz der Schätzwerte auf und erhöhten oft beträchtlich die Rechenzeit, so daß sie in den systematischen Untersuchungen keine Verwendung fanden.

Tabelle 2.5: Übersicht über die verwendeten Optimierungsmethoden.

Bezeichnung	Erklärung	Fehler	Beschreibung
A	„ <i>amoeba</i> “	ϵ_{IF}	Simplexmethode konstant initialisiert
MA_NN	„ <i>multi-amoeba</i> , nicht nachoptimiert“	ϵ_{IF}	Simplexmethode dreifach durchgeführt, initialisiert durch Gewinner der vollständigen Suche
MA	„ <i>multi-amoeba</i> “	ϵ_{IF}	Simplexmethode dreifach durchgeführt, initialisiert durch Gewinner der vollständigen Suche
		ϵ_{lowf}	Nachoptimierung durch Powell-Methode
MA_LF	„ <i>multi-amoeba</i> , <i>low frequency</i> “	ϵ_{lowf}	Simplexmethode dreifach durchgeführt, initialisiert durch Gewinner der vollständigen Suche
RD	R_d	ϵ_{IF}	Eindimensionale Optimierung unter Verwendung von R_d

3. Ergebnisse

3.1 Signaltyp FLT und SYN

3.1.1 Bestimmung der Schwelle σ in der Berechnung von OQ

In der Berechnung von OQ ist die Schwelle $\sigma(\varepsilon) = \varepsilon \cdot U_{\max}$ in Gleichungen (2.12) und (2.13) noch unbestimmt. Zur Ermittlung ihres Wertes wurden 250 LF-Konstellationen zufällig initialisiert und anhand dieser LF-Werte Signale des Typs FLT und SYN generiert, die daraufhin mit der SIM-Methode analysiert wurden. In Abhängigkeit von ε wurde die Rangkorrelation ρ sowie der mittlere Abstand δ zwischen geschätztem und „korrektem“ (d.h. dem mit demselben ε -Wert aus dem vorgegebenen LF-Modell berechneten) Öffnungsquotienten für beide Formulierungen (2.12) und (2.13) untersucht. Die Ergebnisse sind in den Abbildungen 3.2 und 3.1 dargestellt.

Die Abweichungen δ zwischen „korrekten“ und geschätzten Werten weisen für beide Signalarten ein deutliches Extremum bei $\varepsilon = 0.02$ auf. Damit einhergehend ist eine leichte Erhöhung der ohnehin schon sehr hohen Korrelationen in diesem Bereich zu beobachten, so daß der Wert von $\varepsilon = 0.02$ für alle nachfolgenden Berechnungen verwendet wurde.

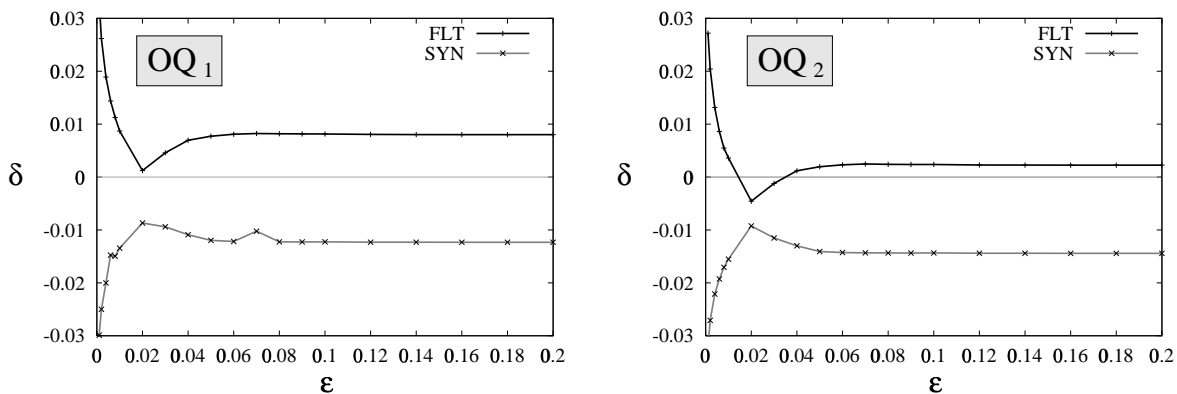


Abbildung 3.1: Mittlerer Abstand δ für OQ_1 (links) und OQ_2 (rechts) zwischen korrekten und geschätzten Werten für Signaltyp FLT und SYN basierend auf 250 Signalen für jeden Typ. Negative Werte bedeuten kleinere Werte des geschätzten Wertes.

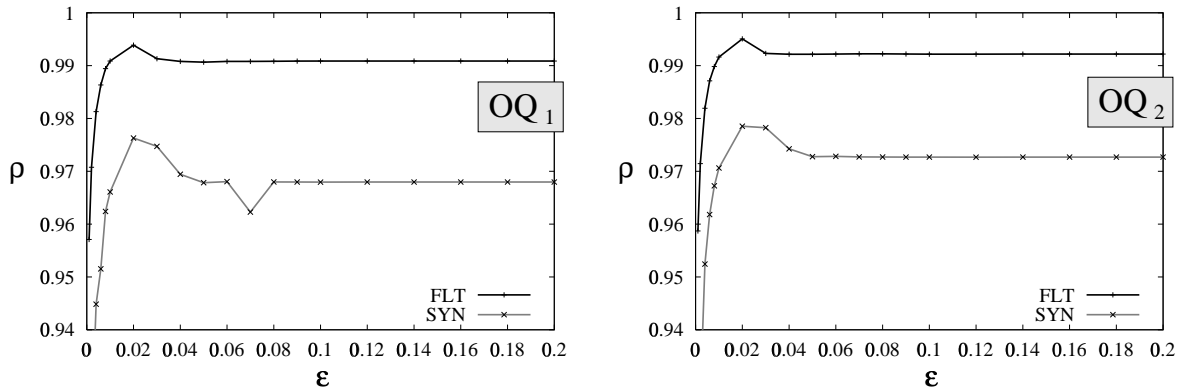


Abbildung 3.2: Rangkorrelation ρ für OQ_1 (links) und OQ_2 (rechts) zwischen korrekten und geschätzten Werten für Signaltyp FLT und SYN basierend auf je 250 Signalen mit denselben, zufällig generierten LF-Konfigurationen für jeden Typ.

3.1.2 Streuung der Schätzwerte

Zur Untersuchung der Streuung der Ergebnisse bei der Analyse desselben Signals in verschiedenen Abschnitten wurden 16 verschiedene LF-Konfigurationen zufällig erzeugt und jeweils ein Glottissignal von 1s Dauer generiert, wobei die Grundfrequenz $f_0 = 107\text{Hz}$ fest vorgegeben war. Dieses Signal wurde mit einem Nulpole-Filter 11. Ordnung, das die Resonanzen des Vokals /a:/ nachbildet, gefiltert (Signaltyp FLT). Jedes dieser 16 Signale wurde im Abschnitt 200ms bis 900ms bei der generell verwendeten Fensterlänge $l_w = 200\text{ms}$ und einem Fenstervorschub $\delta l_w = 100\text{ms}$ analysiert. Hierdurch ergaben sich für jedes Signal 6 Meßwerte, deren Variabilitätskoeffizienten $v = \sigma/\mu$ (Standardabweichung σ bezogen auf den Mittelwert μ) für die verschiedenen LF-Parameter in Tabelle 3.1 aufgeführt sind. Dieselben 16 LF-Konfigurationen wurden außerdem verwendet, um den Vokal /a:/ mit dem Sprachsynthesator zu erzeugen (Signaltyp SYN). Die hieraus erhaltenen Variabilitätskoeffizienten sind in Tabelle 3.2 angegeben.

Beim Vergleich der Signalarten wird ersichtlich, daß in manchen Fällen deutliche Abweichungen in der Konstanz der Schätzung – ausgedrückt durch große v – auftreten. Als Beispiel sei Puls 3 herausgegriffen: Für Signaltyp FLT ergaben sich Schätzwerte, die für alle LF-Parameter in allen 6 Fenstern sehr konsistent waren. Dagegen weist das Signal SYN (mit denselben Parameterkonfigurationen) deutlich höhere Variabilitätskoeffizienten v auf.

Tabelle 3.1: Variabilitätskoeffizienten v für die Schätzung der LF-Parameter. Die Variabilitätskoeffizienten beschreiben die Variation in der Schätzung für 16 zufällig initialisierte Signale des Typs FLT (Gesamtlänge eines Signals 700ms, Fensterlänge 200ms, Fenstervorschub 100ms).

	1	2	3	4	5	6	7	8
t_p	< 0.1%	1.0%	< 0.1%	< 0.1%	0.9%	8.4%	5.1%	1.3%
t_e	< 0.1%	0.3%	< 0.1%	< 0.1%	0.7%	6.4%	4.0%	0.9%
t_a	< 0.1%	1.8%	< 0.1%	< 0.1%	2.4%	3.8%	4.3%	7.2%
E_e	0.4%	0.9%	< 0.1%	< 0.1%	0.9%	0.9%	1.5%	3.2%
$t_e - t_p$	< 0.1%	1.7%	< 0.1%	< 0.1%	0.3%	2.0%	0.3%	4.0%
	9	10	11	12	13	14	15	16
t_p	0.1%	0.2%	< 0.1%	1.9%	1.6%	< 0.1%	< 0.1%	5.9%
t_e	0.3%	< 0.1%	< 0.1%	1.1%	0.6%	< 0.1%	< 0.1%	4.3%
t_a	4.7%	3.0%	< 0.1%	2.9%	6.2%	< 0.1%	< 0.1%	9.6%
E_e	0.1%	1.2%	< 0.1%	0.9%	1.0%	0.1%	< 0.1%	3.4%
$t_e - t_p$	1.9%	0.7%	< 0.1%	1.6%	3.0%	< 0.1%	< 0.1%	4.2%

Tabelle 3.2: Variabilitätskoeffizienten v für die Schätzung für dieselben LF-Konfigurationen wie in Tabelle 3.1 für den Signaltyp SYN (Gesamtlänge eines Signals 700ms, Fensterlänge 200ms, Fenstervorschub 100ms).

	1	2	3	4	5	6	7	8
t_p	< 0.1%	7.1%	13.3%	< 0.1%	< 0.1%	< 0.1%	3.6%	< 0.1%
t_e	< 0.1%	6.7%	13.0%	< 0.1%	< 0.1%	< 0.1%	3.7%	< 0.1%
t_a	< 0.1%	15.7%	11.3%	< 0.1%	< 0.1%	< 0.1%	22.3%	< 0.1%
E_e	< 0.1%	5.7%	5.6%	< 0.1%	< 0.1%	< 0.1%	8.0%	< 0.1%
$t_e - t_p$	< 0.1%	3.3%	11.5%	< 0.1%	< 0.1%	< 0.1%	4.2%	< 0.1%
	9	10	11	12	13	14	15	16
t_p	4.0%	< 0.1%	0.8%	3.2%	< 0.1%	13.6%	< 0.1%	< 0.1%
t_e	4.0%	< 0.1%	0.9%	2.7%	< 0.1%	13.1%	< 0.1%	< 0.1%
t_a	13.8%	< 0.1%	1.5%	14.5%	< 0.1%	7.5%	< 0.1%	< 0.1%
E_e	3.7%	< 0.1%	0.5%	4.6%	< 0.1%	1.3%	< 0.1%	< 0.1%
$t_e - t_p$	3.6%	< 0.1%	1.4%	1.3%	< 0.1%	9.0%	< 0.1%	< 0.1%

Die Ursache hierfür liegt in den Schätzergebnissen in den einzelnen Fenstern, die in Tabelle 3.3 aufgeführt sind. Diese aufgeschlüsselten Werte lassen erkennen, daß in den ersten vier Fenstern konsistent eine bestimmte Konfiguration gefunden wurde, jedoch in Fenstern 5 und 6 eine davon abweichende. Die Änderung der LF-Parameter führt ebenfalls zu einer Erhöhung der Variabilitätskoeffizienten der „abgeleiteten“ Parameter, die mit Ausnahme von SQ_1 in derselben Größenordnung liegen wie für die direkten LF-Parameter. Der Fehler err ist in Fenstern 5 und 6 geringfügig höher, doch ist diese Erhöhung relativ zum möglichen Wertebereich (vgl. Abb. 2.28) als sehr gering zu bewerten.

Tabelle 3.3: Analysewerte für das synthetisierte Signal Nr. 3 aus Tabelle 3.2. Jede Zeile gibt die Werte im k -ten Fenster wieder. Die zugehörigen Variabilitätskoeffizienten v sind in der letzten Zeile aufgeführt.

k	t_p	t_e	t_a	E_e	err	OQ_1	SQ_1	CQ_1	PSP
1	0.526	0.603	0.0174	1.34	0.3812	0.360	3.69	0.076	0.0087
2	0.526	0.603	0.0174	1.34	0.3813	0.360	3.69	0.076	0.0087
3	0.526	0.603	0.0174	1.34	0.3812	0.360	3.69	0.076	0.0087
4	0.526	0.603	0.0174	1.34	0.3813	0.360	3.69	0.076	0.0087
5	0.690	0.787	0.0135	1.19	0.3840	0.446	3.60	0.096	0.0120
6	0.690	0.787	0.0135	1.19	0.3840	0.446	3.60	0.096	0.0120
v	13.25%	13.03%	11.18%	5.63%	0.34%	10.42%	1.21%	11.51%	16.33%

3.1.3 Genauigkeit der Schätzung

Parameter aus dem angepaßten LF-Modell

Zur Beurteilung der Genauigkeit des Verfahrens in der Schätzung der LF-Parameter wurden 500 LF-Konfigurationen zufällig erzeugt und unter Verwendung verschiedener Optimierungsmethoden analysiert. Ergebnisse der Schätzung sind als Rangkorrelationen der Werte mit den korrekten Parameterwerten (im folgenden einfach als „Korrelationen“ bezeichnet) in Tabelle 3.4, als durchschnittliche relative Betragsabweichungen in Tabelle 3.5 zusammengestellt. Den Berechnungen der Abweichungen wurden bei Parameter t_a die logarithmierten Werte $\ln(t_a)$ zugrundegelegt. Für E_e wurden die Abweichungen nicht berechnet, da die geschätzten Werte aufgrund der Filterung nicht die ursprünglichen Größen E_e approximieren, sondern linear transformierte Werte. Da in den geschätzten Werten bei diesem (unbeschränkten) Parameter gelegentlich Ausreißer auftraten, ließ sich eine sinnvolle Definition des mittleren Abstands zwischen Schätz- und Originalwert nicht in einfacher Weise finden.

Tabelle 3.4: Rangkorrelation für die einzelnen LF-Parametern zwischen geschätzten und korrekten Werten bei verschiedenen Optimierungsmethoden und Prädiktionsordnungen p für Signaltyp FLT und SYN. Die Markierung der Signifikanzniveaus \mathbf{p} erfolgt gemäß Tabelle 2.4.

Optim.	p	t_p		t_e		t_a		E_e		$t_e - t_p$	
		FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN
MA	10	0.76	0.80	0.69	0.80	0.76	0.95	0.92	0.98	0.97	0.98
MA	13	0.71	0.87	0.61	0.80	0.03 [†]	0.61	0.76	0.86	0.80	0.97
MA_LF	10	0.77	0.82	0.68	0.79	0.73	0.80	0.90	0.96	0.96	0.97
MA_LF	13	0.69	0.85	0.62	0.80	0.13 ⁺	0.51	0.79	0.83	0.79	0.91
MA_NN	10	0.79	0.72	0.73	0.70	0.81	0.95	0.92	0.98	0.98	0.98
MA_NN	13	0.70	0.76	0.58	0.70	0.02 [†]	0.47	0.76	0.85	0.83	0.97
A	10	0.79	0.74	0.73	0.70	0.67	0.94	0.90	0.98	0.97	0.98
A	13	0.68	0.76	0.57	0.66	0.07 [†]	0.51	0.77	0.85	0.93	0.97

Tabelle 3.5: Mittlere relative Betragsdifferenzen in % der geschätzten bezogen auf die korrekte Werte bei verschiedenen Optimierungsmethoden und Prädiktionsordnungen p für 504 zufällig generierte LF-Konstellationen (Signaltyp FLT und SYN).

Optim.	p	t_p		t_e		t_a		$t_e - t_p$		Mittelwert	
		FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN
MA	10	15.9	14.1	19.0	16.8	13.0	5.5	12.2	5.9	16.02	11.28
MA	13	19.5	12.6	22.7	15.7	26.5	18.6	25.9	11.1	25.22	15.46
MA_LF	10	15.7	14.6	18.9	18.0	12.7	8.5	15.1	7.0	16.64	12.82
MA_LF	13	19.6	13.2	22.1	16.7	27.9	19.4	27.4	13.8	25.86	16.82
MA_NN	10	13.8	19.6	16.6	23.9	13.3	5.7	9.0	7.5	14.05	15.12
MA_NN	13	19.0	15.5	23.0	19.2	26.6	18.1	21.7	10.5	24.08	16.88
A	10	13.1	19.7	15.6	24.2	14.9	6.6	10.2	7.8	14.34	15.54
A	13	20.1	17.4	24.3	21.9	26.7	18.4	22.0	10.8	24.82	18.26

Für die „abgeleiteten“ Parameter OQ , SQ , CQ (jeweils in ihren beiden in Abschnitt 2.7 beschriebenen Formulierungen) und PSP wurden analog die Korrelationen und relativen Abweichungen zwischen den geschätzten Werten und den aus den Prototypsignalen ermittelten Referenzwerten berechnet. Diese Werte sind in Tabellen 3.6 und 3.7 aufgeführt.

Für Signaltyp FLT belegen die Ergebnisse, daß sich alle direkten LF-Parameter weitgehend unabhängig von der verwendeten Optimierungsmethode zuverlässig schätzen lassen. Generell ist auffällig, daß die Prädiktionsordnung $p = 10$ deutlich bessere Schätzungen ermöglicht als $p = 13$. Für t_a führt dies sogar so weit, daß sich bei Prädiktionsordnung 13 keine Korrelationen zwischen geschätzten und korrekten Referenzwerten – unabhängig von der Optimierungsmethode – zeigen. Ebenfalls auffällig ist, daß SYN-Signale eine bessere Schätzung der direkten LF-Parameter zulassen als die Signale FLT (mit denselben Parameterkonstellationen). Insbesondere für t_a und $t_e - t_p$ liegen die durchschnittlichen Abweichungen zwischen geschätzten und korrekten Werten deutlich unter 10%.

Bei Signaltyp SYN werden für Optimierungsmethoden MA und MA_LF in den Parametern $t_e, t_p, t_e - t_p$ und E_e etwas höhere Korrelationen gefunden, für t_a bleiben die Korrelationen bei MA_LF etwas hinter den anderen zurück. Bei der Schätzung von t_a und E_e ergeben sich deutlich höhere Korrelationen für $p = 10$, während für die anderen beiden LF-Parameter t_p, t_e sowie für ihre Differenz hohe bis sehr hohe Korrelationen, unabhängig von der Prädiktionsordnung, gefunden werden. Die mittlere Abweichung bei $t_e - t_p$ ist jedoch ebenfalls für $p = 10$ deutlich geringer. Die höchste durchschnittliche Korrelation für die verschiedenen Optimierungsmethoden ergibt sich für MA (0.902).

In den „abgeleiteten“ Parametern zeigt sich generell eine sehr gute Reproduzierbarkeit der ursprünglichen Werte, unabhängig von der speziellen Optimierungsmethode, wobei wieder die deutlich bessere Schätzung bei Prädiktionsordnung $p = 10$ gegenüber $p = 13$ ins Auge sticht. Für Signaltyp FLT kann der Öffnungsquotient äußerst zuverlässig mit einer mittleren Abweichung von 6-7%, die unabhängig von der Optimierungsmethode erreicht wird, geschätzt werden. Vergleichbar niedrige mittlere Abweichungen werden für Signaltyp SYN bei der Schätzung der beiden CQ -Parameter gefunden. Auffällig sind die relativ hohen mittleren Abweichungen in PSP für beide Signalarten und alle Optimierungsmethoden bei gleichzeitig äußerst hohen Rangkorrelationen.

Parameter aus dem IF-Signal

Bei der Berechnung der „abgeleiteten“ Parameter ist – insbesondere im Hinblick auf ihre oft praktizierte direkte Bestimmung aus dem IF-Signal – ebenfalls von Interesse, in wie weit auf diese Weise berechnete Größen mit den korrekten Werten korrelieren, bzw. davon abweichen. Die Ergebnisse dieser Untersuchung sind in Tabellen 3.8 und 3.9 dargestellt.

Unabhängig vom Signaltyp zeigt sich hierbei eine weitaus geringere Abhängigkeit von der Prädiktionsordnung p im Vergleich zu den Werten in Tabellen 3.6 – 3.9. Hierdurch bedingt sind insbesondere für $p = 13$ teilweise höhere Korrelationen zu finden

Tabelle 3.6: Rangkorrelation der „abgeleiteten“ Parameter mit den korrekten Werten bei verschiedenen Optimierungsmethoden und Prädiktionsordnungen p für Signaltyp FLT und SYN. Alle Werte sind hochsignifikant ($p < 0.001$).

Optim.	p	OQ_1		OQ_2		SQ_1		SQ_2		CQ_1		CQ_2		PSP	
		FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN
MA	10	0.99	0.98	0.99	0.98	0.88	0.91	0.93	0.92	0.97	0.98	0.99	0.99	0.99	0.99
MA	13	0.87	0.98	0.87	0.98	0.59	0.82	0.68	0.88	0.80	0.97	0.83	0.99	0.84	0.99
MA_LF	10	0.98	0.98	0.99	0.98	0.82	0.87	0.90	0.90	0.96	0.97	0.98	0.98	0.99	0.98
MA_LF	13	0.87	0.94	0.87	0.93	0.53	0.74	0.66	0.76	0.79	0.91	0.83	0.92	0.84	0.92
MA_NN	10	0.98	0.96	0.98	0.96	0.91	0.87	0.93	0.90	0.98	0.98	0.98	0.98	0.99	0.98
MA_NN	13	0.89	0.97	0.89	0.97	0.64	0.83	0.69	0.87	0.83	0.97	0.84	0.99	0.85	0.99
A	10	0.98	0.95	0.98	0.95	0.89	0.87	0.92	0.90	0.97	0.98	0.98	0.98	0.99	0.98
A	13	0.96	0.96	0.97	0.96	0.70	0.80	0.78	0.84	0.93	0.97	0.96	0.98	0.97	0.98

Tabelle 3.7: Mittlere relative Differenzen in % der „abgeleiteten“ Parameter bei verschiedenen Optimierungsmethoden und Prädiktionsordnungen p für 504 zufällig generierte LF-Konstellationen (Signaltyp FLT und SYN).

Optim.	p	OQ_1		OQ_2		SQ_1		SQ_2		CQ_1		CQ_2		PSP	
		FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN
MA	10	7.0	11.4	6.2	12.9	12.3	16.6	13.4	16.1	12.2	5.9	13.3	4.8	21.9	9.4
MA	13	13.6	11.4	11.7	13.0	19.3	18.5	21.6	10.5	25.9	11.1	24.8	11.5	43.2	18.1
MA_LF	10	7.3	11.2	6.3	11.2	15.4	19.1	14.2	16.3	15.1	7.0	13.4	5.0	20.9	8.5
MA_LF	13	14.1	11.0	11.5	12.6	21.3	24.1	20.2	15.3	27.4	13.8	23.5	11.9	41.8	17.1
MA_NN	10	6.2	13.2	6.1	13.1	10.0	19.6	11.5	17.3	9.0	7.5	11.5	6.0	19.1	12.5
MA_NN	13	11.7	13.7	10.7	14.8	18.3	19.4	19.9	12.6	21.7	10.5	21.1	11.7	36.1	19.6
A	10	6.2	13.7	5.9	13.7	10.7	19.9	11.0	17.3	10.2	7.8	10.9	6.1	18.8	13.4
A	13	11.4	14.7	9.6	15.8	17.8	21.7	17.2	14.4	22.0	10.8	18.6	11.3	34.6	20.5

Tabelle 3.8: Rangkorrelation zwischen den „abgeleiteten“ Parametern berechnet aus dem inversgefilterten Signal und den korrekten Werten bei verschiedenen Optimierungsmethoden und Prädiktionsordnungen p für Signaltyp FLT und SYN. Alle Werte sind hochsignifikant ($p < 0.001$).

Optim.	p	OQ_1		OQ_2		SQ_1		SQ_2		CQ_1		CQ_2		PSP	
		FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN
MA	10	0.82	0.95	0.79	0.96	0.87	0.90	0.87	0.88	0.97	1.00	0.95	1.00	0.37	0.47
MA	13	0.72	0.96	0.68	0.93	0.85	0.91	0.80	0.88	0.88	1.00	0.91	0.96	0.47	0.55
MA_LF	10	0.82	0.93	0.82	0.95	0.87	0.90	0.88	0.88	0.97	1.00	0.96	1.00	0.45	0.47
MA_LF	13	0.70	0.91	0.63	0.90	0.83	0.92	0.80	0.87	0.87	0.95	0.90	0.91	0.45	0.57
MA_NN	10	0.80	0.94	0.79	0.95	0.87	0.90	0.85	0.89	0.96	1.00	0.95	0.99	0.40	0.49
MA_NN	13	0.75	0.97	0.69	0.95	0.85	0.92	0.80	0.87	0.90	1.00	0.92	0.96	0.49	0.55
A	10	0.81	0.95	0.80	0.96	0.88	0.90	0.85	0.88	0.97	1.00	0.94	1.00	0.40	0.51
A	13	0.81	0.97	0.79	0.94	0.89	0.91	0.83	0.87	0.95	1.00	0.94	0.96	0.51	0.57

Tabelle 3.9: Mittlere relative Betragsdifferenzen in % zwischen den aus dem inversgefilterten Signal berechneten „abgeleiteten“ Parametern und den korrekten Werten bei verschiedenen Optimierungsdurchläufen und Prädiktionsordnungen p für 504 zufällig generierte LF-Konstellationen (Signaltyp FLT und SYN).

Optim.	p	OQ_1		OQ_2		SQ_1		SQ_2		CQ_1		CQ_2		PSP	
		FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN	FLT	SYN
MA	10	58.0	34.6	60.1	35.4	105.3	55.4	61.6	37.0	13.6	4.3	18.3	7.7	605.4	844.1
MA	13	59.1	32.1	66.5	41.6	143.2	69.8	76.2	40.1	20.8	11.7	20.0	18.8	557.0	661.6
MA_LF	10	57.4	37.3	72.3	36.9	99.8	58.0	68.5	38.3	12.9	4.7	17.4	7.6	599.4	1093.8
MA_LF	13	60.8	33.9	78.0	41.1	166.9	69.9	351.1	36.8	24.1	13.4	28.1	21.0	585.0	702.1
MA_NN	10	58.4	35.1	76.2	36.0	100.3	55.6	84.8	36.9	13.1	4.1	20.4	8.0	595.8	806.5
MA_NN	13	58.4	31.7	73.9	38.4	129.2	67.5	85.2	36.3	19.5	10.7	18.2	30.4	538.4	614.8
A	10	57.5	34.7	64.7	35.4	99.3	54.8	68.0	37.2	12.6	4.1	18.4	7.4	597.8	832.2
A	13	55.1	32.0	61.8	38.2	115.4	68.7	76.3	36.7	18.2	11.0	17.8	25.5	543.4	601.4

als bei der Ermittlung der Parameter aus dem angepaßten LF-Modell. Die mittleren Abweichungen sind jedoch mit der Ausnahme von CQ extrem hoch, so daß für diese Größen nicht mehr von einer zuverlässigen Schätzung gesprochen werden kann. Die Verschlußquotienten CQ sind dagegen, insbesondere für Signaltyp SYN, nach wie vor sehr zuverlässig bestimmbar.

Zusammenfassend läßt sich feststellen, daß für Signalarten FLT und SYN, d.h. für Signale, bei denen die Anregungsfunktion bereits durch das LF-Modell parametrisiert ist, eine sehr gute Reproduzierbarkeit der originalen Parameter zu finden ist. Die stärksten Abweichungen ergeben sich in der Bestimmung der genauen Lagen von t_p und t_e , die die entsprechenden Korrelationen auf „nur“ ca. 0.8 sinken lassen, wobei die Differenz $t_e - t_p$ davon unbeeinflusst äußerst zuverlässig geschätzt wird. Eine deutliche Überlegenheit einer einzelnen Optimierungsmethode konnte nicht festgestellt werden. Dagegen findet sich ein eindeutiger Unterschied in der Zuverlässigkeit der Schätzungen in Abhängigkeit von der Prädiktionsordnung p , wobei $p = 10$ eine gute Schätzung ermöglicht, $p = 13$ dagegen nicht. Bei der Ermittlung der „abgeleiteten“ Parameter aus dem IF-Signal anstatt aus dem angepaßten Modell zeigt sich, daß auf diese Weise berechnete Werte aller „abgeleiteten“ Parameter, mit der Ausnahme von CQ , viel höhere Unsicherheiten aufweisen. Interpretiert man die „abgeleiteten“ Parameter im Hinblick auf die Signalform, aus der sie berechnet wurden, so muß die Gestalt des IF-Signals als relativ unabhängig von der Prädiktionsordnung angesehen werden, während die des angepaßten LF-Modells von der Prädiktionsordnung beeinflusst wird.

3.2 Signaltyp 2MMS

3.2.1 Bestimmung der Referenzparameter

Zum Test, in wie weit durch die Zeitbereichsmethode (vgl. Abb. 2.16) zur Anpassung des LF-Modells an den durch das Zweimassenmodell generierten glottalen Fluß akkurate Schätzungen möglich sind, erfolgte eine „Zurückschätzung“ der LF-Parameter aus LF-Signalen. Hierzu wurden 500 zufällig initialisierte LF-Konstellationen generiert und für jede ein glottales Flußsignal der Länge 200ms erzeugt, dessen Parameter daraufhin mit der Zeitbereichsmethode geschätzt wurden. Die Ergebnisse sind in Tabelle 3.10 zusammengefaßt.

Die Werte verdeutlichen, daß die Reproduzierbarkeit der LF-Parameter auf diese Weise fast fehlerfrei erfolgt. Lediglich die Bestimmung von $t = 0$, d.h. des Zeitpunktes, bezüglich dessen t_p und t_e angegeben werden, ist einer höheren Schätzungsgenauigkeit unterworfen (mittlere Abweichung ca. 2.8%). Dies ist einerseits eine Folge des sehr langsamen Anstiegs von $U(t)$ bzw. $E(t)$ zu Beginn der Periode für große Werte von t_p und t_e , der zu erhöhten Schätzwerten des Periodenbeginns $t = 0$ führt (mit der daraus resultierenden Unterschätzung von t_p und t_e), andererseits ein Resultat der durch die Abtastrate begrenzten Auflösung, die für diesen Punkt nicht durch Interpolation verbessert werden kann. Insgesamt wurde die prinzipiell mögliche Genauigkeit der Anpassung nach dieser Methode als gut bewertet.

Tabelle 3.10: Zeitbereichsschätzung des Flußsignals für 500 zufällig initialisierte LF-Konfigurationen. Die mittlere relative Betragsdifferenz $\delta_{|\cdot|}$, die mittlere relative Abweichung δ (Minuszeichen: Referenz < Schätzung) sowie die Rangkorrelationen ρ zwischen Referenz und Schätzung sind für die vier LF-Parameter und die Differenz $t_e - t_p$ angegeben.

	t_p	t_e	t_a	E_e	$t_e - t_p$
$\delta_{ \cdot }$	2.68%	2.80%	0.25%	0.06%	0.15%
δ	-0.64%	-0.79%	0.23%	< 0.01%	-0.15%
ρ	0.98	0.97	1.00	1.00	1.00

3.2.2 Abhängigkeit von θ

Die Abhängigkeit der Schätzwerte vom Anstellwinkel θ im modifizierten Zweimasenmodell wurde ermittelt, indem 1s lange Abschnitte des Vokals /a:/ bei systematischer Variation des Anstellwinkels θ ($\theta \in [0.01^\circ, 1.96^\circ]$, $\delta\theta = 0.05^\circ$) synthetisiert und analysiert wurden. Werte für $\theta > 2^\circ$ ergaben Signale, bei denen kein LF-Modell an den Fluß zur Festlegung des Referenz-LF-Modells angepaßt werden konnte, da sie zunehmend sinusförmige Gestalt aufwiesen und die Schwingungen nicht stationär waren, sondern nach und nach „ausstarben“. Die Signale wurden in nichtüberlappenden, 200ms langen Fenstern mit Prädiktionsordnung $p = 13$ analysiert. Die weiterführenden Analysen basieren auf den Medianwerten aus den resultierenden 5 Fenstern. Als Referenz wurde an jedes Signal das LF-Modell direkt an den glottalen Fluß mit Hilfe der Zeitbereichsmethode nach Abbildung 2.16 angepaßt.

Tabelle 3.11: Rangkorrelation der Medianwerte des im Zeitbereich angepaßten Referenz-LF-Modells zu θ für den gesamten Bereich und für $\theta > 0.85^\circ$ (Signaltyp 2MMS, Prädiktionsordnung $p = 13$). Die Markierung der Signifikanzniveaus **p** erfolgt gemäß Tabelle 2.4.

Bereich	t_p	t_e	t_a	E_e	$t_e - t_p$			
$\theta > 0$	0.78	0.76	0.57	-0.79	< 0.01 [†]			
$\theta > 0.85^\circ$	0.95	0.77	0.95	-0.62*	-0.54*			
	OQ_1	OQ_2	SQ_1	SQ_2	CQ_1	CQ_2	PSP	
$\theta > 0$	0.76	0.77	0.82	0.96	< 0.01 [†]	0.31 [†]	0.68	
$\theta > 0.85^\circ$	0.75 ⁺	0.83	0.99	0.97	-0.54*	-0.30 [†]	0.01 [†]	

Korrelationen zwischen Referenzwerten und θ

Die Abschätzung, in wie weit sich durch Variation von θ systematische Veränderungen einzelner Parameter ergeben, erfolgte durch Rangkorrelation der verschiedenen Parameter des direkt an den glottalen Fluß angepaßten LF-Modells mit θ . Die geschah einerseits für den gesamten untersuchten Wertebereich von θ , andererseits in dem eingeschränkten Bereich $\theta > 0.85^\circ$. Die Ergebnisse sind in Tabelle 3.11 aufgeführt.

Bei der Untersuchung der Parameter zeigte sich, daß sehr deutliche (bei E_e negative) Korrelationen mit θ bestehen. Die Kurvenverläufe von $t_p(\theta)$ und $t_e(\theta)$ wiesen jedoch einen Einbruch im Bereich $[0.56^\circ, 0.81^\circ]$ auf (Abb. 3.3). Bei detaillierterer Analyse der Signale zeigte sich, daß die Rückwirkungen des Vokaltrakts auf das Flußsignal dazu führten, daß im genannten Abschnitt der Periodenbeginn $t = 0$ sprunghaft bei einem höheren Wert ermittelt wurde, wodurch sich simultan t_p und t_e erniedrigen (Abb. 3.4). Dies erklärt die etwas geringeren Korrelationswerte dieser beiden Parameter zu θ in Tabelle 3.11.

Um diesen durch die Anpassungsmethode bedingten, systematischen Veränderung der Werte im Bereich $[0.56^\circ, 0.81^\circ]$ zu vermeiden, wurden die Korrelationen ebenfalls lediglich für den Bereich $\theta > 0.85^\circ$ bestimmt (Tabelle 3.11, jeweils untere Zeile). In diesem Fall sind die meisten Parameter, insbesondere jetzt auch t_a und die „abgeleiteten“

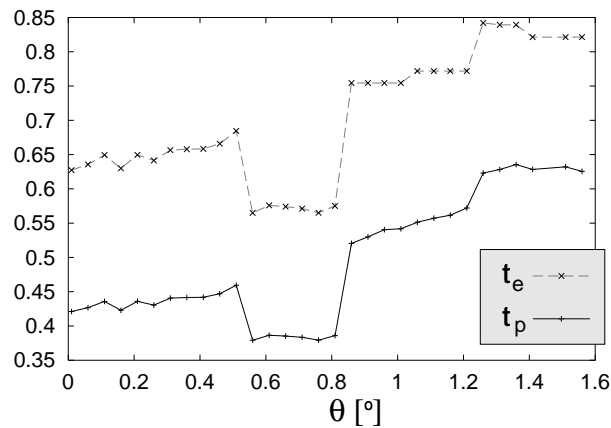


Abbildung 3.3: Abhängigkeit der Referenzparameter t_p und t_e von θ .

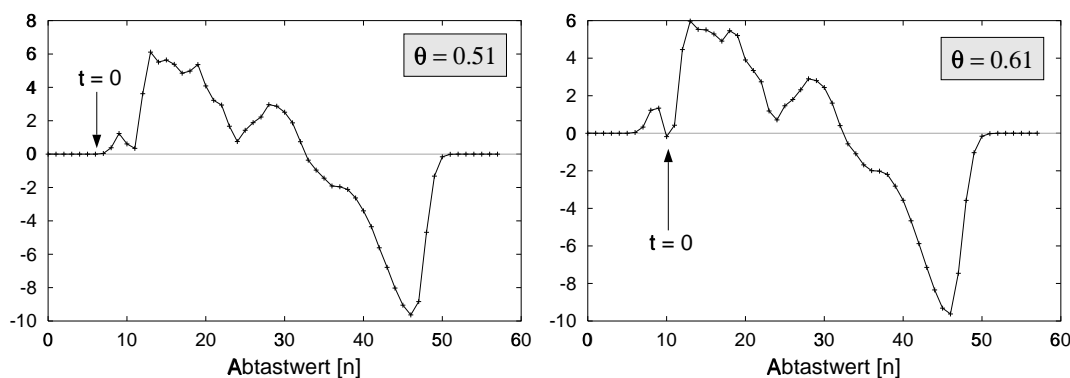


Abbildung 3.4: Eine Periode des mit dem erweiterten Zweimassenmodell generierten Flußsignals für $\theta = 0.51$ (links) und $\theta = 0.61$ (rechts).

Parameter OQ und besonders SQ , sehr hoch mit θ korreliert. Ebenfalls wird ein Unterschied zwischen t_p und t_e deutlich, da die Korrelation von t_p mit θ auf 95% ansteigt, die von t_e jedoch unverändert bei ca. 76% bleibt. Ausnahmen von den hohen Korrelationen stellen weiterhin $t_e - t_p$ sowie die Parameter CQ und PSP dar, die bestenfalls mäßig mit θ korrelieren und somit als weitgehend unabhängig von θ betrachtet werden dürfen.

Ergebnisse der SIM-Schätzung

Der Test der SIM-Anpassung in bezug auf die Güte der geschätzten Werte erfolgte einerseits im Vergleich zu den θ -Werten, andererseits zu den Referenzwerten der direkt an den Fluß angepaßten LF-Modelle. Die Ergebnisse der Rangkorrelationen sind in Tabelle 3.12 zusammengestellt.

Die Korrelationen zu θ zeigen, daß E_e relativ stark (negativ) mit dem wachsenden Öffnungswinkel korreliert, wie dies auch den mit zunehmendem θ deutlich schwächer werdenden Signalen entspricht (Rangkorrelation von θ mit der Signalenergie: -0.98). Die Differenz $t_e - t_p$ korreliert hoch mit θ für alle Optimierungsarten (etwas schwächer bei Optimierung MA_NN), obwohl die einzelnen Parameter t_p und t_e im allgemeinen nicht mit θ korrelieren. In Anbetracht der Referenzkorrelationen in Tabelle 3.11 (hohe Korrelation mit t_e und t_p , jedoch keine mit $t_e - t_p$) läßt sich dies als mangelhafte Schätzung einer der beiden Parameter deuten, während der andere Parameter in hinreichender Genauigkeit ermittelt wird. Der für den spektralen Abfall wichtige Parameter t_a korreliert jedoch kaum mit θ . Alle „abgeleiteten“ Parameter mit Ausnahme von OQ_2 zeigen deutliche Korrelationen mit θ , wobei die für SQ negativ ist. Generell zeigen sich keine deutlichen Unterschiede zwischen den verschiedenen Optimierungsmethoden.

Bei der Untersuchung der Korrelationen zwischen den geschätzten Werten nach dem SIM-Verfahren und den Referenzwerten (Tab. 3.12, rechte Seite) zeigt sich, daß die Werte für E_e in hoher Übereinstimmung durch das SIM-Verfahren geschätzt werden. Verschiedene Optimierungsmethoden ermöglichen ebenfalls die konsistente Schätzung anderer LF-Parameter: Wurde mit Methode A oder MA_NN optimiert, ergaben sich Korrelationen für t_e , wurde MA_LF angewendet, wies die Korrelation zu t_a signifikante Werte auf. Bei den „abgeleiteten“ Parametern wiesen OQ und PSP unabhängig von der Optimierungsmethode Korrelationen von 60-80% zu θ auf. CQ und SQ korrelierten jedoch kaum oder sogar signifikant negativ mit θ . In Anbetracht der durch die Werte in Tabelle 3.11 belegten Unabhängigkeit der Schließquotienten CQ von θ ist dieses Ergebnis für diese Parameter nicht verwunderlich. Dagegen weisen die niedrigen Korrelationen für SQ auf eine wenig vertrauenswürdige Schätzung dieses Parameters hin. Für Parameter E_e belegen die hohen Korrelationen unabhängig von der Optimierungsmethode jedoch die Robustheit der (nachgeschalteten) Schätzung dieses Parameters, die demnach weitgehend unabhängig von der genauen Wahl der anderen drei LF-Parameter zuverlässig vorgenommen werden kann.

Zur Untersuchung, warum sich keine Abhängigkeit zwischen θ und t_a ergab, wurden Signale, Signalspektren und Verläufe der Parameterwerte in Abhängigkeit von θ genauer untersucht. Hierbei wurde festgestellt, daß bezüglich vieler Parameter eine systematische Veränderung der Signale erst ab $\theta > 0.7^\circ$ auftritt, die in Abbildung 3.5 exempla-

Tabelle 3.12: Korrelation der Medianwerte der Inversfilterung zu θ und zur Referenz (Signaltyp 2MMS). Die Markierung der Signifikanzniveaus \mathfrak{p} erfolgt gemäß Tabelle 2.4.

	Korrelation zu θ				Korrelation zur Referenz			
	A	MA	MA_NN	MA_LF	A	MA	MA_NN	MA_LF
t_p	-0.01 [†]	-0.26 [†]	0.09 [†]	-0.33 [†]	0.13 [†]	-0.28 [†]	0.20 [†]	-0.39 [*]
t_e	0.52 ⁺	-0.18 [†]	0.51 ⁺	-0.13 [†]	0.67	-0.10 [†]	0.58	-0.21 [†]
t_a	-0.14 [†]	0.27 [†]	-0.09 [†]	0.42 [*]	-0.24 [†]	0.12 [†]	-0.33 [†]	0.61
E_e	-0.73	-0.73	-0.62	-0.76	0.89	0.90	0.81	0.95
$t_e - t_p$	0.80	0.82	0.68	0.86	0.18 [†]	0.13 [†]	0.09 [†]	0.05 [†]
OQ_1	0.70	0.71	0.69	0.83	0.74	0.63	0.70	0.71
OQ_2	0.71	0.73	0.70	0.86	0.78	0.66	0.74	0.73
SQ_1	-0.65	-0.62	-0.66	-0.61	-0.57	-0.55 ⁺	-0.56	-0.57
SQ_2	-0.64	-0.30 [†]	-0.13 [†]	-0.39 [*]	-0.65	-0.38 [*]	-0.22 [†]	-0.38 [*]
CQ_1	0.80	0.82	0.68	0.86	0.18 [†]	0.13 [†]	0.09 [†]	0.05 [†]
CQ_2	0.87	0.89	0.73	0.89	0.40 [*]	0.38 [*]	0.36 [*]	0.32 [†]
PSP	0.80	0.85	0.73	0.88	0.69	0.71	0.66	0.68

Tabelle 3.13: Korrelation der Medianwerte der Inversfilterung zu θ und zur Referenz für $\theta > 0.85^\circ$ (Signaltyp 2MMS). Die Markierung der Signifikanzniveaus \mathfrak{p} erfolgt gemäß Tabelle 2.4.

	Korrelation zu θ				Korrelation zur Referenz			
	A	MA	MA_NN	MA_LF	A	MA	MA_NN	MA_LF
t_p	0.19 [†]	-0.13 [†]	0.13 [†]	-0.13 [†]	0.30 [†]	-0.03 [†]	0.24 [†]	-0.07 [†]
t_e	0.33 [†]	-0.14 [†]	0.16 [†]	-0.16 [†]	0.44 [†]	0.01 [†]	0.32 [†]	-0.10 [†]
t_a	0.57 [*]	0.87	0.26 [†]	0.83	0.54 [*]	0.82	0.25 [†]	0.80
E_e	-0.53 [*]	-0.47 [†]	-0.56 [*]	-0.50 [†]	0.92	0.86	0.95	0.89
$t_e - t_p$	0.55 [*]	0.38 [†]	0.39 [†]	0.63 [*]	-0.10 [†]	0.04 [†]	-0.03 [†]	-0.19 [†]
OQ_1	0.48 [†]	0.21 [†]	0.38 [†]	0.32 [†]	0.55 [*]	0.35 [†]	0.45 [†]	0.30 [†]
OQ_2	0.52 [*]	0.39 [†]	0.41 [†]	0.64 [*]	0.62 [*]	0.55 [*]	0.61 [*]	0.55 [*]
SQ_1	-0.07 [†]	0.15 [†]	-0.12 [†]	0.01 [†]	-0.10 [†]	0.14 [†]	-0.12 [†]	0.04 [†]
SQ_2	-0.29 [†]	-0.50 [†]	-0.01 [†]	-0.53 [*]	-0.35 [†]	-0.48 [†]	0.01 [†]	-0.49 [†]
CQ_1	0.55 [*]	0.38 [†]	0.39 [†]	0.63 [*]	-0.10 [†]	0.04 [†]	-0.03 [†]	-0.19 [†]
CQ_2	0.62 [*]	0.60 [*]	0.54 [*]	0.91	0.09 [†]	-0.02 [†]	0.16 [†]	-0.21 [†]
PSP	0.59 [*]	0.62 [*]	0.54 [*]	0.87	0.38 [†]	0.27 [†]	0.37 [†]	0.08 [†]

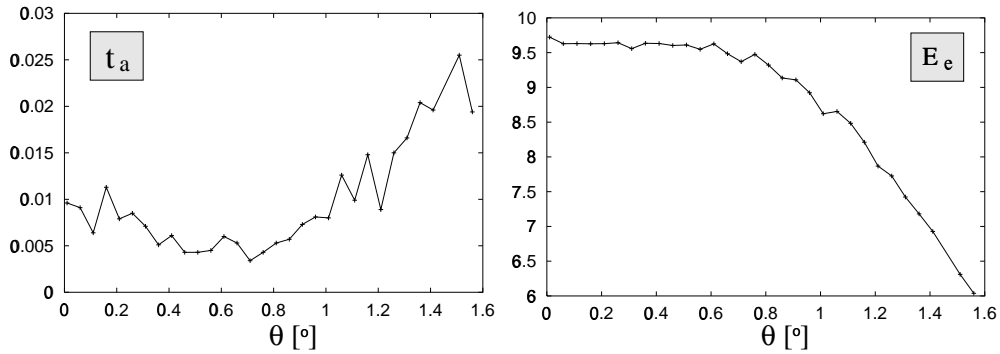


Abbildung 3.5: Abhängigkeit der Referenzparameter t_a und E_e von θ .

risch für die Parameter des Referenzmodells t_a und E_e dargestellt sind. Um die systematischen Veränderungen einiger Parameter im Bereich $\theta \in [0.56^\circ, 0.81^\circ]$ auszuschließen (vgl. Abb. 3.3), wurden die Analysen nur für den Bereich $\theta \in [0.86^\circ, 1.96^\circ]$, $\delta\theta = 0.05^\circ$ wiederholt. Die Ergebnisse sind in Tabelle 3.13 dargestellt.

Hierbei zeigte sich, daß sich nun bei den Optimierungen MA und MA_LF hohe Korrelationen von t_a mit θ ergaben. Dagegen sanken die Korrelationen mit θ für andere Werte, so auch für E_e , teilweise ab. Optimierung MA_LF führte zu sehr hohen Korrelationen bei CQ_2 und PSP . Für die Übereinstimmung zwischen Referenz- und geschätzten Parametern zeigte sich nun eine gute Übereinstimmung in den Parametern t_a und E_e , dagegen in allen anderen Parametern keine (Ausnahme: OQ_2 mit mäßiger Korrelation bei allen Optimierungsmethoden).

In Anbetracht der Untersuchungsergebnisse in Abschnitt 3.1, bei der eine deutlich bessere Schätzung bei Prädiktionsordnung $p = 10$ erfolgte, wurden dieselben Analysen für $p = 10$ wiederholt. Die Ergebnisse sind in Anhang C in Tabellen C.1 und C.2 detailliert aufgeführt. In diesem Fall ergaben sich jedoch keine Verbesserungen der Schätzungen, mit Ausnahme von E_e , dessen Korrelation zu θ sich für die verschiedenen Optimierungsmethoden auf -74% bis -79% im Betrag leicht erhöhte, einhergehend mit erhöhten Korrelationen zu den Referenzwerten von 89% bis 96%. Die „abgeleiteten“ Parameter wiesen jedoch bei dieser Prädiktionsordnung für Optimierung A um ca. 10% geringere Korrelationswerte bezüglich θ auf, während sich für die anderen Optimierungsarten überhaupt keine Korrelationen ergaben.

Zusammenfassend läßt sich sagen, daß durch die Einführung des Anstellwinkels θ im modifizierten Zweimassenmodell die Simulation verschiedener Phonationsarten möglich wurde, die sich in deutlichen Korrelationen zwischen θ und den verschiedenen „direkten“ und „abgeleiteten“ Parameterwerten des Referenzmodells widerspiegeln. Die Signalenergie wird konsistent durch den zuverlässig schätzbaren Parameter E_e erfaßt. Die hohen Korrelationen zwischen t_a und θ für das Referenzsignal bestätigen, daß sich mit der Einführung von θ insbesondere der spektrale Abfall, der eine

charakteristische Größe in der Simulation hypofunktioneller Stimmgebung darstellt, durch Veränderung der Abruptheit des Verschlusses modellieren läßt, wobei sich dieses Verhalten erst ab $\theta > 0.7^\circ$ deutlich äußert. Ausgehend von den vorliegenden Daten kann keiner der Optimierungsmethoden ein definitiver Vorzug gegeben werden, obwohl sich im eingeschränkten Bereich $\theta > 0.85^\circ$ Optimierungen MA und MA_LF durch hohe Korrelationen der geschätzten t_a -Werte sowohl mit θ als auch mit den Referenzwerten hervorhoben. Die ansonsten generell niedrigen Korrelationen zwischen mit der SIM-Methode geschätzten Modellparametern und den direkt an das Flußsignal angepaßten Referenzmodellparametern deuten darauf hin, daß die verschiedenen Anpassungsmethoden für Referenzmodell einerseits und das Modell innerhalb der SIM-Methode andererseits als nicht äquivalent zu betrachten sind. Dies ist angesichts der teilweise stark vom LF-Modell abweichenden Flußformen (vgl. Abb. 2.8) wenig verwunderlich, weist aber auf den Unterschied des SIM-Ansatzes im Hinblick auf andere Anpassungsmethoden im Zeitbereich (z.B. Strik et al., 1993) hin.

3.3 Signaltyp 2MMF

Die Verwendung von Signalen des Typs 2MMF bedeutet wieder eine Annäherung an das Quelle-Filter-Prinzip, da hier die Resonanzen des Vokaltrakts durch ein Nurple-Filter 11. Ordnung beschrieben werden, nachdem der anregende Fluß innerhalb des Sprachsynthesemodells unter Verwendung des selbstschwingenden Zweimassenmodells generiert wurde. Die Ergebnisse, die sich in völliger Analogie zu den in Abschnitt 3.2 gewonnenen Analysen für Signaltyp 2MMS ergaben, sind im Anhang C in Tabellen C.3 – C.6 aufgeführt.

Es zeigt sich, daß die geschätzten Parameter t_p und t_e mäßig negativ mit θ korrelierten (-0.63 bis -0.71 für Optimierungsmethoden A, MA, MA_NN). Die „abgeleiteten“ Parameter wiesen ähnliche (negative) Korrelationen in denselben Größenordnungen für Methoden MA, MA_NN, MA_LF auf. Lediglich SQ_1 zeigte höhere (positive) Korrelationen für die Optimierungsmethoden MA und MA_NN (0.81 bzw. 0.83). Bei Einschränkung auf $\theta > 0.85^\circ$ sanken alle vormaligen Korrelationen ab. Lediglich t_a korrelierte bei Verwendung von Methode A noch nennenswert (0.71) mit θ und wies auch eine Korrelation von 65% zu den Referenzwerten auf. Die Schätzung des Parameters E_e konvergierte für alle θ auf denselben Wert, so daß keine Rangkorrelation durchgeführt werden konnte.

Wie für Signaltyp 2MMS wurde die gesamte Analyse bei Prädiktionsordnung $p = 10$ wiederholt. Hierbei ergaben sich weiterhin negative Korrelationen zu θ um -0.6 für t_p und t_e bei den Optimierungen A und MA. Deutlich waren nun die positiven Korrelationen zwischen t_a und θ (A: 0.74, MA_LF: 0.79, MA: 0.91). Die hierdurch ausgedrückte Zuverlässigkeit der t_a -Schätzung äußerte sich auch in den Korrelationen zwischen geschätzten t_a - und Referenzwerten (MA: 0.59, MA_LF: 0.69), die für den eingeschränkten Bereich $\theta > 0.85^\circ$ auf 0.72 bzw. 0.76 anstiegen. Bei den „abgeleiteten“ Parametern wurden keine deutlichen Korrelationen mit θ oder den entsprechenden Referenzwerten beobachtet. Parameter E_e ließ sich weiterhin nicht ermitteln.

Für Signaltyp 2MMF zeigt sich, daß die SIM-Methode nur eingeschränkt „passen-

de“ Werte einer LF-Modellanpassung liefert. So kann E_e gar nicht geschätzt werden, während sich für viele Parameter inkonsistente Schätzwerte, d.h. niedrige Korrelationswerte zu θ ergeben. Eine genaue Ursache für das Konvergieren des E_e -Wertes auf denselben Punkt konnte nicht ermittelt werden. Jedoch zeigte sich insbesondere bei Methode MA mit Prädiktionsordnung $p = 10$ eine hohe Korrelation zwischen θ und den aus dem angepaßten LF-Modell geschätzten t_a -Werten, die ein charakteristisches Signalmerkmal (nämlich den spektralen Abfall) widerspiegeln. Diese Zuverlässigkeit äußert sich ebenfalls in den deutlichen Korrelationen der Schätzwerte mit den Referenzwerten für diesen Parameter.

3.4 Signaltyp ECHT

3.4.1 Unterschiede zwischen Phonationsarten

Zur Beurteilung des SIM-Verfahrens bei der Analyse echter Stimmproben in verschiedenen Phonationsarten (Signaltyp ECHT) wurden zunächst die Verschiedenheiten der Mittelwerte für die Verteilungen, die sich für die unterschiedlichen Parameter ergaben, durch den t -Test statistisch überprüft. Die Analyseergebnisse für die verschiedenen Vokale (/ε:/ gegenüber /a:/) wiesen nur vereinzelt Unterschiede auf, so daß für die weiterführenden Interpretationen alle Vokale einer Aufnahme zusammen analysiert wurden. Die Ergebnisse der Analyse für alle fünf Aufnahmen gemeinsam sind in Anhang C in Tabellen C.7 – C.17 in Abhängigkeit von der Optimierungsmethode und von der Prädiktionsordnung aufgeführt.

Generell zeigten sich bei einer großen Zahl der LF- und „abgeleiteten“ Parameter signifikante Unterschiede zwischen den verschiedenen Phonationsarten. Bei t_p und t_e waren die deutlichsten Unterschiede unabhängig von der Optimierungsmethode für Ordnung 10 zu beobachten, wobei zusätzlich die Methode RD bei Ordnung 16 äußerst konsistente Ergebnisse aufwies. Die Verteilungen für t_a wiesen nur für hypofunktionelle Stimmgebung bei Ordnung 16 signifikant abweichende, konsistent erhöhte Werte auf. Auch bei diesem Parameter wurden weitere signifikante Unterschiede bei Verwendung von Optimierung RD mit der Prädiktionsordnung $p = 16$ gefunden. Parameter E_e wies vereinzelt Unterschiede auf, die jedoch kein deutliches Muster erkennen ließen. Die relativ geringe Anzahl der signifikanten Unterschiede zwischen den Phonationsarten für diesen Parameter spiegelt wider, daß die Aufnahmen der verschiedenen Stimmproben jeweils optimal ausgesteuert waren und somit keine absolute Kalibrierung der Signalenergie vorgenommen werden konnte. Bei den „abgeleiteten“ Parametern zeigten sich keine Unterschiede hinsichtlich verschiedener Optimierungsmethoden oder Prädiktionsordnungen. Normale und hypofunktionelle Stimmgebung nahmen bei den OQ -Parametern eine Sonderstellung ein, da sich außer für Optimierung RD in allen Fällen signifikante Unterschiede zu allen anderen Phonationsarten ergaben. Für SQ wurden geringfügig mehr signifikante Unterschiede bei Ordnung 16 für alle Optimierungsmethoden gefunden. Insbesondere Optimierung RD bei $p = 16$ führte auch hier zu generell signifikant verschiedenen Werten zwischen den Phonationsarten.

Unabhängig von der durch die große Anzahl signifikanter Unterschiede demonstrierten Anwendbarkeit der SIM-Methode zur Differenzierung zwischen verschiedenen Phonationsmechanismen ist die Deutung der qualitativen Beziehungen einzelner Parameter zueinander von Interesse. So kann für t_a eine Erhöhung bei hypofunktioneller Phonation erwartet werden, die durch die Ergebnisse der statistischen Auswertungen bestätigt wird. Dagegen läßt sich die Verringerung von t_a bei gepreßter Phonation gegenüber normaler Phonation generell nicht mit hoher Signifikanz bestätigen und zeigt sich zudem nur bei Prädiktionsordnung $p = 10$. Die erwartete Absenkung des Öffnungsquotienten bei gepreßter Phonation spiegelte sich in den Ergebnissen nicht wider, im Gegenteil wurde eher eine gelegentlich signifikante Erhöhung beobachtet. Der erwartungsgemäß höhere CQ bei gepreßter Phonation gegenüber normaler Phonation wurde jedoch für alle Methoden unabhängig von der Prädiktionsordnung (Ausnahmen: MA_NN bei $p = 16$), gegenüber hypofunktioneller Phonation für alle Methoden außer RD bei $p = 10$ gefunden.

Um herauszufinden, in wie weit die teilweise fehlenden Übereinstimmungen zwischen erwarteten und tatsächlichen Ergebnissen auf die gemeinsame Untersuchung von Stimmproben verschiedener Versuchspersonen zurückzuführen sind, wurden die Analysen ausschließlich mit den Daten der drei Aufnahmen desselben Sprechers (Sprecher 1) wiederholt. Hierbei stellte sich heraus, daß tatsächlich trotz der geringeren Datenmenge mehr signifikante Unterschiede zwischen verschiedenen Phonationsarten in den unterschiedlichen Parametern auftraten. Dies untermauert, daß die Stichprobenzahl für eine statistische Mittelung über verschiedene Testsprecher als zu gering zu betrachten ist. Prinzipielle Veränderungen der Ergebnisse, die abweichende Interpretationen nach sich ziehen würden, ergaben sich jedoch nicht.

3.4.2 Übergänge zwischen Phonationsarten

Basierend auf den Ergebnissen des vorigen Abschnitts wurden die Aufnahmen, bei denen eine Versuchsperson (Sprecher 1) die Stimme möglichst kontinuierlich von hypofunktioneller über normale zu gepreßter Phonation und in gleicher Weise wieder zurück zu hypofunktioneller Stimmgebung variierte, nach zwei der „besten“ Vorgehensweisen analysiert. Zur Ermittlung der erfolgversprechendsten Methoden wurde die höchste Anzahl an signifikanten Unterschieden zwischen den verschiedenen Phonationsarten betrachtet, woraus sich Optimierungsmethoden MA bei Ordnung 10 und RD bei Ordnung 16 als „Gewinner“ ergaben. Exemplarisch sind die Ergebnisse dieser Analysen für jeweils einen Übergang in Abbildung 3.6 für Optimierung MA und in Abbildung 3.7 für Optimierung RD dargestellt.

Es zeigte sich, daß bei beiden Methoden die Mehrzahl der Parameter den Übergang zwischen den Phonationsarten sehr deutlich widerspiegeln. Auch die Richtung der Veränderungen, d.h. die qualitative Beziehung zwischen Parametern und Phonationsart, deckt sich ausnahmslos mit den Erwartungen. So führt die hypofunktionelle Phonation zu einer Erhöhung der Verschußzeit $t_e - t_p$ bzw. CQ und der Rückkehrzeit t_a . Ebenfalls erhöht sind OQ und PSP , wohingegen E_e erwartungsgemäß bei kleinen Werten liegt. Umgekehrt sind bei gepreßter Phonation die Verschußzeiten und CQ -Werte

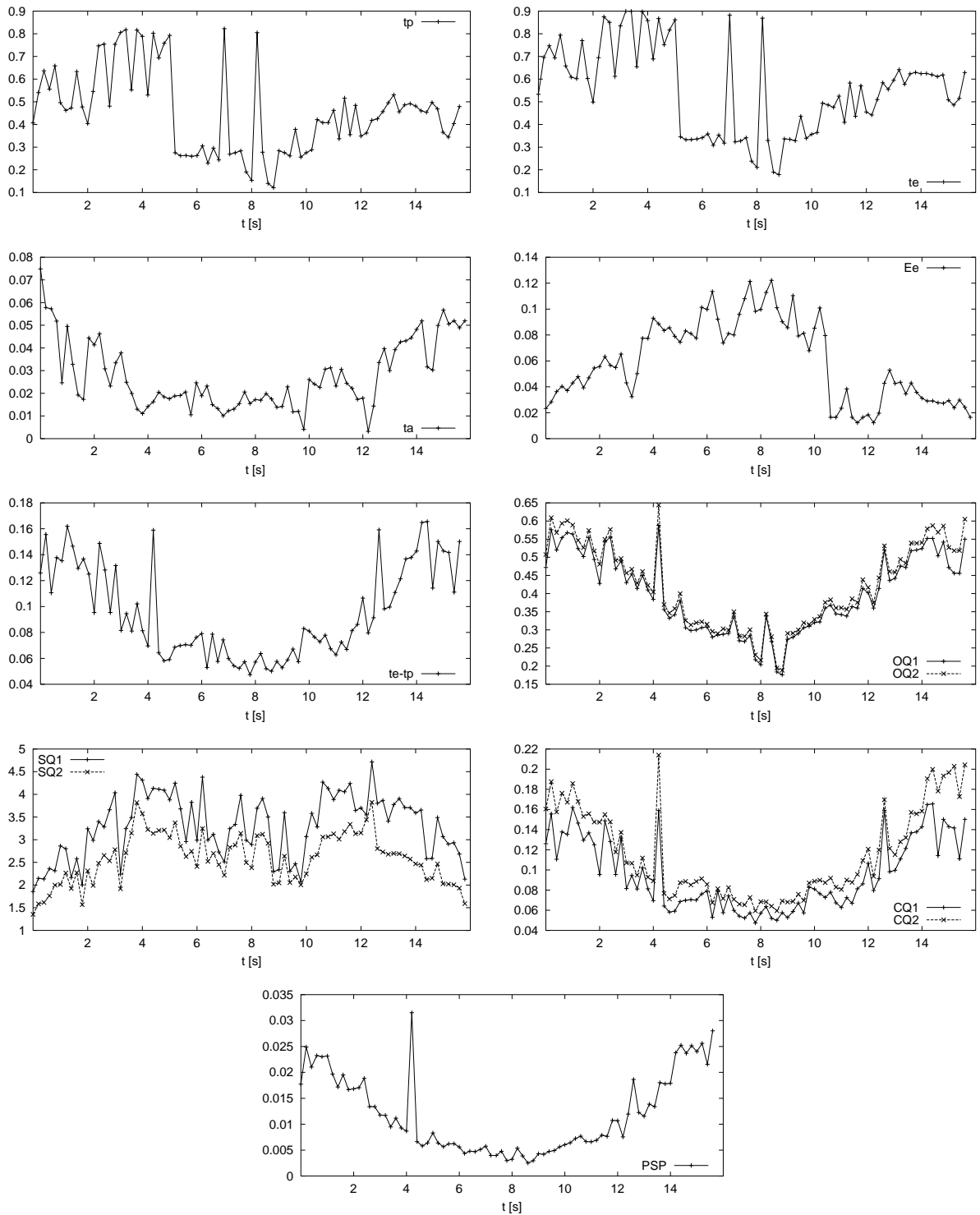


Abbildung 3.6: Ergebnisse der Parameterschätzungen bei Optimierungsmethode MA (Prädiktionsordnung $p = 10$) für den kontinuierlichen Übergang von hypofunktioneller Phonation zu gepreßter Phonation (um $t = 8$ s) und wieder zurück.

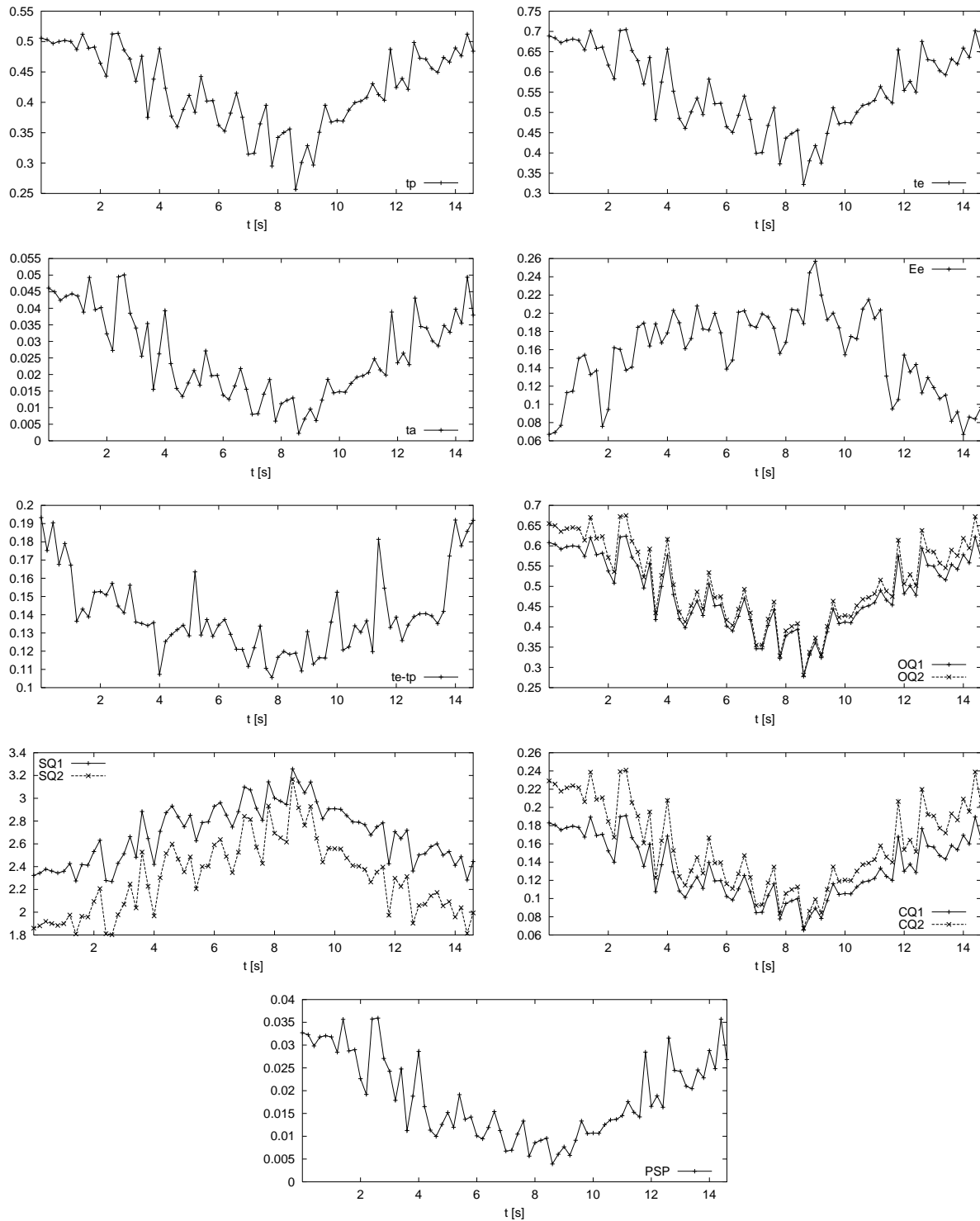


Abbildung 3.7: Ergebnisse der Parameterschätzungen bei Optimierungsmethode RD (Prädiktionsordnung $p = 16$) für den kontinuierlichen Übergang von hypofunktioneller Phonation zu gepreßter Phonation (um $t = 8\text{s}$) und wieder zurück.

niedrig, ebenfalls die Rückkehrzeiten t_a , was den harten Verschleiß korrekt wiedergibt. Öffnungsquotienten und PSP weisen niedrige Werte auf, während E_e bei hohen Werten liegt, die der erhöhten Signalenergie entsprechen (Rangkorrelation zwischen E_e und Signalenergie um 80% bei den verschiedenen Aufnahmen des Stimmübergangs).

3.4.3 Zusammensetzung des Datenraums

Singulärwertzerlegung

Um eine Information über die Beiträge der einzelnen Parameter in der Beschreibung des Datenraumes zu erhalten, wurde eine Hauptkomponentenanalyse durch Singulärwertzerlegung (*singular value decomposition*, SVD) durchgeführt. Dies wurde für die Analyseergebnisse aller Sprecher sowie jeweils einzeln für die verschiedenen Testpersonen vorgenommen. Der Eingangsdatenraum bestand hierbei aus N 10-dimensionalen Vektoren mit den Komponenten $t_p, t_e, \ln(t_a), OQ_1, OQ_2, SQ_1, SQ_2, CQ_1, CQ_2, PSP$ mit $N = 7188$ für alle Sprecher, $N = 4307, 1456, 1425$ für Sprecher 1,2,3. Parameter E_e wurde aufgrund der verschieden angesteuerten Aufnahmen von der Analyse ausgeschlossen. Zur Reduktion der Dimensionalität wurde eine Schranke von 95% für die Varianz des Eingangsdatenraums angesetzt.

In allen vier Fällen ergaben sich sehr übereinstimmende Hauptkomponenten, so daß im folgenden lediglich die Ergebnisse der Untersuchungen von Sprecher 1 dargestellt werden, da für ihn die meisten Datenpunkte vorlagen, jedoch die Beeinflussung der Ergebnisse durch die gemeinsame Auswertung verschiedener Sprecher ausgeschlossen wurde. Für diesen Sprecher ergibt sich ein vierdimensionaler Unterraum (Abb. 3.8), der durch die in Abbildung 3.9 dargestellten Eigenvektoren aufgespannt wird. In den

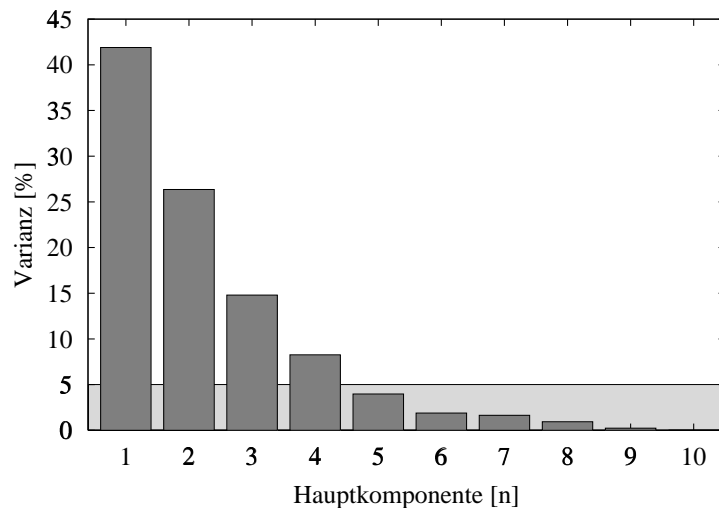


Abbildung 3.8: Beschreibung der Varianz des 10-dimensionalen Eingangsraums durch die Eigenvektoren (Hauptkomponenten) der SVD.

ersten, wichtigsten Eigenvektor gehen alle Parameter ein, wobei die „abgeleiteten“ Parameter etwas stärker repräsentiert sind als die direkten Parameter des LF-Modells. Dieser Vektor spiegelt die vergleichbaren Wichtigkeiten der Parameter einerseits, die Gegenläufigkeit von SQ bezüglich der anderen „abgeleiteten“ Parameter andererseits wider. Eigenvektor 2 besitzt eine starke Komponente in t_p , t_e und (diesmal mit demselben Vorzeichen) in SQ . Vektor 3 wird im wesentlichen durch t_a aufgespannt. Somit differenzieren Vektoren 2 und 3 zwischen den verschiedenen direkten Parametern des LF-Modells. Vektor 4 setzt sich aus den gleich orientierten Beiträgen von t_p und t_e sowie den dazu entgegengesetzten von SQ , PSP und, in etwas geringerem Maß, CQ zusammen.

Diskriminanzanalyse

Basierend auf der neuen, niedrigerdimensionalen Beschreibung des Datenraums wurden Fishersche Diskriminanzanalysen durchgeführt, um zu überprüfen, ob anhand der SVD-Zerlegung eine phonationsspezifische Charakterisierung der Daten vorgenommen werden kann. Die Untersuchungen ergaben, daß generell signifikante Diskriminationen zwischen den untersuchten Phonationsarten gefunden wurden, die sich als weitgehend

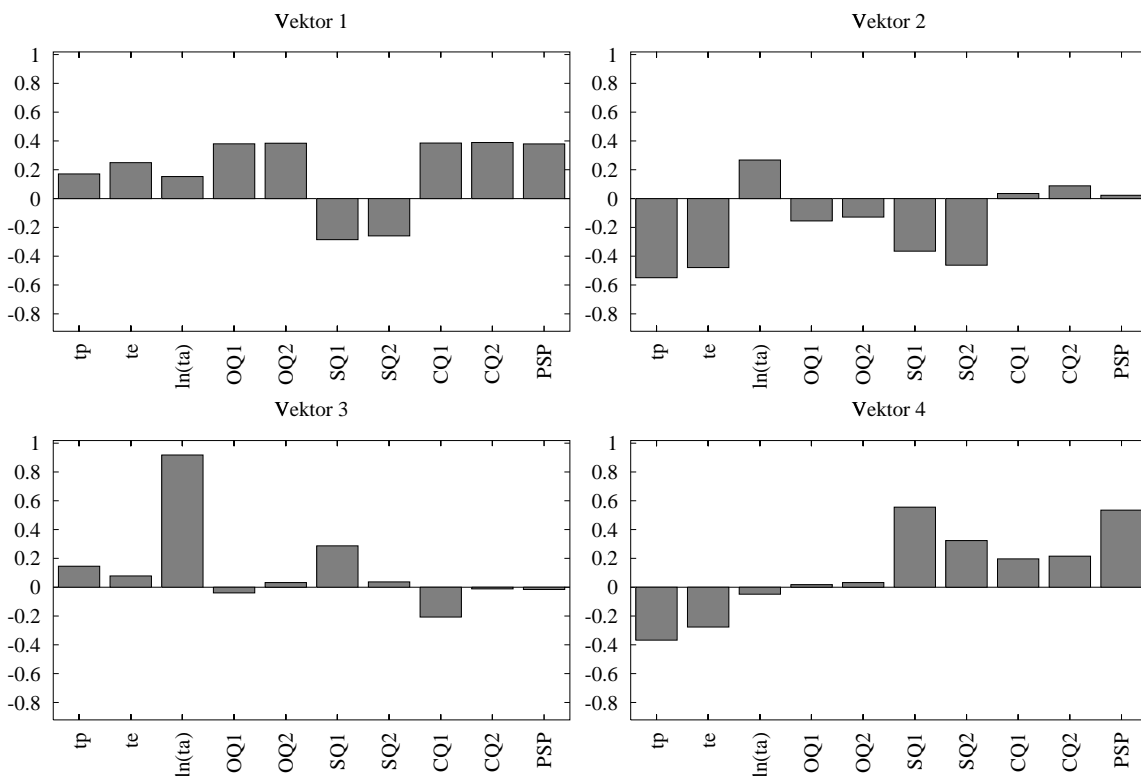


Abbildung 3.9: Zusammensetzung der Eigenvektoren aus der SVD (Numerierung entsprechend Abbildung 3.8).

unabhängig von der verwendeten Optimierung und der Prädiktionsordnung herausstellen. Exemplarisch sind die Ergebnisse für Optimierung MA_NN bei Prädiktionsordnung $p = 13$ in Tabelle 3.14 aufgeführt. In diesem Fall ergab die sukzessive Reduktion der Merkmale in der Analyse der verschiedenen Testsprecher nur zwei Phonationspaare, bei denen eine signifikante Diskriminierung auf der Grundlage von lediglich 2 der Vektoren möglich ist. Dies bestätigt die Ergebnisse der SVD, daß der zugrundeliegende Datenraum weitgehend als vierdimensional zu betrachten ist.

Bei der kombinierten Untersuchung der Daten aller Sprecher zeigte sich jedoch ein anderes Bild. Für die Optimierungsmethode MA wurde bei keiner der drei Prädiktionsordnungen $p = 10, 13, 16$ eine signifikante Diskriminierung zwischen gepreßter Phonation und einer anderen Phonationsart erreicht. Die Ergebnisse bei den anderen drei Optimierungsarten waren untereinander vergleichbar, so daß sie exemplarisch für MA_NN mit $p = 10$ (zu Vergleichszwecken mit Tabelle 3.14) und für RD mit $p = 16$ (zu Vergleichszwecken mit Abbildung 3.7) in Anhang C in Tabellen C.18 und C.19 aufgeführt sind. Die sukzessive Reduktion der Merkmale ergab dabei kein einheitliches Bild: So scheint im Fall von MA_NN ($p = 10$) Vektor 3 einen vergleichsweise schwächeren Beitrag zur Diskriminierung zu liefern, während für RD ($p = 16$) Vektor 4 eher diese Rolle zukommt. Bei der Untersuchung der anderen (nicht dargestellten) Ergebnisse liegen ähnlich schwer zu interpretierende Verhältnisse vor. Die Prädiktionsordnung 10 zeichnete sich aber für die drei Methoden A, MA_NN, RD generell durch eine höhere Güte der Diskriminationen im Vergleich zu den Ordnungen $p = 13$ oder 16 aus.

Zusammenfassend läßt sich sagen, daß die Beschreibung der verschiedenen Phonationsarten durch die Koordinaten bezüglich der SVD-Vektoren auch bei gemeinsamer Analyse aller Sprecher für viele Optimierungsarten eine signifikante Diskrimination zufriedenstellender Güte liefert. Im Vergleich zu den Ergebnissen für die Einzelpersonen bei Verwendung derselben Methode (MA_NN bei $p = 10$, Tabelle 3.14) ist die Güte für fast alle Phonationspaare jedoch deutlich abgesunken und für die Diskrimination gepreßt–laut nicht mehr signifikant. Dies läßt sich als Hinweis interpretieren, daß durch die SIM-Methode glottale Merkmale extrahierbar sind, die in beträchtlichem Maß sprecherspezifisch sind und so einen potentiellen Beitrag zur Binnendifferenzierung bestimmter Pathologiegruppen liefern können.

Tabelle 3.14: Ergebnisse der Fisherschen Diskriminanzanalyse für Optimierungsmethode MA_NN bei Prädiktionsordnung $p = 10$ auf der Grundlage der vierdimensionalen Datenbeschreibung anhand der Eigenvektoren aus der SVD. Angegeben ist die Güte der Trennung (signifikante Werte zum Signifikanzniveau $\mathbf{p} < 0.01$ sind durch \bullet markiert). In Klammern angegebene Ziffern bezeichnen die Indizes der Eigenvektoren, bei deren sukzessiver Weglassung (von links nach rechts) die Diskrimination weiterhin signifikant bleibt.

	gepreßt	laut	hypo
<u>Sprecher 1 (1)</u>			
normal	100.0% \bullet	100.0% \bullet (4, 3)	94.6% \bullet
gepreßt		89.3% \bullet	96.4% \bullet
laut			85.7% \bullet
<u>Sprecher 1 (2)</u>			
normal	75.0% \bullet	75.0% \bullet	83.3% \bullet
gepreßt		75.0% \bullet	79.6% \bullet
laut			79.6% \bullet
<u>Sprecher 1 (3)</u>			
normal	76.8% \bullet	96.4% \bullet	96.4% \bullet
gepreßt		94.6% \bullet	96.4% \bullet
laut			92.9% \bullet
<u>Sprecher 2</u>			
normal	92.9% \bullet	96.4% \bullet	96.4% \bullet
gepreßt		89.3% \bullet	94.6% \bullet
laut			100.0% \bullet (1, 2)
<u>Sprecher 3</u>			
normal	91.1% \bullet	73.2% \bullet	98.2% \bullet
gepreßt		76.8% \bullet	100.0% \bullet
laut			98.2% \bullet

4. Diskussion und Ausblick

Glottale Parameter, die nach der vorgestellten SIM-Methode berechnet werden, weisen je nach Art des untersuchten Testsignals unterschiedlich gute Übereinstimmungen mit den korrekten Referenzwerten auf. Für künstlich erzeugte Signale gilt, daß im Falle der Anregung mit einer Pulsform, die durch das LF-Modell bereits vorgegeben ist (Signaltyp FLT und SYN), die korrekten Werte generell mit hoher Genauigkeit reproduziert werden können. Eine gewisse Ausnahme bilden die Parameter t_p und t_e , deren absolute Lagen auch bei dieser Art von Signalen oft fehlerhaft geschätzt werden. Gründe hierfür sind einerseits die durch das Modell bedingte geringe Signalamplitude zu Beginn der glottalen Periode bei großen Werten dieser beiden Parameter, so daß ihre Variation nur geringfügige Änderungen der resultierenden Signalform bewirkt, was wiederum eine starke Sensitivität bezüglich geringfügiger Pulsformänderungen in der Umkehrfunktion bedeutet. Andererseits wird bei der Inversfilterung die Phase des Signals nicht explizit modelliert, so daß sich hierdurch ebenfalls Mehrdeutigkeiten insbesondere in der Schätzung dieser beiden Parameter ergeben können. Dies entspricht den Ergebnissen von Veldhuis, der durch Variationsrechnung und Hauptkomponentenanalyse ermittelte, daß für viele LF-Parameterkonstellationen t_a die höchste spektrale Relevanz besitzt, t_e und t_p (bzw. aus diesen Parametern abgeleitete Funktionen) dagegen weit weniger zur Beschreibung der Varianz beitragen und somit umgekehrt schlechter aus dem Spektrum geschätzt werden können (Veldhuis, 1998). Die Differenz $t_e - t_p$ ist durch dieses Problem jedoch nicht berührt und wird bei der Schätzung durch das SIM-Verfahren äußerst zuverlässig ermittelt. Die Beschreibung des Vokaltrakts durch ein AR-Modell stellt sich bei dieser Art Anregung als geeignet heraus, so daß auch bei Verwendung eines Vokaltraktmodells mit komplizierterer Übertragungsfunktion (Signaltyp SYN) die Parameter der ursprünglichen Pulsform gut geschätzt werden.

Anders stellt sich der Fall dar, wenn sich das anregende Signal nur näherungsweise durch eine LF-Parametrisierung beschreiben läßt (Signaltyp 2MMS und 2MMF). In diesen Fällen zeigt sich, daß nur wenige Parameter zuverlässig geschätzt werden. Ein Problem in der Auswertung dieser Art von Signalen liegt in der Bestimmung der Referenzpulsform, die durch Anpassung des LF-Modells im Zeitbereich an das Flußsignal erfolgte. Die Wahl dieser Methode bot sich aufgrund der speziellen Eigenschaften des Flußsignals an, wobei andere Autoren ähnliche Verfahren auch für die Anpassung des LF-Modells an echte Inversfiltersignale verwendet haben (Strik et al., 1993). Fehlende Korrelationen zwischen SIM-ermittelten Parametern und Referenzwerten können daher auch in diesem Vorgehen zur Definition des Referenzpulses begründet liegen. Bei

Anregung des Vokaltraktmodells mit der durch das selbstschwingende Zweimassenmodell generierten Pulsform (Signaltyp 2MMS) wurden kritische Parameter wie t_a und E_e dennoch mit hoher Zuverlässigkeit geschätzt. Diese Parameter korrelierten ebenfalls stark mit dem Ausmaß der simulierten Hypofunktion (realisiert durch wachsenden Anstellwinkel θ des selbstschwingenden Stimmlippenmodells).

Bei denselben Anregungspulsen ergab die Ersetzung des Vokaltraktmodells durch ein reines AR-Modell keine Verbesserung der Schätzwerte, obwohl in diesem Fall (Signaltyp 2MMF) die Inversfilterung theoretisch die fehlerfreie Rekonstruktion des Eingangssignals ermöglichen sollte. Die auf diese Weise erzeugten, äußerst artifiziellen Signale wiesen perzeptiv weiterhin den Charakter synthetisierter Vokale auf. Die mit der SIM-Methode geschätzten Werte der einzelnen Parameter führten jedoch, mit Ausnahme des Parameters t_a , zu inkonsistenten Schätzwerten. Dies ist in Anbetracht der Tatsache, daß das SIM-Verfahren bereits nicht zur Anpassung des LF-Modells an den durch das Zweimassenmodell generierten glottalen Fluß zur Definition der Referenzpulsform verwendet werden konnte, sondern durch eine Anpassung im Zeitbereich ersetzt werden mußte, nicht verwunderlich. Die lineare Filterung eines Signals, das dem angesetzten Glottismodell nur bedingt entspricht, kann zu einer Fehlertopologie führen, die unter den der Schätzmethode zugrundeliegenden Annahmen (Anregung durch eine Pulsform, die sich durch das LF-Modell parametrisieren läßt) kein ausgeprägtes Minimum besitzt, welches sich durch multidimensionale Extremwertsuche ermitteln ließe. In Anbetracht der Resultate bei den anderen Testsignalarten sollten die fehlenden Korrelationen für diesen Signaltyp als Hinweis auf die Voraussetzung der „Gutartigkeit“ des analysierten Signals bei Verwendung der SIM-Methode interpretiert werden. Positiv bewerten läßt sich selbst in diesem Fall die Zuverlässigkeit der Schätzung des Parameters t_a , der in der Literatur als schwierig zu schätzende Größe genannt wird (Fant, 1993).

Bei der Analyse echter Stimmproben zeigt sich ein hohes Maß an Differenzierungsmöglichkeiten zwischen verschiedenen Phonationsarten auf der Grundlage der berechneten Parameter. Am eindrucklichsten ist dies während der kontinuierlichen Veränderung der Phonation zwischen hypofunktioneller und gepreßter Stimmgebung zu beobachten. Untermauert von statistischen Untersuchungen weist t_a für hypofunktionelle Phonation deutlich erhöhte, für gepreßte Phonation erniedrigte Werte auf und steht damit im Einklang zu den Untersuchungen von Gobl und Chasaide (1992) sowie von Childers und Ahn (1995). Entgegen den Ergebnissen von Alku und Vilkmán (1996) stellt damit t_a einen direkt interpretierbaren Parameter zur Unterscheidung zwischen diesen beiden Phonationsarten dar. Das Absinken von E_e für hypofunktionelle Phonation bestätigt ebenfalls die Ergebnisse von Gobl und Chasaide (1992) und ist im Einklang mit der generellen Interpretation von E_e als Maß für die Energie der höheren Formanten (Fant, 1995) und damit der Gesamtenergie, mit der E_e hoch korrelierte. Das Abfallen von OQ und CQ beim Übergang von hypofunktioneller zu gepreßter Phonation spiegelt die physiologischen Veränderungen bei diesem Übergang wider und entspricht den Ergebnissen von Alku und Vilkmán (1996). Die zunehmende Asymmetrie bei demselben Übergang, die durch den anwachsenden Asymmetriequotient SQ ausgedrückt wird, wurde ebenfalls von Alku und Vilkmán (1996) sowie von Childers und Ahn (1995) beschrieben. Die relativen Veränderungen der drei „abgeleiteten“ Parameter OQ , CQ , SQ waren jedoch

in Alku und Vilkman (1996) deutlich geringer.

Der untersuchte hochdimensionale Raum aus 10 Parametern kann durch SVD auf einen vierdimensionalen Raum projiziert werden, in dem eine sehr hohe Differenzierbarkeit zwischen den verschiedenen Phonationsarten möglich ist. Der annähernd gleich starke Beitrag aller Parameter zum Eigenvektor mit dem höchsten Eigenwert kann als Entsprechung der Ergebnisse von Childers und Ahn (1995) aufgefaßt werden, die auf der Analyse lediglich der vier LF-Parameter basierten. Angesichts des deutlich höchsten Eigenwerts für diesen Vektor ist eine mögliche Interpretation, daß die verschiedenen Parameter nicht unabhängig voneinander variieren, sondern in Abhängigkeit von der Phonationsart typische Beziehungen zueinander aufweisen. Dies würde auch die guten Resultate bei der Optimierung auf der Grundlage des R_d -Parameters erklären, in dessen Definition bereits statistische Zusammenhänge zwischen Modellparametern für verschiedene Phonationsarten eingehen (Fant, 1995). Die Struktur der weiteren Eigenvektoren bestätigt die von Fant (1993) postulierte Wichtigkeit des Parameters t_a in der Unterscheidung zwischen den verschiedenen Phonationsarten.

Eine weitere Folgerung, die sich aus der Hauptkomponentenanalyse ergibt, ist, daß offensichtlich mehr als die ursprünglichen vier direkten Parameter des LF-Modells notwendig sind, um die Varianz der Daten in ausreichender Genauigkeit zu beschreiben. Die „abgeleiteten“ Parameter erfassen daher Stimmeigenschaften, die sich nicht durch lineare Transformationen aus den ursprünglichen vier LF-Parametern ableiten lassen, die ja ihrerseits bereits eine niedrigerdimensionale Einbettung aufweisen (Veldhuis, 1998). Die verschiedenen Definitionen der „abgeleiteten“ Parameter zeigen keine nennenswerten Unterschiede, so daß ihre genaue Definition generell als unkritisch betrachtet werden kann. Der relativ junge Parameter PSP konnte als aussagekräftiger glottaler Parameter bestätigt werden.

Für die Anwendung der SIM-Methode zur Analyse pathologischer Stimmen muß beachtet werden, daß an verschiedenen Punkten des Algorithmus die relative Konstanz der Grundperiode über 10 aufeinanderfolgende Perioden vorausgesetzt wird. Somit muß als Bedingung für die Anwendung der SIM-Methode bei pathologischen Stimmen der Jitterwert (d.h. die lokale Schwankung der Grundfrequenz) bzw. die Irregularitätskomponente nach Michaelis et al. (1998a) oder Fröhlich et al. (2000) bestimmt und diese Voraussetzung überprüft werden. Da die SIM-Methode jedoch die Differenzierung zwischen verschiedenen Phonationsarten erlaubt, die bei höhergradigen Stimmstörungen mit stärkeren Unregelmäßigkeiten in der Stimmgebung oft gar nicht mehr realisiert werden können, ist dies nicht als gravierende Einschränkung der potentiellen phoniatischen Einsetzbarkeit zu werten, die primär auf die Beschreibung leichtgradiger Stimmstörungen ausgerichtet wäre.

Daneben darf für eine routinemäßige Anwendung der SIM-Methode im klinischen Alltag der hohe Rechenbedarf nicht unberücksichtigt bleiben. Aufgrund der multidimensionalen Optimierungsverfahren und die iterativ erfolgende Inversfilterung beträgt bei einer Grundfrequenz von 110Hz die ungefähre Rechendauer für ein 200ms-Fenster auf einer SPARC-Station Ultra-1 für Optimierungsmethode A 1:15min, für MA 5:30min, für MA_NN 2:40min, für RD 50s und für MA_LF 3:10min. Ist Rechenzeit ein wichtiger Faktor, sollte daher auf Methode RD zurückgegriffen werden, die bei relativ hoher

Prädiktionsordnung gute Ergebnisse aufweist. Ansonsten führt Methode MA bei Prädiktionsordnung 10 zu glatteren Ergebnissen und bietet den Vorteil, nicht auf die vielfältigen Möglichkeiten an Parameterkonstellation zur Beschreibung des glottalen Modells verzichten zu müssen. Dieser Aspekt scheint insbesondere bei der Untersuchung pathologischer Stimmen wichtig, d.h. von Stimmen, die Abweichungen von einer normalen Stimmgebung aufweisen und daher von Parameter R_d eventuell nur unzureichend parametrisiert werden.

In dieser Hinsicht stellt sich ein interessanter Ausblick auf zukünftige Anwendungsmöglichkeiten dar, indem spezielle, an die Pathologie angepasste Parametrisierungen des Glottisflusses das LF-Modell in der SIM-Methode ersetzen könnten. Die Wahl des LF-Modells erfolgte aufgrund seiner weiten Verbreitung und seiner Möglichkeit, verschiedene Phonationsarten modellieren zu können (Gobl und Chasaide, 1992; Fant, 1993). Für die Beschreibung anderer, echter Stimmpathologien bietet es jedoch nicht notwendigerweise eine akkurate Grundlage. Daher könnten in weiterführenden Untersuchungen neue Modelle entwickelt werden, die spezielle pathologische Phonationsmechanismen angemessen beschreiben. Der Zusammenhang zwischen glottaler Fläche und glottalem Fluß ist zwar nur bei hoher Dämpfung näherungsweise durch eine lineare Beziehung beschreibbar (Titze, 1984) und weist im allgemeinen einen komplizierten Zusammenhang auf (Hertegård und Gauffin, 1995), doch ist die glottale Fläche durch Hochgeschwindigkeitsglottografie direkt zugänglich, während die direkte Messung des glottalen Flusses nur invasiv möglich ist. Aus diesem Grund könnte versucht werden, spezifische, stimpfpathologiebezogene glottale Parametrisierungen auf der Basis von hochgeschwindigkeitsglottografischen Aufnahmen zu entwerfen, die sich problemlos in die SIM-Methode einbinden ließen. Einen möglichen Ausgangspunkt könnte das Glottismodell von Titze (1989) darstellen, das sowohl die Parametrisierung der glottalen Öffnungsfläche als auch des glottalen Flusses durch denselben Satz an Steuerparametern erlaubt.

Die mit Hilfe der SIM-Methode berechneten Parameter können sinnvolle Ergänzungen zu bisher verwendeten und etablierten, niedrigdimensionalen Beschreibungen pathologischer Stimmigenschaften wie dem *Göttinger Heiserkeits-Diagramm* (Michaelis et al., 1998a; Fröhlich et al., 2000) darstellen. Somit könnten sie zur weiteren Differenzierung zwischen verschiedenen Stimmstörungsbildern beitragen, die angesichts der hohen Dimensionalität des Raums der perzeptiven Stimmgüte (Hammarberg et al., 1980) in einem äquivalenten, hochdimensionalen Raum akustischer Stimmparameter erfolgen sollte.

5. Zusammenfassung und Schlußfolgerungen

Das vorgestellte Verfahren zur simultanen Schätzung des LF-Glottismodells während der Inversfilterung (SIM-Verfahren) erlaubt die vollautomatische Berechnung glottaler Parameter, die phonationsspezifische Aussagen über die individuelle Stimmgebung ermöglichen. Ausgehend von der DAP-basierten Schätzung der Vokaltraktresonanzen erfolgt eine iterative Anpassung der Modellparameter, die zu einem Residuum mit optimal flachem Spektrum im Sinne des Itakura-Saito-Fehlers führt. Der Erfolg der multidimensionalen Parameteroptimierung zeigte sich abhängig von der gewählten Initialisierung, Optimierungsmethode und Prädiktionsordnung. Untersuchungen an synthetisierten Signalen und echten Stimmproben ergaben, daß bei künstlich generierten Signalen die korrekten zugrundeliegenden Parameterwerte generell mit hoher Zuverlässigkeit ermittelt werden konnten, sofern die Erzeugung der Stimmprobe auf einer durch das LF-Modell beschriebenen Signalform beruhte. Bei davon abweichenden Pulsformen, wie sie sich z.B. durch das selbstschwingende Zweimassenmodell der Stimmlippenbewegung in einer erweiterten Form zur Modellierung hypofunktioneller Phonation ergab, wurden dagegen nur wenige Parameter zuverlässig geschätzt, die jedoch weiterhin starke Aussagekraft bezüglich der simulierten Phonation besaßen. In der Analyse echter Stimmen ergaben sich deutliche, statistisch abgesicherte Zusammenhänge zwischen verschiedenen Phonationsarten und den berechneten Parametern, die somit die spezifische Charakterisierung individueller Stimmgebungsunterschiede ermöglichen. Die Untersuchung der Dimensionalität des zugrundeliegenden Parameterraums durch Hauptkomponentenanalyse ergab, daß eine vierdimensionale Beschreibung der Datenverteilung angemessen ist. Die Methode der sukzessiven Reduktion der Merkmale bei der auf dem vierdimensionalen Unterraum durchgeführten Diskriminanzanalyse zeigte, daß diese vier Dimensionen alle generell auch für die Diskriminierung notwendig sind, wobei sich individuelle Unterschiede bei einzelnen Sprechern zeigten. Das SIM-Verfahren liefert somit einen Beitrag zur unüberwachten, akustischen Charakterisierung von Stimmen, die durch andere Methoden bislang nicht geleistet wurde. Die sich ergebenden Parameter lassen sich direkt im Sinne der Stimmgebung interpretieren und ermöglichen so die zukünftige Anwendung der Methode in der phoniatischen Praxis.

Literaturverzeichnis

- Alipour, F. und Titze, I. (1996). Combined simulation of two-dimensional airflow and vocal fold vibration. In: Davis, P. und Fletcher, N., Hrsgb., *Vocal Fold Physiology: Controlling Complexity and Chaos*, S. 17–29. Singular Publishing Group, Inc., San Diego.
- Alku, P., Strik, H. und Vilkman, E. (1997). Parabolic spectral parameter - a new method for quantification of the glottal flow. *Speech Comm.* **22**, 67–79.
- Alku, P. und Vilkman, E. (1995). Effects of bandwidth on glottal airflow waveforms estimated by inverse filtering. *J. Acoust. Soc. Am.* **98**, 763–767.
- Alku, P. und Vilkman, E. (1996). A comparison of glottal voice source quantification parameters in breathy, normal and pressed phonation of female and male speakers. *Folia Phoniatr. Logop.* **48**, 240–254.
- Alku, P. (1992). Glottal wave analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering. *Speech Comm.* **11**, 109–118.
- Ananthapadmanabha, T. (1984). Acoustic analysis of voice source dynamics. *STL-QPSR* **2-3/1984**, 1–24.
- Arends, N., Povel, D.-J., van Os, E. und Speth, L. (1990). Predicting voice quality of deaf speakers on the basis of glottal characteristics. *J. Speech Hear. Res.* **33**, 116–122.
- Atal, B. S. und Schroeder, M. R. (1970). Adaptive predictive coding of speech signals. *Bell Systems Technical Journal* **49**, 1973–1986.
- Atal, B. S. und Schroeder, M. R. (1979). Predictive coding of speech signals and subjective error criteria. *IEEE Trans. Acoust., Speech, Sig. Process.* **ASSP-27**, 247–254.
- Baken, R. J. (1992). Electroglottography. *Journal of Voice* **6**, 98–110.
- Beasley, D., Bull, D. und Martin, R. (1993a). An overview of genetic algorithms: Part 1, fundamentals. *University Computing* **15**, 58–69.
- Beasley, D., Bull, D. und Martin, R. (1993b). An overview of genetic algorithms: Part 2, research topics. *University Computing* **15**, 170–181.

- Berry, D. A. und Titze, I. R. (1996). Normal modes in a continuum model of vocal fold tissues. *J. Acoust. Soc. Am.* **100**, 3345–3354.
- Childers, D. und Ahn, C. (1995). Modeling the glottal volume-velocity waveform for three voice types. *J. Acoust. Soc. Am.* **97**, 505–519.
- Childers, D. und Lee, C. (1991). Vocal quality factors: Analysis, synthesis, and perception. *J. Acoust. Soc. Am.* **90**, 2394–2410.
- Dueck, G. und Scheuer, T. (1990). Threshold accepting: A general purpose optimization algorithm appearing superior to simulated annealing. *Journal of Computational Physics* **90**, 161–175.
- Dueck, G. (1993). New optimization heuristics-the great deluge algorithm and the record-to-record travel. *Journal of Computational Physics* **104**, 86–92.
- El-Jaroudi, A. und Makhoul, J. (1991). Discrete all-pole modeling. *IEEE Trans. Sig. Process.* **39**, 411–423.
- Eskenazi, L., Childers, D. und Hicks, D. (1990). Acoustic correlates of vocal quality. *J. Speech Hear. Res.* **33**, 298–306.
- Fant, G., Liljencrants, J. und Lin, Q. (1986). A four-parameter model of glottal flow. *STL-QPSR* **4/1985**, 1–13.
- Fant, G. (1979). Vocal source analysis – a progress report. *STL-QPSR* **3-4/1979**, 31–54.
- Fant, G. (1993). Some problems in voice source analysis. *Speech Comm.* **13**, 7–22.
- Fant, G. (1995). The LF-model revisited. Transformations and frequency domain analysis. *STL-QPSR* **2-3/1995**, 119–156.
- Fettweis, A. und Meerkötter, K. (1975). On adaptors for wave digital filters. *IEEE Trans. Acoust., Speech, Sig. Process.* **ASSP-23**, 516–525.
- Fettweis, A. (1986). Wave digital filters: Theory and practice. *Proceedings of the IEEE* **74**, 270–327.
- Flanagan, J. und Landgraf, L. (1968). Self-oscillating source for vocal-tract synthesizers. *IEEE Trans. Audio Electroac.* **AU-16**, 57–64.
- Fröhlich, M., Michaelis, D., Strube, H. W. und Kruse, E. (1997). Acoustic voice quality description: Case studies for different regions of the hoarseness diagram. In: Wittenberg, T., Mergell, P., Tigges, M. und Eysholdt, U., Hrsgb., *Advances in Quantitative Laryngoscopy, 2nd ‘Round Table’*, S. 143–150. Dept. Phoniatics, Erlangen.

- Fröhlich, M., Michaelis, D. und Kruse, E. (1998a). Image sequences as necessary supplement to a pathological voice data base. In: de Krom, G., Hrsgb., *Proceedings of VOICEDATA98*, S. 64–69. Utrecht Institute of Linguistics OTS, Utrecht.
- Fröhlich, M., Michaelis, D. und Kruse, E. (1998b). Objektive Beschreibung der Stimmgüte unter Verwendung des Heiserkeits-Diagramms. *HNO* **46**, 684–689.
- Fröhlich, M., Michaelis, D. und Strube, H. W. (1998c). Acoustic “breathiness measures” in the description of pathologic voices. In: *Proceedings ICASSP 98*, Band 2, S. 937–940. Seattle, WA.
- Fröhlich, M., Michaelis, D., Strube, H. W. und Kruse, E. (1998d). Stimmgütebeschreibung mit Hilfe des Heiserkeits-Diagramms: Untersuchung verschiedener pathologischer Gruppen. In: Gross, M., Hrsgb., *Aktuelle phoniatisch-pädaudiologische Aspekte 1997/98*, Band 5, S. 42–48. Median Verlag, Heidelberg.
- Fröhlich, M., Michaelis, D., Strube, H. W. und Kruse, E. (1999). Perturbationsmaße und das asymmetrische Zweimassen-Glottismodell. In: *Aktuelle Phoniatisch-Pädaudiologische Aspekte 1999*. Im Druck.
- Fröhlich, M., Michaelis, D., Strube, H. W. und Kruse, E. (2000). Voice quality assessment by means of the hoarseness diagram. Angenommen zur Veröffentlichung bei *J. Speech Lang. Hear. Res.*, 2000.
- Fröhlich, M. und Strube, H. W. (1996). Inversfilterung an pathologischen Stimmen: Automatische Anpassung des LF-Modells. In: Portele, T. und Hess, W., Hrsgb., *Fortschritte der Akustik: DAGA 96*, S. 488–489. Oldenburg.
- Funaki, F., Miyanaga, Y. und Tochinal, K. (1996). A time-varying ARMAX speech analysis method based on glottal source model. In: *ASA and ASJ Third Joint Meeting*, S. 1267–1272.
- Gobl, C. und Chasaide, A. N. (1992). Acoustic characteristics of voice quality. *Speech Comm.* **11**, 481–490.
- Gobl, C. (1988). Voice source dynamics in connected speech. *STL-QPSR* **1/1988**, 123–159.
- Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J. und Wedin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol.* (Stockh.) **90**, 441–451.
- Hartung, und Elpelt, (1995). *Multivariate Statistik*. Oldenbourg Verlag, München, 5. Auflage.
- Hertegård, S. und Gauffin, J. (1995). Glottal area and vibratory patterns studied with simultaneous stroboscopy, flow glottography, and electroglottography. *J. Speech Hear. Res.* **38**, 85–100.

- Hess, W. und Indefrey, H. (1987). Accurate time-domain pitch determination of speech signals by means of a laryngograph. *Speech Communication* **6**, 55–68.
- Hess, W. (1983). *Pitch Determination of Speech Signals*. Springer Verlag, Berlin.
- Holmberg, E. B., Hillman, R. E., Perkell, J. S. und Gress, C. (1994). Relationships between intra-speaker variation in aerodynamic measures of voice production and variation in SPL across repeated recordings. *J. Speech Hear. Res.* **37**, 484–495.
- Holmberg, E., Hillman, R. und Perkell, J. (1995a). Measures of the glottal airflow waveform, EGG, and acoustic spectral slope for female voice. *Proceedings ICPhS 95* **3**, 178–181.
- Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C. und Goldman, S. L. (1995b). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures. *J. Speech Hear. Res.* **38**, 1212–1223.
- Ishizaka, K. und Flanagan, J. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Sys. Tech. J.* **51**, 1233–1268.
- Ishizaka, K. und Isshiki, N. (1976). Computer simulation of pathological vocal-cord vibration. *J. Acoust. Soc. Am.* **60**, 1193–1198.
- Karjalainen, M., Altosaar, T. und Vainio, M. (1998). Speech synthesis using warped linear prediction and neural networks. In: *Proceedings of ICASSP '98*, Band 2, S. 877–880. Seattle, WA.
- Karlsson, I. und Liljencrants, J. (1996). Diverse voice qualities: Models and data. *TMH-QPSR* **2/1996**, 143–146.
- Kay Elemetrics Corp., Pine Brook, NJ (1992). *Computerized Speech Lab Operations Manual*, April 1992.
- Kirkpatrick, S., Gelatt, Jr., C. und Vecchi, M. (1983). Optimization by simulated annealing. *Science* **220**, 671–680.
- Klasmeyer, G. (1998). The acoustic analysis of emotional speech. In: Lehmann, T., Palm, C., Spitzer, K. und Tolxdorff, T., Hrsgb., *Advances in Quantitative Laryngoscopy, Voice and Speech Research. Proceedings of the 3rd International Workshop*, S. 127–133. RWTH University of Technology, Aachen.
- Klatt, D. H. und Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* **87**, 820–857.
- Kobler, J. B., Hillman, R. E., Zeitels, S. M. und Kuo, J. (1998). Assessment of vocal function using simultaneous aerodynamic and calibrated videostroboscopic measures. *Ann. Otol. Rhinol. Laryngol.* **107**, 477–485.

- Kreiman, J. und Gerratt, B. R. (1994). The multidimensional nature of pathologic vocal quality. *J. Acoust. Soc. Am.* **96**, 1291–1302.
- Kreiman, J. und Gerratt, B. R. (1996). The perceptual structure of pathologic voice quality. *J. Acoust. Soc. Am.* **100**, 1787–1795.
- Kreiman, J. und Gerratt, B. R. (1998). Validity of rating scale measures of voice quality. *J. Acoust. Soc. Am.* **140**, 1598–1608.
- Kröger, B. (1991). Zur Auswirkung der Glottis-Sprechtrakt-Kopplung auf die Stimmreinheit. *Sprache-Stimme-Gehör* **15**, 139–142.
- Kruse, E., Michaelis, D., Zwirner, P. und Bender, E. (1997). Stimmfunktionelle Qualitätssicherung in der kurativen Mikrochirurgie der Larynxmalignome auf der Basis der „Laryngealen Doppelventilfunktion“. *HNO* **45**, 712–718.
- Kruse, E., Fröhlich, M. und Michaelis, D. (1998). Phonatory conditions and acoustic analysis of pathologic voices: Is there a correspondence? In: *Program and Abstract Book, 24th IALP congress*, S. 127. Amsterdam.
- Lehmann, T., Palm, C., Spitzer, K. und Tolxdorff, T., Hrsgb. (1998). *Advances in Quantitative Laryngoscopy, Voice and Speech Research. Proceedings of the 3rd International Workshop*, Aachen, Juni 19-20 1998. RWTH University of Technology.
- Lous, N., Hofmans, G., Veldhuis, R. und Hischberg, A. (1998). A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design. *ACUSTICA - acta acustica* **84**, 1135–1150.
- Ma, C., Kamp, Y. und Willems, L. F. (1994). A Frobenius norm approach to glottal closure detection from the speech signal. *IEEE Trans. Acoust., Speech, Sig. Process.* **2**, 258–265.
- Makhoul, J. (1975). Linear prediction: A tutorial review. *Proceedings of the IEEE* **63**, 561–580.
- Markel, J. und Gray, Jr., A. (1976). *Linear Prediction of Speech*. Band 12 von *Communication and Cybernetics*. Springer-Verlag, Berlin.
- Mathews, M., Miller, J. E. und David, Jr., E. (1961). Pitch synchronous analysis of voiced sounds. *J. Acoust. Soc. Am.* **33**, 179–186.
- Medan, Y., Yair, E. und Chazan, D. (1991). Super resolution pitch determination of speech signals. *IEEE Trans. Sig. Process.* **39**, 40–48.
- Mergell, P. und Herzel, H. (1997). Modelling biphonation – the role of the vocal tract. *Speech Comm.* **22**, 141–154.
- Meyer, P., Wilhelms, R. und Strube, H. (1989). A quasiarticulatory speech synthesizer for German language running in real time. *J. Acoust. Soc. Am.* **86**, 523–539.

- Meyer, P. (1987). *Sprachsynthese mit einem einfachen Artikulatoremodell und einem vokalktraktanalogen Echtzeitsynthetisator*. Dissertation, Georg-August-Universität Göttingen.
- Michaelis, D., Strube, H. W. und Kruse, E. (1997). Reliabilität und Validität des Heiserkeits-Diagramms. In: Gross, M. und Eysholdt, U., Hrsgb., *Aktuelle phoniatriisch-pädaudiologische Aspekte 1996*, Band 4, S. 25–26. Median Verlag, Heidelberg.
- Michaelis, D., Fröhlich, M. und Strube, H. W. (1998a). Selection and combination of acoustic features for the description of pathologic voices. *J. Acoust. Soc. Am.* **103**, 1628–1639.
- Michaelis, D., Fröhlich, M., Strube, H. W. und Kruse, E. (1998b). Reliabilität akustischer Stimmgütebeschreibung bei reduziertem Umfang der Stimmaufnahmen. In: Gross, M., Hrsgb., *Aktuelle phoniatriisch-pädaudiologische Aspekte 1997/98*, Band 5, S. 48–53. Median Verlag, Heidelberg.
- Michaelis, D., Fröhlich, M., Strube, H. W., Kruse, E., Story, B. und Titze, I. R. (1998c). Grenzen der Jitter- und Shimmer-Messung pathologischer Stimmen mit dem unüberwachten Waveform-Matching Verfahren. In: Sill, A., Hrsgb., *Fortschritte der Akustik – DAGA 98 (Zürich)*, S. 382–383. Oldenburg.
- Michaelis, D., Fröhlich, M., Strube, H. W., Kruse, E., Story, B. und Titze, I. R. (1998d). Some simulations concerning jitter and shimmer measurement. In: Lehmann, T., Palm, C., Spitzer, K. und Tolxdorff, T., Hrsgb., *Advances in Quantitative Laryngoscopy, Voice and Speech Research. Proceedings of the 3rd International Workshop*, S. 71–80. RWTH University of Technology, Aachen.
- Michaelis, D. und Strube, H. W. (1995). Empirical study to test the independence of different acoustic voice parameters on a large voice database. In: Pardo, J., Enriquez, E., Ortega, J., Ferreiros, J., Macias, J. und Valverde, F., Hrsgb., *Euro-speech '95*, Band 3, S. 1891–1894.
- Milenkovic, P. (1986). Glottal inverse filtering by joint estimation of an AR system with a linear input model. *IEEE Trans. Acoust., Speech, Sig. Process.* **ASSP-34**, 28–41.
- Milenkovic, P. (1987). Least mean square measures of voice perturbation. *J. Speech Hear. Res.* **30**, 529–538.
- Miller, R. (1959). Nature of the vocal cord wave. *J. Acoust. Soc. Am.* **31**, 667–677.
- Nelder, J. und Mead, R. (1965). Zitiert von Press et al. (1988).
- Parsa, V. und Jamieson, D. G. (1999). A comparison of high precision F0 extraction algorithms for sustained vowels. *J. Speech Lang. Hear. Res.* **42**, 112–126.

- Pelorsen, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J. und Auregan, Y. (1994). Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. *J. Acoust. Soc. Am.* **96**, 3416–3431.
- PEVOC-II, (1997). Pan european voice conference. Regensburg.
- Press, W., Flannery, B., Teukolsky, S. und Vetterling, W. (1988). *Numerical Recipes in C*. Cambridge University Press, Cambridge.
- Price, P. (1989). Male and female voice source characteristics: Inverse filtering results. *Speech Comm.* **8**, 261–277.
- Qi, Y. und Hillman, R. E. (1997). Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals. *J. Acoust. Soc. Am.* **102**, 537–543.
- Rosenberg, A. (1970). Effect of glottal pulse shape on the quality of natural vowels. *J. Acoust. Soc. Am.* **49**, 583–590.
- Rothenberg, M. (1973). A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *J. Acoust. Soc. Am.* **53**, 1632–1645.
- Schroeder, M. R. und Atal, B. S. (1985). Code-excited linear prediction (CELP): High quality speech at very low bit rates. *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 937–940.
- Schroeder, M. R. und Sloane, N. J. A. (1987). New permutation codes using Hadamard unscrambling. *IEEE Transactions on Information Theory* **IT-33**, 144–146.
- Schroeder, M. R. (1968). Period histogram and product spectrum: New methods for fundamental-frequency measurement. *J. Acoust. Soc. Am.* **43**, 829–834.
- Schroeder, M. R. (1981). Direct (nonrecursive) relations between cepstrum and predictor coefficients. *IEEE Trans. Acoust., Speech, Sig. Process.* **ASSP-29**, 297–301.
- Schroeder, M. R. (1982). Linear prediction, extremal entropy and prior information in speech signal analysis and synthesis. *Speech Comm.* **1**, 9–20.
- Schroeder, M. R. (1986). Predictive coding of speech: Historical review and directions for future research. *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 3157–3164.
- Schroeder, M. R. (1999). *Computer Speech. Recognition, Compression, Synthesis*. Band 35 von *Springer Series in Information Sciences*. Springer Verlag, Berlin.
- Schroeter, J. und Sondhi, M. (1992). Speech coding based on physiological models of speech production. In: Furui, S. und Sondhi, M. M., Hrsgb., *Advances in Speech Signal Processing*, S. 231–267. Marcel Dekker, Inc., New York.

- Skoglund, J. (1998). Analysis and quantization of glottal pulse shapes. *Speech Comm.* **24**, 133–152.
- Södersten, M., Håkansson, A. und Hammarberg, B. (1999). Comparison between automatic and manual inverse filtering procedures for healthy female voices. *Log. Phon. Vocol.* **24**, 26–38.
- Sondhi, M. und Schroeter, J. (1987). A hybrid time-frequency domain articulatory speech synthesizer. *IEEE Trans. Acoust., Speech, Sig. Process.* **ASSP-35**, 955–967.
- Story, B. H., Titze, I. R. und Hoffman, E. A. (1998). Vocal tract area functions for an adult female speaker based on volumetric imaging. *J. Acoust. Soc. Am.* **14**, 471–487.
- Story, B. und Titze, I. (1995). Voice simulation with a body-cover model of the vocal folds. *J. Acoust. Soc. Am.* **97**, 1249–1260.
- Strik, H., Cranen, B. und Boves, L. (1993). Fitting a LF-model to inverse filter signals. In: *Eurospeech '93*, Band 1, S. 103–106.
- Strik, H. (1994). *Physiological control and behaviour of the voice source in the production of prosody*. Dissertation, University Nijmegen.
- Strik, H. (1996). Testing two automatic methods for estimation of voice source parameters. In: Strik, H., Oostdijk, N., Cucchiari, C. und Coppen, P., Hrsgb., *Proceedings of the Department of Language and Speech*, Band 19, S. 105–127. Nijmegen.
- Strik, H. (1998). Automatic parametrization of differentiated glottal flow: Comparing methods by means of synthetic flow pulses. *J. Acoust. Soc. Am.* **103**, 2659–2669.
- Strube, H. W., Michaelis, D. und Fröhlich, M. (1996). Akustische Sprachparameter zur Bewertung glottaler Pathologien. In: Mehnert, D., Hrsgb., *Elektronische Sprachsignalverarbeitung*, Nr. 13 in Studentexte zur Sprachkommunikation, S. 52–58. Dresden.
- Strube, H. W. (1974). Determination of the instant of glottal closure from the speech-wave. *J. Acoust. Soc. Am.* **56**, 1625–1629.
- Strube, H. W. (1980). Linear prediction on a warped frequency scale. *J. Acoust. Soc. Am.* **68**, 1071–1076.
- Sulter, A. M. und Wit, H. P. (1996). Glottal volume velocity waveform characteristics in subjects with and without vocal training, related to gender, sound intensity, fundamental frequency, and age. *J. Acoust. Soc. Am.* **100**, 3360–3373.
- Svec, J. und Schutte, H. (1996). Videokymography: High-speed line scanning of vocal fold vibration. *Journal of Voice* **10**, 201–205.

- Titze, I. R., Baer, T., Cooper, D. und Scherer, R. (1983). Automated extraction of glottographic waveform parameters and regression to acoustic and physiologic variables. In: Bless, D. und Abbs, J., Hrsgb., *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, Kapitel 11, S. 146–154. College-Hill Press, San Diego.
- Titze, I. und Liang, H. (1992). Comparison of F_0 extraction methods for high precision voice perturbation measurements. NCVS Status and Progress Report **3**, 97–115.
- Titze, I. (1984). Parameterization of the glottal area, glottal flow, and vocal fold contact area. *J. Acoust. Soc. Am.* **75**, 570–580.
- Titze, I. R. (1989). A four-parameter model of the glottis and vocal fold contact area. *Speech Comm.* **8**, 191–201.
- Titze, I. R. (1990). Interpretation of the electroglottographic signal. *Journal of Voice* **4**, 1–9.
- Titze, I. R. (1994). *Principles of Voice Production*. Prentice Hall, Englewood Cliffs, N.J. 07632.
- Varho, S. und Alku, P. (1998). Separated Linear Prediction – a new all-pole modeling technique for speech analysis. *Speech Comm.* **24**, 111–121.
- Veldhuis, R. N. (1998). The spectral relevance of glottal-pulse parameters. In: *Proceedings of ICASSP '98*, Band 2, S. 873–876. Seattle, WA.
- Wakita, H. und Fant, G. (1978). Toward a better vocal tract model. *STL-QPSR* **1/1978**, 9–29.
- Wittenberg, T., Mergell, P., Tigges, M. und Eysholdt, U., Hrsgb. (1997). *Advances in Quantitative Laryngoscopy, 2nd 'Round Table'*, Erlangen, 1997. Dept. Phoniatics.

A. Erweiterung des Zweimassenmodells auf eine nicht-rechteckige Glottisfläche

Im folgenden soll die Erweiterung des Zweimassenmodells auf ein trapezförmige bzw. dreieckige Glottisfläche beschrieben werden. Zunächst werden in Abschnitt A.1 die Gleichungen des originalen Modells referiert, woraufhin in Abschnitt A.2 die Modifikationen und Ergänzungen beim Übergang zur nicht-rechteckigen Glottisfläche ausgeführt werden.

A.1 Ursprüngliches Zweimassenmodell

Das verwendete Modell der Stimmlippen als Schwingungssystem basiert auf dem in der Literatur als „Zweimassenmodell“ (*two mass model*) bekannten Modell (Ishizaka und Flanagan, 1972), das durch Sondhi und Schroeter (1987) und Schroeter und Sondhi (1992) noch einige leichte Verbesserungen in der numerischen Beschreibung erfuhr. Die Schwingungsgleichungen für die beiden Federpendel lauten (die Indizes beziehen sich auf die jeweiligen Pendelkörper):

$$\begin{aligned} m_1 \ddot{x}_1 + r_1 \dot{x}_1 + s_1 + k_c(x_1 - x_2) &= f_1 \\ m_2 \ddot{x}_2 + r_2 \dot{x}_2 + s_2 + k_c(x_2 - x_1) &= f_2. \end{aligned} \tag{A.1}$$

Hierbei lauten die Federkräfte s für $i = 1, 2$:

$$s_i = \begin{cases} k_i x_i + s'_i, & x_i > -d_0, \\ (k_i + h_i) x_i + s'_i, & x_i \leq -d_0, \end{cases} \tag{A.2}$$

wobei d_0 die Ruheauslenkung der Pendel darstellt und s'_i nichtlineare (kubische) Terme (s.u.). Somit beschreiben (A.1) zwei nichtlineare, gekoppelte Differentialgleichungen.

Die anregenden Kräfte auf der rechten Seite in (A.1) sind durch folgende Gleichun-

gen gegeben:

$$f_1 = \begin{cases} P_{m1} l_g d_1, & x_1 > d_0, x_2 > d_0 \\ P_s l_g d_1 & \text{sonst} \end{cases} \quad (\text{A.3})$$

$$f_2 = \begin{cases} P_{m2} l_g d_2, & x_1 > d_0, x_2 > d_0 \\ P_s l_g d_2, & x_1 > d_0, x_2 \leq d_0 \\ 0 & \text{sonst} \end{cases} \quad (\text{A.4})$$

mit

$$P_{m1} = P_s - 1,37 \frac{\rho}{2} \left(\frac{U_g}{A_{g1}} \right)^2 - \frac{1}{2} \left(R_{v1} U_g + L_{g1} \frac{dU_g}{dt} \right) \quad (\text{A.5})$$

$$P_{m2} = P_{m1} - \frac{1}{2} \left\{ (R_{v1} + R_{v2}) U_g + (L_{g1} + L_{g2}) \frac{dU_g}{dt} \right\} - \frac{\rho}{2} U_g^2 \left(\frac{1}{A_{g2}^2} - \frac{1}{A_{g1}^2} \right). \quad (\text{A.6})$$

ρ bezeichnet hierbei die Dichte der Luft, U_g den glottalen Fluß, P_s den subglottalen Druck, L_g und R_v die Induktivitäten bzw. Widerstände, die sich aus den glottalen Öffnungsflächen A_g für die Massen 1 und 2 ergeben.¹

Indem die Terme höherer Ordnung durch die Werte des vorangegangenen Iterationsschritts ersetzt werden, lassen sich Gleichungen (A.1) in ein lineares Gleichungssystem überführen:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_{(n);1} \\ x_{(n);2} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}. \quad (\text{A.7})$$

Hierbei bezeichnen $x_{(n);1}$ und $x_{(n);2}$ die Auslenkungen der 1. bzw. 2. Masse zum Zeitpunkt (Abtastwert) n . Die Matrixkoeffizienten lauten

$$a_{ii} = (k_i + h_i + k_c) t_s^2 + r_i t_s + m_i, \quad i = 1, 2 \quad (\text{A.8})$$

$$a_{12} = a_{21} = -k_c t_s^2. \quad (\text{A.9})$$

Die rechte Seite von (A.7) ist für $i = 1, 2$ durch folgende Ausdrücke gegeben:

$$b_i = (2m_i + r_i t_s) x_{(n-1);i} - m_i x_{(n-2);i} - s'_i t_s^2 + f_{(n-1);i} t_s^2 \quad (\text{A.10})$$

$$s'_i = \begin{cases} k_i \eta_{ki} x_{(n-1);i}^3, & x_{(n-1);i} > d_0 \\ k_i \eta_{ki} x_{(n-1);i}^3 + h_i \left[d_0 + \eta_{hi} (x_{(n-1);i} + d_0)^3 \right], & x_{(n-1);i} \leq d_0. \end{cases} \quad (\text{A.11})$$

Die Werte der Konstanten sind in Tabelle 2.1 angegeben.

¹Zur Definition von L_g und R_v siehe Ishizaka und Flanagan (1972) Gl. (8), Sondhi und Schroeter (1987) Gln. (3)-(4) oder Schroeter und Sondhi (1992) Gln. (6)-(7) unter Berücksichtigung von Schroeter und Sondhi (1992) Gl. (9).

A.2 Erweitertes Zweimassenmodell

Durch Einführung des Anstellwinkels θ (vgl. Abb. 2.6) wird die Beschreibung der Bewegungsgleichungen komplizierter, da sich nun die anregenden Kräfte f_1, f_2 , die Federkräfte s_1, s_2 , sowie die Dämpfungskonstanten r_1, r_2 nicht mehr abrupt bei Verschluss verändern, sondern sich eine kontinuierliche Abhängigkeit vom aktuellen Öffnungsgrad ergibt. Diese Größen werden daher als Integrale der zugrundeliegenden Liniendichten über die Länge der Stimmlippen in y -Richtung berechnet (Abb. A.1). Die Gleichungen sind für die „obere“ und „untere“ Masse dieselben, so daß im folgenden auf die Indizes 1 und 2 verzichtet wird.

Die Federsteife wird ausgedehnt über die gesamte Länge des Pendels betrachtet:

$$k = \int_{-\frac{l_g}{2} \cos \theta}^{\frac{l_g}{2} \cos \theta} \frac{k}{l_g \cos \theta} dy. \quad (\text{A.12})$$

Für die offene Glottis wird in (A.2) und (A.11) x durch y parametrisiert, wobei die momentane Auslenkung ξ des Punktes $y = 0$ sowie der Winkel θ eingehen (vgl. Abb. A.1):

$$x(y) = \xi - y \tan \theta. \quad (\text{A.13})$$

Aus den sich ergebenden Ausdrücken werden die Federkräfte unter Verwendung von

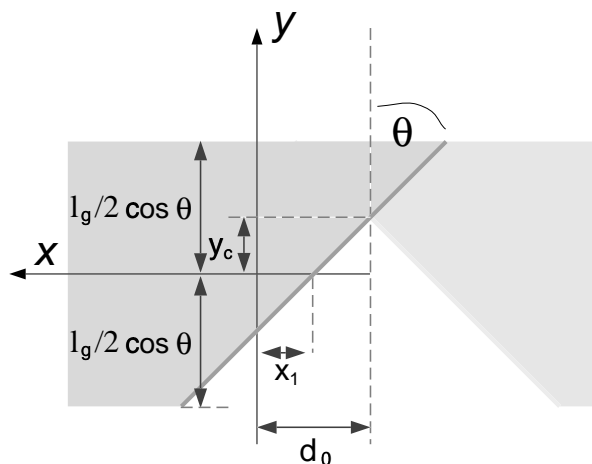


Abbildung A.1: Integration in y -Richtung bei teilweise geschlossener Glottis.

(A.12) wie folgt berechnet:

$$\begin{aligned}
s(x) &= \int_{y_{\min,o}}^{y_{\max,o}} k(x + \eta_k x^3) \frac{dy}{\cos \theta} + \int_{y_{\min,c}}^{y_{\max,c}} h(x + \eta_h x^3) \frac{dy}{\cos \theta} \\
&= \frac{1}{l_g \cos \theta} \left(\int_{y_{\min,o}}^{y_{\max,o}} k(\xi - y \tan \theta + \eta_k (\xi - y \tan \theta)^3) dy \right. \\
&\quad \left. + \int_{y_{\min,c}}^{y_{\max,c}} h(\xi + d_0 - y \tan \theta + \eta_h (\xi + d_0 - y \tan \theta)^3) dy \right) \quad (A.14)
\end{aligned}$$

$$= \left[s_o \right]_{y_{\min,o}}^{y_{\max,o}} + \left[s_c \right]_{y_{\min,c}}^{y_{\max,c}} \quad (A.15)$$

mit

$$\begin{aligned}
s_o &= \frac{k}{l_g \cos \theta} \left(\xi y - \frac{1}{2} y^2 \tan \theta + \eta_k \left(\xi^3 y - \frac{3}{2} \xi^2 y^2 \tan \theta + \xi y^3 \tan^2 \theta - \frac{1}{4} y^4 \tan^3 \theta \right) \right) \\
s_c &= \frac{h}{l_g \cos \theta} \left((\xi + d_0) y - \frac{1}{2} y^2 \tan \theta + \eta_h \left((\xi + d_0)^3 y \right. \right. \\
&\quad \left. \left. - \frac{3}{2} (\xi + d_0)^2 y^2 \tan \theta + (\xi + d_0) y^3 \tan^2 \theta - \frac{1}{4} y^4 \tan^3 \theta \right) \right). \quad (A.16)
\end{aligned}$$

Die Grenzen für die Integration in (A.15) hängen vom momentanen Grad des Verschlusses ab (Index „o“ bezeichnet den offenen Teil der Glottis, Index „c“ den geschlossenen). Für s_o wird zu jedem Zeitpunkt über die volle Länge der Stimmlippe integriert, wodurch sich die geradzahigen Potenzen in y wegheben:

$$\begin{aligned}
s_o &= \left[k \left(\xi y - \frac{1}{2} y^2 \tan \theta + \eta_k \left(\xi^3 y - \frac{3}{2} \xi^2 y^2 \tan \theta + \xi y^3 \tan^2 \theta - \frac{1}{4} y^4 \tan^3 \theta \right) \right) \right]_{-\frac{l_g}{2} \cos \theta}^{\frac{l_g}{2} \cos \theta} \\
&= l_g \cos \theta \cdot k \left(\xi + \eta_k \xi^3 + \eta_k \xi \left(\frac{l_g}{2} \right)^2 \sin^2 \theta \right) \quad (A.17)
\end{aligned}$$

Für den geschlossenen Anteil ist $y_{\max,c} \equiv \frac{l_g}{2} \cos \theta$ unabhängig von ξ , doch hängt $y_{\min,c}$ von der momentanen Auslenkung ξ und dem Öffnungswinkel θ ab (vgl. Abb. A.1). Die untere Integrationsgrenze für s_c lautet damit

$$y_{\min,c} = \begin{cases} (\xi + d_0) \frac{1}{\tan \theta}, & \xi + d_0 < \frac{l_g}{2} \sin \theta \\ \frac{l_g}{2} \cos \theta, & \xi + d_0 \geq \frac{l_g}{2} \sin \theta. \end{cases} \quad (A.18)$$

Wie im ursprünglichen Modell werden die Terme höherer Ordnung in ξ aus den zeitverzögerten Werten $\xi_{(n-1);1}$ berechnet. Die Koeffizienten des Gleichungssystems (A.7)

lauten daher für $i = 1, 2$:

$$a_{ii} = \left(k_i + \eta_{ki} \left(\frac{l_g \sin \theta}{2} \right)^2 + \frac{h_i}{2} - \frac{h_i y_{0i}}{l_g \cos \theta} + k_c \right) t_s^2 + r_i t_s + m_i \quad (\text{A.19})$$

$$a_{12} = a_{21} = -k_c t_s^2 \quad (\text{A.20})$$

$$\begin{aligned} b_i = & (2m_i + r_i t_s) \xi_{(n-1);i} - m_i \xi_{(n-2);i} + f_{(n-1);i} t_s^2 \\ & - t_s^2 \left[k_i \eta_{ki} \xi_{(n-1);i}^3 + \frac{h_i}{2} \left(d_0 - \frac{1}{4} l_g \sin \theta \right) - \frac{h_i}{l_g \cos \theta} \left(d_0 y_0 - \frac{1}{2} y_0^2 \tan \theta \right) \right. \\ & + h_i \eta_{hi} \left(\frac{1}{2} (\xi_{(n-1);i} + d_0)^3 - \frac{3}{8} (\xi_{(n-1);i} + d_0)^2 l_g \sin \theta \right. \\ & \quad \left. \left. + \frac{1}{8} (\xi_{(n-1);i} + d_0) (l_g \sin \theta)^2 - \frac{1}{64} (l_g \sin \theta)^3 \right) \right. \\ & - \frac{h_i \eta_{hi}}{l_g \cos \theta} \left(\frac{1}{2} (\xi_{(n-1);i} + d_0)^3 y_0 - \frac{3}{2} (\xi_{(n-1);i} + d_0)^2 y_0^2 \tan \theta \right. \\ & \quad \left. \left. + (\xi_{(n-1);i} + d_0) y_0^3 \tan^2 \theta - \frac{1}{4} y_0^4 \tan^3 \theta \right) \right]. \quad (\text{A.21}) \end{aligned}$$

B. Der DAP-Algorithmus

Ausgehend von der Darstellung in Abschnitt 2.9.1 werden im folgenden die mathematischen Grundlagen des DAP-Algorithmus genauer ausgeführt.

Um die Filterkoeffizienten a_i zu berechnen, wird zunächst die Nulpole-Bedingung (2.22) in Gl. (2.34) eingesetzt:

$$\begin{aligned} \epsilon_{\text{IS}} &= \frac{1}{N} \sum_{m=1}^N \left[P(\omega_m) \left| \sum_{k=0}^p a_k e^{-i\omega_m k} \right|^2 - \ln \left\{ P(\omega_m) \left| \sum_{k=0}^p a_k e^{-i\omega_m k} \right|^2 \right\} - 1 \right] \\ &= \frac{1}{N} \sum_{m=1}^N \left[P(\omega_m) \sum_{k=0}^p \sum_{l=0}^p a_k a_l e^{i\omega_m(k-l)} \right. \\ &\quad \left. - \ln \left\{ P(\omega_m) \sum_{k=0}^p \sum_{l=0}^p a_k a_l e^{i\omega_m(k-l)} \right\} - 1 \right]. \end{aligned} \quad (\text{B.1})$$

Die Ableitung der Doppelsumme nach den einzelnen Koeffizienten lautet:

$$\frac{\partial}{\partial a_i} P(\omega_m) \sum_{k=0}^p \sum_{l=0}^p a_k a_l e^{i\omega_m(k-l)} = P(\omega_m) \cdot 2 \sum_{k=0}^p a_k e^{i\omega_m(k-i)}. \quad (\text{B.2})$$

Damit ergibt sich die Minimierungsbedingung (2.21) zu:

$$\begin{aligned} \frac{\partial \epsilon_{\text{IS}}}{\partial a_i} &= \frac{1}{N} \sum_{m=1}^N P(\omega_m) \cdot 2 \sum_{k=0}^p a_k e^{i\omega_m(k-i)} - \frac{1}{N} \sum_{m=1}^N \frac{P(\omega_m) \cdot 2 \sum_{k=0}^p a_k e^{i\omega_m(k-i)}}{P(\omega_m) \left| \sum_{l=0}^p a_l e^{i\omega_m l} \right|^2} \\ &\stackrel{(2.27),(2.32)}{=} 2 \sum_{k=0}^p a_k R(k-i) - 2 \sum_{k=0}^p a_k \hat{R}(k-i) \\ &= 2 \sum_{k=0}^p a_k \left[R(k-i) - \hat{R}(k-i) \right] = 0, \quad 0 \leq i \leq p. \end{aligned} \quad (\text{B.3})$$

Da $\hat{R}(i)$ nach Gl. (2.32) ebenfalls eine Funktion der Filterkoeffizienten a_i ist, stellen (B.3) ein nichtlineares Gleichungssystem in den $p+1$ Unbekannten a_i dar.

Um dieses Gleichungssystem zu lösen, betrachten wir zunächst die Umformung der Identität:

$$\begin{aligned}
& \hat{H}(\omega_m) \cdot A(\omega_m) = 1 \\
\stackrel{\cdot \hat{H}^*}{\Leftrightarrow (2.22)} & \hat{P}(\omega_m) \cdot A(\omega_m) = \hat{H}^*(\omega_m) \\
\stackrel{\cdot e^{i\omega_m i}}{\Leftrightarrow (2.22)} & \sum_{k=0}^p a_k \hat{P}(\omega_m) e^{i\omega_m(i-k)} = \hat{H}^*(\omega_m) e^{i\omega_m i} \\
\stackrel{\frac{1}{N} \sum_m}{\Leftrightarrow (2.31)} & \sum_{k=0}^p a_k \hat{R}(i-k) = \frac{1}{N} \sum_{m=1}^N \hat{H}^*(\omega_m) e^{i\omega_m i} \\
\stackrel{(\cdot)^*}{\Leftrightarrow} & \sum_{k=0}^p a_k \hat{R}(i-k) = \frac{1}{N} \sum_{m=1}^N \hat{H}(\omega_m) e^{-i\omega_m i} \\
\stackrel{\Leftrightarrow}{\Leftrightarrow} & \sum_{k=0}^p a_k \hat{R}(i-k) = \hat{h}(-i), \quad \forall i. \quad (\text{B.4})
\end{aligned}$$

Hierbei ist $\hat{h}(-i)$ die zeitinvertierte Impulsantwort des diskret abgetasteten Nulpole-Modells:

$$\hat{h}(-i) = \frac{1}{N} \sum_{m=1}^N \frac{e^{-i\omega_m i}}{A(\omega_m)}. \quad (\text{B.5})$$

Durch Einsetzen von (B.4) und (B.5) in die Minimierungsbedingung (B.3) erhält man folgendes Gleichungssystem:

$$\begin{aligned}
& \sum_{k=0}^p a_k R(k-i) = \sum_{k=0}^p a_k \hat{R}(k-i), \quad 0 \leq i \leq p \\
\stackrel{\Leftrightarrow}{\Leftrightarrow} & \sum_{k=0}^p a_k R(i-k) = \hat{h}(-i), \quad 0 \leq i \leq p. \quad (\text{B.6})
\end{aligned}$$

In Matrixschreibweise läßt sich dies schreiben als

$$\mathbf{R}\mathbf{a} = \hat{\mathbf{h}}, \quad (\text{B.7})$$

wobei $\mathbf{a} = (a_0, a_1, \dots, a_p)^T$ und $\hat{\mathbf{h}} = (h(0), h(-1), \dots, h(-p))^T$ p -dimensionale Spaltenvektoren und \mathbf{R} eine symmetrische Toeplitz-Matrix mit den Elementen der ersten $p+1$ AKF-Terme darstellen. Da $\hat{h}(-i)$ nach Gl. (B.5) und (2.22) ebenfalls von den Filterkoeffizienten a_i abhängt, stellt (B.7) immer noch ein nichtlineares Gleichungssystem in den Koeffizienten a_i dar, das iterativ gelöst wird. Um die Konvergenz des Algorithmus zu verbessern, werden die neuen Koeffizienten nicht direkt aus (B.7) berechnet, sondern durch Einführung eines Konvergenzparameters α gemäß

$$\mathbf{a}_{m+1} = \mathbf{a}_m(1 - \alpha) + \alpha \mathbf{R}^{-1} \hat{\mathbf{h}}_m. \quad (\text{B.8})$$

In den Untersuchungen von El-Jaroudi und Makhoul (1991) konnte die Konvergenzgeschwindigkeit durch geeignete Wahl von α im Bereich $0.4 \leq \alpha \leq 0.8$ deutlich erhöht werden. In dieser Arbeit wurde daher $\alpha = 0.5$ verwendet.

C. Analyseergebnisse für verschiedene Signalarten

Die folgenden Seiten enthalten Tabellen, in denen Ergebnisse aus Kapitel 3 ausführlich angegeben werden.

Tabelle C.1: Rangkorrelation der Medianwerte aus dem SIM-Verfahren zu θ und zur Referenz für verschiedene Optimierungsmethoden (Prädiktionsordnung 10, Signaltyp 2MMS). Die Markierung der Signifikanzniveaus \mathbf{p} erfolgt gemäß Tabelle 2.4.

	Korrelation zu θ				Korrelation zur Referenz			
	A	MA	MA_NN	MA_LF	A	MA	MA_NN	MA_LF
t_p	0.05 [†]	0.19 [†]	-0.05 [†]	-0.17 [†]	0.03 [†]	0.03 [†]	-0.08 [†]	-0.08 [†]
t_e	0.42 [*]	0.26 [†]	0.10 [†]	-0.03 [†]	0.36 [*]	-0.02 [†]	-0.02 [†]	-0.05 [†]
t_a	-0.34 [†]	-0.30 [†]	-0.04 [†]	-0.29 [†]	-0.34 [†]	-0.29 [†]	-0.10 [†]	-0.28 [†]
E_e	-0.79	-0.76	-0.74	-0.79	0.95	0.94	0.89	0.96
$t_e - t_p$	0.65	0.23 [†]	0.29 [†]	0.06 [†]	0.16 [†]	-0.01 [†]	-0.43 [*]	0.22 [†]
OQ_1	0.65	0.18 [†]	0.16 [†]	-0.16 [†]	0.62	-0.14 [†]	-0.06 [†]	-0.05 [†]
OQ_2	0.67	0.11 [†]	0.20 [†]	-0.12 [†]	0.69	-0.14 [†]	0.06 [†]	-0.17 [†]
SQ_1	-0.57	-0.11 [†]	-0.51 [†]	-0.08 [†]	-0.58	-0.21 [†]	-0.49 [†]	0.05 [†]
SQ_2	-0.32 [†]	0.16 [†]	-0.44 [*]	0.33 [†]	-0.32 [†]	0.14 [†]	-0.47 [†]	0.34 [†]
CQ_1	0.65	0.23 [†]	0.29 [†]	0.06 [†]	0.16 [†]	-0.01 [†]	-0.43 [*]	0.22 [†]
CQ_2	0.61	-0.11 [†]	0.34 [†]	-0.08 [†]	0.31 [†]	-0.07 [†]	-0.31 [†]	0.09 [†]
PSP	0.62	-0.03 [†]	0.32 [†]	-0.07 [†]	0.60	-0.27 [†]	-0.10 [†]	-0.09 [†]

Tabelle C.2: Rangkorrelation der Medianwerte aus dem SIM-Verfahren zu θ und zur Referenz für $\theta > 0.85^\circ$ für verschiedene Optimierungsmethoden (Prädiktionsordnung 10, Signaltyp 2MMS) Die Markierung der Signifikanzniveaus \mathbf{p} erfolgt gemäß Tabelle 2.4.

	Korrelation zu θ				Korrelation zur Referenz			
	A	MA	MA_NN	MA_LF	A	MA	MA_NN	MA_LF
t_p	0.34 [†]	0.39 [†]	0.59 [*]	-0.00 [†]	0.25 [†]	0.41 [†]	0.64 ⁺	-0.08 [†]
t_e	0.66 ⁺	0.36 [†]	0.58 [*]	-0.13 [†]	0.51 [†]	-0.00 [†]	0.32 [†]	-0.27 [†]
t_a	-0.31 [†]	0.11 [†]	-0.13 [†]	-0.20 [†]	-0.26 [†]	0.14 [†]	-0.25 [†]	-0.06 [†]
E_e	-0.57 [*]	-0.45 [†]	-0.49 [†]	-0.60 [*]	0.95	0.84	0.87	0.98
$t_e - t_p$	0.80	0.34 [†]	0.78	-0.08 [†]	-0.34 [†]	-0.05 [†]	-0.60 [*]	0.53 [*]
OQ_1	0.77	0.41 [†]	0.63 [*]	-0.09 [†]	0.31 [†]	-0.25 [†]	0.12 [†]	0.36 [†]
OQ_2	0.69 ⁺	0.30 [†]	0.61 [*]	-0.06 [†]	0.57 [*]	0.03 [†]	0.41 [†]	0.20 [†]
SQ_1	-0.54 [*]	-0.02 [†]	-0.41 [†]	0.20 [†]	-0.54 [*]	0.04 [†]	-0.45 [†]	0.30 [†]
SQ_2	-0.13 [†]	0.53 [*]	0.27 [†]	0.63 [*]	-0.13 [†]	0.58 [*]	0.27 [†]	0.74 ⁺
CQ_1	0.80	0.34 [†]	0.78	-0.08 [†]	-0.34 [†]	-0.05 [†]	-0.60 [*]	0.53 [*]
CQ_2	0.60 [*]	0.03 [†]	0.67 ⁺	-0.13 [†]	-0.05 [†]	0.10 [†]	-0.28 [†]	0.59 [*]
PSP	0.53 [*]	0.06 [†]	0.65 ⁺	-0.13 [†]	0.06 [†]	-0.28 [†]	-0.03 [†]	0.43 [†]

Tabelle C.3: Rangkorrelation der Medianwerte aus dem SIM-Verfahren zu θ und zur Referenz für verschiedene Optimierungsmethoden (Prädiktionsordnung 13, Signaltyp 2MMF). Die Markierung der Signifikanzniveaus \mathbf{p} erfolgt gemäß Tabelle 2.4.

	Korrelation zu θ				Korrelation zur Referenz			
	A	MA	MA_NN	MA_LF	A	MA	MA_NN	MA_LF
t_p	-0.63	-0.71	-0.65	-0.14 [†]	-0.39*	-0.41*	-0.46 ⁺	-0.08 [†]
t_e	-0.59	-0.70	-0.64	-0.31 [†]	-0.36*	-0.44*	-0.46 ⁺	-0.24 [†]
t_a	0.66	0.53 ⁺	0.52 ⁺	-0.33 [†]	0.63	0.10 [†]	0.32 [†]	-0.19 [†]
E_e	—	—	—	—	—	—	-0.59	—
$t_e - t_p$	0.14 [†]	-0.67	-0.64	-0.60	-0.13 [†]	0.15 [†]	0.24 [†]	0.12 [†]
OQ_1	0.03 [†]	-0.67	-0.63	-0.60	0.17 [†]	-0.44*	-0.44*	-0.33 [†]
OQ_2	0.11 [†]	-0.63	-0.65	-0.56	0.23 [†]	-0.39*	-0.44*	-0.33 [†]
SQ_1	-0.06 [†]	0.81	0.83	0.76	-0.25 [†]	0.72	0.67	0.59
SQ_2	-0.26 [†]	0.45*	0.37*	0.63	-0.21 [†]	0.43*	0.41*	0.60
CQ_1	0.14 [†]	-0.67	-0.64	-0.60	-0.13 [†]	0.15 [†]	0.24 [†]	0.12 [†]
CQ_2	0.29 [†]	-0.57	-0.66	-0.58	0.05 [†]	-0.15 [†]	-0.10 [†]	-0.11 [†]
PSP	0.22 [†]	-0.59	-0.63	-0.60	0.24 [†]	-0.42*	-0.38*	-0.35*

Tabelle C.4: Rangkorrelation der Medianwerte aus dem SIM-Verfahren zu θ und zur Referenz für $\theta > 0.85^\circ$ für verschiedene Optimierungsmethoden (Prädiktionsordnung 13, Signaltyp 2MMF). Die Markierung der Signifikanzniveaus \mathbf{p} erfolgt gemäß Tabelle 2.4.

	Korrelation zu θ				Korrelation zur Referenz			
	A	MA	MA_NN	MA_LF	A	MA	MA_NN	MA_LF
t_p	0.19 [†]	-0.44 [†]	-0.38 [†]	0.55*	0.26 [†]	-0.41 [†]	-0.35 [†]	0.56*
t_e	0.25 [†]	-0.43 [†]	-0.31 [†]	0.55*	0.47 [†]	-0.52*	-0.53*	0.26 [†]
t_a	0.71 ⁺	0.17 [†]	-0.07 [†]	0.24 [†]	0.65 ⁺	0.17 [†]	-0.10 [†]	0.19 [†]
E_e	—	—	—	—	—	—	—	—
$t_e - t_p$	0.62*	-0.04 [†]	-0.27 [†]	0.36 [†]	-0.37 [†]	0.23 [†]	0.38 [†]	0.09 [†]
OQ_1	0.60*	-0.09 [†]	-0.30 [†]	0.47 [†]	0.53*	-0.34 [†]	-0.43 [†]	0.51 [†]
OQ_2	0.54*	-0.01 [†]	-0.28 [†]	0.50 [†]	0.52*	-0.16 [†]	-0.40 [†]	0.58*
SQ_1	0.12 [†]	0.38 [†]	0.53*	0.00 [†]	0.08 [†]	0.35 [†]	0.43 [†]	0.06 [†]
SQ_2	0.04 [†]	0.28 [†]	0.25 [†]	-0.09 [†]	0.00 [†]	0.19 [†]	0.16 [†]	-0.03 [†]
CQ_1	0.62*	-0.04 [†]	-0.27 [†]	0.36 [†]	-0.37 [†]	0.23 [†]	0.38 [†]	0.09 [†]
CQ_2	0.65 ⁺	0.10 [†]	-0.19 [†]	0.42 [†]	-0.33 [†]	-0.00 [†]	0.19 [†]	0.25 [†]
PSP	0.59*	0.09 [†]	-0.21 [†]	0.43 [†]	-0.05 [†]	0.01 [†]	0.05 [†]	0.27 [†]

Tabelle C.5: Rangkorrelation der Medianwerte aus dem SIM-Verfahren zu θ und zur Referenz für verschiedene Optimierungsmethoden (Prädiktionsordnung 10, Signaltyp 2MMF). Die Markierung der Signifikanzniveaus \mathbf{p} erfolgt gemäß Tabelle 2.4.

	Korrelation zu θ				Korrelation zur Referenz			
	A	MA	MA_NN	MA_LF	A	MA	MA_NN	MA_LF
t_p	-0.61	-0.58	-0.42*	0.36*	-0.57	-0.54 [†]	-0.25 [†]	0.17 [†]
t_e	-0.64	-0.61	-0.48 [†]	0.37*	-0.51 [†]	-0.55 [†]	-0.32 [†]	0.18 [†]
t_a	0.74	0.91	0.45*	0.79	0.35*	0.59	0.11 [†]	0.69
E_e	—	—	—	—	—	—	—	—
$t_e - t_p$	-0.53 [†]	-0.06 [†]	-0.43*	0.50 [†]	0.06 [†]	0.17 [†]	0.13 [†]	0.08 [†]
OQ_1	-0.60	-0.27 [†]	-0.41*	0.50 [†]	-0.39*	-0.14 [†]	-0.18 [†]	0.37*
OQ_2	-0.52 [†]	-0.18 [†]	-0.41*	0.53 [†]	-0.29 [†]	-0.02 [†]	-0.19 [†]	0.44*
SQ_1	0.48 [†]	-0.02 [†]	0.51 [†]	-0.23 [†]	0.51 [†]	-0.12 [†]	0.43*	-0.26 [†]
SQ_2	0.11 [†]	-0.46 [†]	-0.01 [†]	-0.50 [†]	0.13 [†]	-0.47 [†]	0.00 [†]	-0.51 [†]
CQ_1	-0.53 [†]	-0.06 [†]	-0.43*	0.50 [†]	0.06 [†]	0.17 [†]	0.13 [†]	0.08 [†]
CQ_2	-0.37*	0.13 [†]	-0.33 [†]	0.50 [†]	-0.00 [†]	0.21 [†]	-0.07 [†]	0.31 [†]
PSP	-0.37*	0.13 [†]	-0.33 [†]	0.50 [†]	-0.15 [†]	0.21 [†]	-0.16 [†]	0.42*

Tabelle C.6: Rangkorrelation der Medianwerte aus dem SIM-Verfahren zu θ und zur Referenz für $\theta > 0.85^\circ$ für verschiedene Optimierungsmethoden (Prädiktionsordnung 10, Signaltyp 2MMF) Die Markierung der Signifikanzniveaus \mathbf{p} erfolgt gemäß Tabelle 2.4.

	Korrelation zu θ				Korrelation zur Referenz			
	A	MA	MA_NN	MA_LF	A	MA	MA_NN	MA_LF
t_p	-0.27 [†]	-0.14 [†]	-0.31 [†]	0.16 [†]	-0.21 [†]	-0.06 [†]	-0.26 [†]	0.15 [†]
t_e	-0.17 [†]	0.06 [†]	-0.26 [†]	0.21 [†]	-0.23 [†]	-0.04 [†]	-0.34 [†]	0.15 [†]
t_a	0.49 [†]	0.71 [†]	0.62*	0.78	0.45 [†]	0.72 [†]	0.63*	0.76 [†]
E_e	—	—	—	—	—	—	—	—
$t_e - t_p$	0.46 [†]	0.42 [†]	0.27 [†]	0.49 [†]	-0.30 [†]	-0.24 [†]	-0.07 [†]	0.03 [†]
OQ_1	0.08 [†]	0.15 [†]	0.10 [†]	0.33 [†]	-0.21 [†]	0.01 [†]	-0.25 [†]	0.29 [†]
OQ_2	0.12 [†]	0.26 [†]	0.14 [†]	0.40 [†]	-0.13 [†]	0.16 [†]	-0.19 [†]	0.47 [†]
SQ_1	-0.31 [†]	-0.04 [†]	-0.10 [†]	-0.01 [†]	-0.31 [†]	-0.06 [†]	-0.09 [†]	-0.03 [†]
SQ_2	-0.36 [†]	-0.28 [†]	-0.30 [†]	-0.19 [†]	-0.32 [†]	-0.31 [†]	-0.26 [†]	-0.17 [†]
CQ_1	0.46 [†]	0.42 [†]	0.27 [†]	0.49 [†]	-0.30 [†]	-0.24 [†]	-0.07 [†]	0.03 [†]
CQ_2	0.46 [†]	0.46 [†]	0.26 [†]	0.52*	-0.42 [†]	-0.12 [†]	-0.26 [†]	0.25 [†]
PSP	0.34 [†]	0.42 [†]	0.20 [†]	0.52*	-0.20 [†]	0.06 [†]	-0.14 [†]	0.42 [†]

Tabelle C.7: Verschiedenheit der Verteilungen nach dem t -Test für Parameter t_p berechnet aus allen Daten aller Sprecher. Ein „-“ zeigt an, daß der Mittelwert für die Phonationsart in der linken Spalte kleiner als derjenige für die Phonationsart in der zweitobersten Zeile ist. Die Signifikanzniveaus \mathbf{p} der Unterschiede zwischen den Verteilungen sind durch verschiedene Symbole markiert (\bullet : $\mathbf{p} < 0.001$ (hochsignifikant); $+$: $0.001 < \mathbf{p} < 0.01$ (signifikant); $*$: $0.01 < \mathbf{p} < 0.05$ (schwach signifikant); kein Symbol: $\mathbf{p} \geq 0.05$ (insignifikant)).

	Ordnung 10			Ordnung 13			Ordnung 16		
	gepreßt	laut	hypo	gepreßt	laut	hypo	gepreßt	laut	hypo
RD									
normal	\bullet	.	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet
gepreßt		-+	.		-	.		- \bullet	\bullet
laut			+.+			*			\bullet
A									
normal	+	\bullet	\bullet	-	+.+	\bullet	\bullet	.	*
gepreßt		*	\bullet		\bullet	\bullet		-*	-
laut			\bullet			.			.
MA_NN									
normal	+	\bullet	\bullet	.	\bullet	\bullet	*	-	*
gepreßt		+.+	\bullet		*	\bullet		-+	.
laut			\bullet			+.+			+.+
MA									
normal	*	\bullet	\bullet	.	.	\bullet	+.+	-*	\bullet
gepreßt		+.+	\bullet		.	\bullet		- \bullet	*
laut			\bullet			+.+			\bullet

Tabelle C.8: Verschiedenheit der Verteilungen nach dem t -Test für Parameter t_e berechnet aus allen Daten aller Sprecher. Ein „-“ zeigt an, daß der Mittelwert für die Phonationsart in der linken Spalte kleiner als derjenige für die Phonationsart in der zweitobersten Zeile ist. Die Signifikanzniveaus \mathbf{p} der Unterschiede zwischen den Verteilungen sind durch verschiedene Symbole markiert (\bullet : $\mathbf{p} < 0.001$ (hochsignifikant); $+$: $0.001 < \mathbf{p} < 0.01$ (signifikant); $*$: $0.01 < \mathbf{p} < 0.05$ (schwach signifikant); kein Symbol: $\mathbf{p} \geq 0.05$ (insignifikant)).

	Ordnung 10			Ordnung 13			Ordnung 16		
	gepreßt	laut	hypo	gepreßt	laut	hypo	gepreßt	laut	hypo
RD									
normal	\bullet	.	$+$	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet
gepreßt		$-+$.		$-$.		$-$	\bullet
laut			$*$.			\bullet
A									
normal	\bullet	\bullet	\bullet	.	\bullet	\bullet	\bullet	$*$	\bullet
gepreßt		$*$	\bullet		$+$	$+$		$-*$.
laut			$+$.			$+$
MA_NN									
normal	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	$*$	$-$	\bullet
gepreßt		$*$	\bullet		$*$	\bullet		$-+$	$*$
laut			$+$			$+$			\bullet
MA									
normal	\bullet	\bullet	\bullet	$*$	$+$	\bullet	\bullet	$-$	\bullet
gepreßt		$+$	\bullet		.	\bullet		$-$	\bullet
laut			$+$			$+$			\bullet

Tabelle C.9: Verschiedenheit der Verteilungen nach dem t -Test für Parameter t_a berechnet aus allen Daten aller Sprecher. Ein „-“ zeigt an, daß der Mittelwert für die Phonationsart in der linken Spalte kleiner als derjenige für die Phonationsart in der zweitobersten Zeile ist. Die Signifikanzniveaus \mathbf{p} der Unterschiede zwischen den Verteilungen sind durch verschiedene Symbole markiert (\bullet : $\mathbf{p} < 0.001$ (hochsignifikant); $+$: $0.001 < \mathbf{p} < 0.01$ (signifikant); $*$: $0.01 < \mathbf{p} < 0.05$ (schwach signifikant); kein Symbol: $\mathbf{p} \geq 0.05$ (insignifikant)).

	Ordnung 10			Ordnung 13			Ordnung 16		
	gepreßt	laut	hypo	gepreßt	laut	hypo	gepreßt	laut	hypo
RD									
normal	\bullet	.	$-\bullet$	\bullet	\bullet	-	\bullet	\bullet	\bullet
gepreßt		$-+$	$-\bullet$		-	$-\bullet$		$-\bullet$	\bullet
laut			$-\bullet$			$-\bullet$			\bullet
A									
normal	-	$-\bullet$	$-+$	-	-	-	$+$.	\bullet
gepreßt		$-*$	$-*$		-	.		$-+$	\bullet
laut			-			.			\bullet
MA_NN									
normal	-	-	$-+$.	.	.	\bullet	$+$	\bullet
gepreßt		-	$-*$		-	.		-	\bullet
laut			$-*$.			\bullet
MA									
normal	$-*$	$-\bullet$	$-\bullet$	$+$	-	.	\bullet	.	\bullet
gepreßt		-	$-+$		$-+$	-		$-+$	$+$
laut			$-+$.			\bullet

Tabelle C.10: Verschiedenheit der Verteilungen nach dem t -Test für Parameter E_e berechnet aus allen Daten aller Sprecher. Ein „-“ zeigt an, daß der Mittelwert für die Phonationsart in der linken Spalte kleiner als derjenige für die Phonationsart in der zweitobersten Zeile ist. Die Signifikanzniveaus \mathbf{p} der Unterschiede zwischen den Verteilungen sind durch verschiedene Symbole markiert (\bullet : $\mathbf{p} < 0.001$ (hochsignifikant); $+$: $0.001 < \mathbf{p} < 0.01$ (signifikant); $*$: $0.01 < \mathbf{p} < 0.05$ (schwach signifikant); kein Symbol: $\mathbf{p} \geq 0.05$ (insignifikant)).

	Ordnung 10			Ordnung 13			Ordnung 16		
	gepreßt	laut	hypo	gepreßt	laut	hypo	gepreßt	laut	hypo
RD									
normal	-	- \bullet	\bullet	\bullet *	-	\bullet	\bullet +	-	\bullet +
gepreßt		-	\bullet		-+	\bullet		-+	\bullet *
laut			\bullet			\bullet			\bullet +
A									
normal	-*	- \bullet	.	-*	- \bullet	.	-	-*	\bullet
gepreßt		-	\bullet *		.	\bullet *		.	\bullet +
laut			\bullet +			\bullet *			\bullet
MA_NN									
normal	-*	- \bullet	.	-	- \bullet	\bullet	.	.	.
gepreßt		-	.		.	\bullet +		.	.
laut			\bullet *			\bullet			.
MA									
normal	-*	- \bullet	\bullet	.	- \bullet	\bullet	-	-	\bullet
gepreßt		-	\bullet		- \bullet	\bullet		-	\bullet
laut			\bullet			\bullet			\bullet

Tabelle C.11: Verschiedenheit der Verteilungen nach dem t -Test für Parameter OQ_1 berechnet aus allen Daten aller Sprecher. Ein „-“ zeigt an, daß der Mittelwert für die Phonationsart in der linken Spalte kleiner als derjenige für die Phonationsart in der zweitobersten Zeile ist. Die Signifikanzniveaus \mathbf{p} der Unterschiede zwischen den Verteilungen sind durch verschiedene Symbole markiert (\bullet : $\mathbf{p} < 0.001$ (hochsignifikant); $+$: $0.001 < \mathbf{p} < 0.01$ (signifikant); $*$: $0.01 < \mathbf{p} < 0.05$ (schwach signifikant); kein Symbol: $\mathbf{p} \geq 0.05$ (insignifikant)).

	Ordnung 10			Ordnung 13			Ordnung 16		
	gepreßt	laut	hypo	gepreßt	laut	hypo	gepreßt	laut	hypo
RD									
normal	\bullet	.	.	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet
gepreßt		-+	-		-	.		- \bullet	\bullet
laut			.			.			\bullet
A									
normal	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet
gepreßt		-	-		.	.		-	\bullet
laut			.			.			\bullet
MA_NN									
normal	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	+	\bullet
gepreßt		.	.		.	\bullet		-	\bullet
laut			.			+			\bullet
MA									
normal	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	.	\bullet
gepreßt		.	.		-	\bullet		- \bullet	\bullet
laut			.			\bullet			\bullet

Tabelle C.12: Verschiedenheit der Verteilungen nach dem t -Test für Parameter OQ_2 berechnet aus allen Daten aller Sprecher. Ein „-“ zeigt an, daß der Mittelwert für die Phonationsart in der linken Spalte kleiner als derjenige für die Phonationsart in der zweitobersten Zeile ist. Die Signifikanzniveaus \mathbf{p} der Unterschiede zwischen den Verteilungen sind durch verschiedene Symbole markiert (\bullet : $\mathbf{p} < 0.001$ (hochsignifikant); $+$: $0.001 < \mathbf{p} < 0.01$ (signifikant); $*$: $0.01 < \mathbf{p} < 0.05$ (schwach signifikant); kein Symbol: $\mathbf{p} \geq 0.05$ (insignifikant)).

	Ordnung 10			Ordnung 13			Ordnung 16		
	gepreßt	laut	hypo	gepreßt	laut	hypo	gepreßt	laut	hypo
RD									
normal	\bullet	.	.	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet
gepreßt		-+	-		-	-		- \bullet	\bullet
laut			.			.			\bullet
A									
normal	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet
gepreßt		-	-		.	.		-*	\bullet
laut			.			.			\bullet
MA_NN									
normal	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet *	\bullet
gepreßt		.	.		.	\bullet		-*	\bullet
laut			.			\bullet +			\bullet
MA									
normal	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	.	\bullet
gepreßt		.	.		-	\bullet		- \bullet	\bullet
laut			.			\bullet			\bullet

Tabelle C.13: Verschiedenheit der Verteilungen nach dem t -Test für Parameter SQ_1 berechnet aus allen Daten aller Sprecher. Ein „-“ zeigt an, daß der Mittelwert für die Phonationsart in der linken Spalte kleiner als derjenige für die Phonationsart in der zweitobersten Zeile ist. Die Signifikanzniveaus \mathbf{p} der Unterschiede zwischen den Verteilungen sind durch verschiedene Symbole markiert (\bullet : $\mathbf{p} < 0.001$ (hochsignifikant); $+$: $0.001 < \mathbf{p} < 0.01$ (signifikant); $*$: $0.01 < \mathbf{p} < 0.05$ (schwach signifikant); kein Symbol: $\mathbf{p} \geq 0.05$ (insignifikant)).

	Ordnung 10			Ordnung 13			Ordnung 16		
	gepreßt	laut	hypo	gepreßt	laut	hypo	gepreßt	laut	hypo
RD									
normal	- \bullet	-*	- \bullet	- \bullet	- \bullet	- \bullet	- \bullet	-+	- \bullet
gepreßt		.	- \bullet		.	-*		.	- \bullet
laut			- \bullet			- \bullet			- \bullet
A									
normal	- \bullet	- \bullet	.	- \bullet	- \bullet	- \bullet	-	-+	- \bullet
gepreßt			-	- \bullet
laut			+			.			- \bullet
MA_NN									
normal	- \bullet	-	.	- \bullet	-+	- \bullet	-	-+	- \bullet
gepreßt		.	.		+	-		-	- \bullet
laut			.			-+			- \bullet
MA									
normal	- \bullet	-	.	- \bullet	- \bullet	- \bullet	- \bullet	-*	- \bullet
gepreßt		+	.		*	-		.	- \bullet
laut			*			- \bullet			- \bullet

Tabelle C.14: Verschiedenheit der Verteilungen nach dem t -Test für Parameter SQ_2 berechnet aus allen Daten aller Sprecher. Ein „-“ zeigt an, daß der Mittelwert für die Phonationsart in der linken Spalte kleiner als derjenige für die Phonationsart in der zweitobersten Zeile ist. Die Signifikanzniveaus \mathbf{p} der Unterschiede zwischen den Verteilungen sind durch verschiedene Symbole markiert (\bullet : $\mathbf{p} < 0.001$ (hochsignifikant); $+$: $0.001 < \mathbf{p} < 0.01$ (signifikant); $*$: $0.01 < \mathbf{p} < 0.05$ (schwach signifikant); kein Symbol: $\mathbf{p} \geq 0.05$ (insignifikant)).

	Ordnung 10			Ordnung 13			Ordnung 16		
	gepreßt	laut	hypo	gepreßt	laut	hypo	gepreßt	laut	hypo
RD									
normal	- \bullet	-*	-*	- \bullet	- \bullet	- \bullet	- \bullet	- \bullet	- \bullet
gepreßt		.+	-		.	-		. \bullet	- \bullet
laut			-			-			- \bullet
A									
normal	- \bullet	-	.	- \bullet	- \bullet	-*	-.+	-.+	- \bullet
gepreßt		.+	. \bullet		.*	.		.	- \bullet
laut			.+			.			- \bullet
MA_NN									
normal	-.+	-	.	- \bullet	-	- \bullet	-*	- \bullet	- \bullet
gepreßt		.+	. \bullet		.+	-		-	- \bullet
laut			.*			-.+			- \bullet
MA									
normal	-*	.	. \bullet	- \bullet	-.+	- \bullet	- \bullet	-.+	- \bullet
gepreßt		.+	. \bullet		.+	-		.	- \bullet
laut			.+			-.+			- \bullet

Tabelle C.15: Verschiedenheit der Verteilungen nach dem t -Test für Parameter CQ_1 berechnet aus allen Daten aller Sprecher. Ein „-“ zeigt an, daß der Mittelwert für die Phonationsart in der linken Spalte kleiner als derjenige für die Phonationsart in der zweitobersten Zeile ist. Die Signifikanzniveaus \mathbf{p} der Unterschiede zwischen den Verteilungen sind durch verschiedene Symbole markiert (\bullet : $\mathbf{p} < 0.001$ (hochsignifikant); $+$: $0.001 < \mathbf{p} < 0.01$ (signifikant); $*$: $0.01 < \mathbf{p} < 0.05$ (schwach signifikant); kein Symbol: $\mathbf{p} \geq 0.05$ (insignifikant)).

	Ordnung 10			Ordnung 13			Ordnung 16		
	gepreßt	laut	hypo	gepreßt	laut	hypo	gepreßt	laut	hypo
RD									
normal	\bullet	\bullet^*	.	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet
gepreßt		$-+$	-		-	-		\bullet	\bullet
laut			.			.			\bullet
A									
normal	\bullet	\bullet	.	\bullet	\bullet	.	\bullet	\bullet	\bullet
gepreßt		-	\bullet		.	$-+$		-	\bullet
laut			\bullet			$-+$			\bullet
MA_NN									
normal	\bullet	\bullet	$+$	\bullet	\bullet	\bullet	.	.	\bullet
gepreßt		-	$-+$.	.		-	\bullet
laut			$-*$.			\bullet
MA									
normal	\bullet	\bullet	\bullet^*	\bullet	\bullet	\bullet	\bullet	.	\bullet
gepreßt		-	$-+$		-	.		\bullet	\bullet
laut			$-*$.			\bullet

Tabelle C.16: Verschiedenheit der Verteilungen nach dem t -Test für Parameter CQ_2 berechnet aus allen Daten aller Sprecher. Ein „-“ zeigt an, daß der Mittelwert für die Phonationsart in der linken Spalte kleiner als derjenige für die Phonationsart in der zweitobersten Zeile ist. Die Signifikanzniveaus \mathbf{p} der Unterschiede zwischen den Verteilungen sind durch verschiedene Symbole markiert (\bullet : $\mathbf{p} < 0.001$ (hochsignifikant); $+$: $0.001 < \mathbf{p} < 0.01$ (signifikant); $*$: $0.01 < \mathbf{p} < 0.05$ (schwach signifikant); kein Symbol: $\mathbf{p} \geq 0.05$ (insignifikant)).

	Ordnung 10			Ordnung 13			Ordnung 16		
	gepreßt	laut	hypo	gepreßt	laut	hypo	gepreßt	laut	hypo
RD									
normal	\bullet	\bullet^*	-	\bullet	\bullet	\bullet^+	\bullet	\bullet	\bullet
gepreßt		$-^+$	$-^*$		-	-		$-^*$	\bullet
laut			-			-			\bullet
A									
normal	\bullet	\bullet	-	\bullet	\bullet	-	\bullet	\bullet	\bullet
gepreßt		$-^*$	$-^*$		\bullet	$-^*$		-	\bullet
laut			$-^*$			$-^*$			\bullet
MA_NN									
normal	\bullet	\bullet	-	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet
gepreßt		-	$-^*$		-	\bullet		-	\bullet^*
laut			-			\bullet			\bullet^*
MA									
normal	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet^*	\bullet
gepreßt		-	$-^+$		-	\bullet		$-^*$	\bullet
laut			$-^*$			\bullet			\bullet

Tabelle C.17: Verschiedenheit der Verteilungen nach dem t -Test für Parameter PSP berechnet aus allen Daten aller Sprecher. Ein „-“ zeigt an, daß der Mittelwert für die Phonationsart in der linken Spalte kleiner als derjenige für die Phonationsart in der zweitobersten Zeile ist. Die Signifikanzniveaus \mathbf{p} der Unterschiede zwischen den Verteilungen sind durch verschiedene Symbole markiert (\bullet : $\mathbf{p} < 0.001$ (hochsignifikant); $+$: $0.001 < \mathbf{p} < 0.01$ (signifikant); $*$: $0.01 < \mathbf{p} < 0.05$ (schwach signifikant); kein Symbol: $\mathbf{p} \geq 0.05$ (insignifikant)).

	Ordnung 10			Ordnung 13			Ordnung 16		
	gepreßt	laut	hypo	gepreßt	laut	hypo	gepreßt	laut	hypo
RD									
normal	\bullet	\bullet^*	$-^*$	\bullet	\bullet	\cdot	\bullet	\bullet	\bullet
gepreßt		$-^*$	$-^{\bullet}$		$-$	$-^+$		$-^{\bullet}$	\cdot^{\bullet}
laut			$-^{\bullet}$			$-^+$			\cdot^{\bullet}
A									
normal	\bullet	\bullet	$-$	\bullet	\bullet	$-$	\bullet	\bullet	\bullet
gepreßt		$-^*$	$-$		\cdot	$-$		$-$	\cdot^+
laut			$-$			$-$			\cdot^{\bullet}
MA_NN									
normal	\bullet	\bullet	$-$	\bullet	\bullet	\bullet	\bullet	\cdot^+	$-$
gepreßt		$-$	$-$		\cdot	$-$		$-$	$-$
laut			$-$			$-$			$-$
MA									
normal	\bullet	\bullet	$-$	\bullet	\bullet	$-$	\bullet	\cdot^*	$-$
gepreßt		$-$	$-$		$-$	$-$		$-^{\bullet}$	$-$
laut			$-$			$-$			$-$

Tabelle C.18: Sukzessive Reduktion der Hauptkomponenten bei der Diskrimination zwischen verschiedenen Phonationsarten unter Verwendung aller Stimmdaten (Optimierungsmethode MA_NN, $p = 10$). Die Güte der Diskrimination unter Verwendung aller 4 SVD-Vektoren ist in runden Klammern angegeben. Die weiteren Zeilen stellen die „Unentbehrlichkeiten“ nach Hartung und Elpelt (1995) für die einzelnen Vektoren dar, wobei der als nächstes wegzulassende Parameter durch $\langle \rangle$ markiert ist. Die letzte Spalte gibt an, ob nach Weglassen des markierten Parameters die Diskrimination immer noch signifikant zum Niveau $\mathbf{p} < 0.01$ ist.

Phonationsarten	(Güte)	Vektor 1	Vektor 2	Vektor 3	Vektor 4	Signifikanz
normal – laut	(64.3%)					+
		0.066	0.071	0.008	$\langle 0.005 \rangle$	+
		0.223	0.081	$\langle 0.012 \rangle$	-	+
		0.253	$\langle 0.101 \rangle$	-	-	-
normal – hypo	(76.3%)					+
		0.338	0.413	$\langle 0.026 \rangle$	0.150	+
		0.328	0.407	-	$\langle 0.168 \rangle$	+
		$\langle 0.187 \rangle$	0.386	-	-	-
gepreßt – normal	(63.2%)					+
		0.078	0.036	$\langle 0.002 \rangle$	0.003	+
		0.085	0.038	-	$\langle 0.005 \rangle$	+
		0.265	$\langle 0.041 \rangle$	-	-	-
gepreßt – laut	(55.0%)					-
gepreßt – hypo	(82.7%)					+
		$\langle 0.037 \rangle$	0.249	0.124	0.463	+
		-	0.221	$\langle 0.129 \rangle$	0.426	+
		-	$\langle 0.217 \rangle$	-	0.368	-
laut – hypo	(82.7%)					+
		$\langle 0.030 \rangle$	0.129	0.101	0.484	+
		-	0.114	$\langle 0.102 \rangle$	0.455	+
		-	$\langle 0.115 \rangle$	-	0.421	-

Tabelle C.19: Sukzessive Reduktion der Hauptkomponenten bei der Diskrimination zwischen verschiedenen Phonationsarten unter Verwendung aller Stimmdaten (Optimierungsmethode RD, $p = 16$). Die Güte der Diskrimination unter Verwendung aller 4 SVD-Vektoren ist in runden Klammern angegeben. Die weiteren Zeilen stellen die „Unentbehrlichkeiten“ nach Hartung und Elpelt (1995) für die einzelnen Vektoren dar, wobei der als nächstes wegzulassende Parameter durch $\langle \rangle$ markiert ist. Die letzte Spalte gibt an, ob nach Weglassen des markierten Parameters die Diskrimination immer noch signifikant zum Niveau $p < 0.01$ ist.

Phonationsarten	(Güte)	Vektor 1	Vektor 2	Vektor 3	Vektor 4	Signifikanz
normal – laut	(67.9%)					+
		0.028	0.042	0.013	$\langle 0.001 \rangle$	+
		0.084	0.092	$\langle 0.016 \rangle$	-	+
		0.149	$\langle 0.078 \rangle$	-	-	-
normal – hypo	(85.0%)					+
		0.075	0.116	0.266	$\langle 0.024 \rangle$	+
		$\langle 0.263 \rangle$	0.425	0.446	-	+
		-	0.408	$\langle 0.382 \rangle$	-	-
gepreßt – normal	(68.2%)					+
		0.033	0.015	0.016	$\langle 0.000 \rangle$	+
		0.093	0.050	$\langle 0.024 \rangle$	-	+
		0.171	$\langle 0.029 \rangle$	-	-	-
gepreßt – laut	(57.5%)					+
		0.003	$\langle 0.000 \rangle$	0.003	0.000	+
		0.069	-	0.014	$\langle 0.000 \rangle$	-
gepreßt – hypo	(78.6%)					+
		0.071	$\langle 0.061 \rangle$	0.166	0.112	+
		$\langle 0.016 \rangle$	-	0.199	0.342	+
		-	-	$\langle 0.307 \rangle$	0.385	-
laut – hypo	(85.0%)					+
		0.155	0.197	0.347	$\langle 0.142 \rangle$	+
		$\langle 0.472 \rangle$	0.664	0.567	-	+
		-	0.448	$\langle 0.129 \rangle$	-	-

D. Liste der häufig verwendeten Abkürzungen

AR	<i>auto regressive</i>	S. 15
CQ	<i>closing quotient</i>	S. 38
DAP	<i>discrete all-pole modeling</i>	S. 41
EGG	Elektroglottogramm	S. 9
FFT	<i>fast Fourier transform</i>	S. 47
FIR	<i>finite impulse response</i>	S. 50
GA	genetischer Algorithmus	S. 17
hypo	hypofunktionelle („schwache“) Phonation	S. 34
IF	Inversfilterung	S. 13
IIR	<i>infinite impulse response</i>	S. 50
LF	Liljencrants-Fant (Glottismodell)	S. 16
LP	<i>linear prediction</i>	S. 11
LPC	<i>linear predictive coding</i>	S. 11
MA	<i>moving average</i>	S. 15
OQ	<i>open quotient</i>	S. 37
PSP	<i>parabolic spectral parameter</i>	S. 39
QF	Quelle-Filter	S. 10
SA	<i>simulated annealing</i>	S. 17
SIM	simultane Inversfilterung und automatische Anpassung eines Glottismodells	S. 49
SQ	<i>speed quotient</i>	S. 38
SVD	<i>singular value decomposition</i>	S. 40

Danksagung

An dieser Stelle bedanke ich mich bei Prof. Manfred R. Schroeder, dessen fortwährendes Engagement im Institut und Seminar den Grundstock für diese Arbeit legte und mich den Zusammenhang zu anderen physikalischen Gebieten nicht aus den Augen verlieren ließ, und bei Dr. Hans Werner Strube, der die kritischen Punkte in wissenschaftlichen Fragestellungen sofort erkannte und stets ein offenes Ohr bei sich neu ergebenden Problemen besaß, für die Betreuung dieser Arbeit.

Mein weiterer Dank gilt Prof. Eberhard Kruse, dem Anreger des Projektes, innerhalb dessen diese Dissertation angefertigt wurde, und der unter anderem durch die direkte medizinische Anbindung der Ergebnisse sehr zur Motivation in der Durchführung der Arbeit beitrug.

Im direkten Arbeitsgruppenumfeld danke ich vor allem Dirk Michaelis für die enge, gute Zusammenarbeit, die fruchtbaren Diskussionen und seine Zuverlässigkeit bei der Lösung von fachlichen wie organisatorischen Problemen.

Den weiteren, momentanen wie ehemaligen Mitgliedern meiner Arbeitsgruppe Knut Müller, Heiko Freienstein, Olaf Schreiner, Jannis Kiosses, Tilmann Illg, Jan Lessing, Joachim Kaufmann und Hans Wilmers sowie Dr. Ingo Stuermer von der Forschergruppe „Stimme und Sinnesentwicklung“ danke ich für die angenehme Arbeitsatmosphäre und die Möglichkeit zum regen inhaltlichen Austausch zwischen thematisch benachbarten Forschungsgebieten.

Ein ganz persönliches Dankeschön geht an Cornelia Mascher, die mir den emotionalen Rückhalt für meine Arbeit gab.

