

Vocal Fold Dynamics –  
Quantification and Model-based  
Classification

Der Technischen Fakultät der  
Universität Erlangen-Nürnberg  
zur Erlangung des Grades

DOKTOR-INGENIEUR

vorgelegt von  
Dipl.-Ing. Tobias Wurzbacher

Erlangen 2007

Als Dissertation genehmigt von  
der Technischen Fakultät der  
Universität Erlangen-Nürnberg

Tag der Einreichung : 11.10.2007

Tag der Promotion : 19.12.2007

Dekan : Prof. Dr. Huber

Berichterstatter : Prof. Dr. Dr. U. Eysholdt

Prof. Dr. A. Delgado

Die Dynamik der Stimmlippen –  
Quantifizierung und Modellbasierte  
Klassifikation

Der Technischen Fakultät der  
Universität Erlangen-Nürnberg  
zur Erlangung des Grades

DOKTOR-INGENIEUR

vorgelegt von  
Dipl.-Ing. Tobias Wurzbacher

Erlangen 2007



Kommunikationsstörungen - Berichte aus Phoniatrie und  
Pädaudiologie

Band 17

**Tobias Wurzbacher**

**Vocal Fold Dynamics –  
Quantification and Model-based Classification**

D 29 (Diss. Universität Erlangen-Nürnberg)

Shaker Verlag  
Aachen 2008

**Bibliographic information published by the Deutsche Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Zugl.: Erlangen-Nürnberg, Univ., Diss., 2007

Copyright Shaker Verlag 2008

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8322-6854-1

ISSN 1436-1175

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen

Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9

Internet: [www.shaker.de](http://www.shaker.de) • e-mail: [info@shaker.de](mailto:info@shaker.de)

# Danksagung

Meinem Doktorvater Prof. Dr. Dr. Ulrich Eysholdt danke ich herzlich für die Aufnahme in seine interdisziplinäre Arbeitsgruppe und die Überlassung des Themas. Ich danke für die Freiheit eigene Ideen einzubringen und für die stete Unterstützung mit hilfreichen Ratschlägen.

Bei Herrn Prof Dr. Antonio Delgado möchte ich mich für die gewährte Unterstützung und die Übernahme des Korreferats bedanken.

Ich danke meinen beiden Betreuern PD Dr. Jörg Lohscheller und PD Dr. Michael Döllinger, die wertvolle Hinweise für die wissenschaftliche Durchführung der Arbeit gaben und sich viel Zeit für Diskussionen nahmen. Insbesondere danke ich Ihnen für die konstruktiven Kritiken, die sich bei der Durchsicht meiner Manuskripte ergaben.

Meinen Kollegen Herrn Dr. Raphael Schwarz, Martin Burger und Daniel Voigt danke ich für die angenehme und fruchtbare Arbeitsatmosphäre. Vielen Dank für die vielen Diskussionen, Anregungen und die Zusammenarbeit. Ein herzlicher Dank geht auch an den Studienarbeiter Ingmar Voigt.

Weiterhin bedanke ich mich bei Dr. Hikmet Toy, PD Dr. Maria Schuster und Prof. Dr. Kummer für die Durchführung der Hochgeschwindigkeitsaufnahmen und für die Erklärung medizinischer Gesichtspunkte. Ebenso bedanke ich mich bei Prof. Dr. Dr. Ulrich Hoppe für die intensive Zusammenarbeit.

Ganz besonders möchte ich mich bei meiner Frau Stephanie Wurzbacher, meinen Eltern und meiner Familie sowie meinen Freunden bedanken, die mir den Rücken während der Anfertigung dieser Arbeit gestärkt haben.

Außerdem bedanke ich mich bei allen Mitarbeiterinnen und Mitarbeitern der Abteilung Phoniatrie und Pädaudiologie des Universitätsklinikums Erlangen und all denen, die auf unterschiedliche Weise zum Gelingen dieser Arbeit beigetragen haben und hier nicht namentlich erwähnt sind.

Die vorliegende Arbeit wurde im Rahmen des Projekts DFG HO 2177/1-3 der Deutschen Forschungsgemeinschaft gefördert.

Erlangen, Oktober 2007



# Zusammenfassung

Heiserkeit, als typisches Symptom einer Stimmstörung, wird unter anderem auf aperiodische und asymmetrische Stimmlippenschwingungen zurückgeführt. Mit Hilfe von endoskopischen Hochgeschwindigkeitsvideos lässt sich die Dynamik der Stimmlippen während der Phonation in Echtzeit aufnehmen. Die so gewonnenen digitalen Bilddaten werden für eine rechnergestützte Analyse des Schwingungsmusters genutzt. Ziel der Auswertung ist sowohl eine objektive als auch quantitative Beschreibung der Stimmlippendynamik zu erreichen, um automatisch pathologische von gesunden Schwingungsformen unterscheiden zu können.

Für die Beurteilung der Schwingungsform wird üblicherweise ein stationäres Phonationsparadigma verwendet, d. h. die Untersuchung erfolgt bei konstanter Grundfrequenz und Intensität eines gehaltenen Vokals. Darüber hinaus wird in dieser Arbeit eine nicht-stationäre Phonation in Form einer aktiven Tonhöhenänderung untersucht, um auch eine variierende Stimmlippendynamik in die Bewertung einfließen zu lassen. Anhand von klinischen Daten wird gezeigt, dass eine nicht-stationäre Phonation Vorteile bei der Bewertung von Stimmlippenschwingungen bietet.

Ein modellbasierter Ansatz mit zeitvarianten Parametern stellt die Grundlage für die Analyse der aus den Bilddaten extrahierten Schwingungen dar. Die Modelldynamik wird segmentweise an die beobachteten stationären und nicht-stationären Schwingungsmuster angepasst. Dies geschieht mittels einer automatischen Parameteroptimierung, an deren Ende die Größen der ermittelten Modellparameter Auskunft über Symmetrie und Regularität der Stimmlippenschwingungen geben. Vorgestellt werden zwei biomechanische Stimmlippenmodelle. Das erste spezialisiert sich auf links-rechts Asymmetrien aus dem medialen Stimmlippendrittel. Darauf aufbauend wird beim zweiten Modell auch die longitudinale Schwingungscharakteristik aus dem dorsalen, medialen und ventralen Stimmlippenbereich ausgewertet. Die Funktionalität wird mit synthetischen Daten überprüft und auf reale klinische Daten zur Klassifizierung der Stimmlippenschwingungen angewandt.

Für eine Quantifizierung der Stimmlippendynamik sind insbesondere auch metrische Größen wie Länge, Auslenkung und Geschwindigkeiten von Interesse. Um diese aus den Endoskopieaufnahmen zu erhalten, wird ein Laserprojektionssystem vorgestellt, das, angebracht am Endoskop, zwei parallele Linien über die gesamte Stimmlippenbreite in die Aufnahmen einblendet. Eine Bildverarbeitung wird präsentiert, die eine automatische und robuste Detektion der Laserlinien und somit eine Kalibrierung der Hochgeschwindigkeitsvideos ermöglicht. Die Funktionalität wird anhand klinischer Aufnahmen demonstriert.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Fundamentals</b>	<b>5</b>
2.1	Anatomy and Physiology . . . . .	5
2.2	Voice . . . . .	8
2.2.1	Production . . . . .	8
2.2.2	Pathology . . . . .	9
2.3	Examination Methods . . . . .	10
2.3.1	Acoustic analysis . . . . .	10
2.4	High-Speed Recordings . . . . .	12
2.4.1	Origins of HS recordings . . . . .	12
2.4.2	HS camera system . . . . .	12
2.4.3	HS image series . . . . .	13
2.4.4	Image processing . . . . .	14
2.5	Phonation Paradigms . . . . .	19
2.6	Basic Vocal Fold Modeling . . . . .	21
2.6.1	Two-mass model: 2MM . . . . .	22
<b>3</b>	<b>Model-based Classification</b>	<b>27</b>
3.1	Corpus of Clinical Data . . . . .	27
3.2	Time-dependent Two-Mass Model: TTM . . . . .	29
3.3	Optimization Parameter Set . . . . .	30
3.3.1	Time-dependent scaling . . . . .	30
3.3.2	Initialization . . . . .	31
3.4	Objective Function . . . . .	32
3.5	Optimization . . . . .	33
3.5.1	Block-based optimization . . . . .	33
3.6	Classification Criteria . . . . .	35
3.7	Validation . . . . .	37
3.7.1	Methodology . . . . .	37

3.7.2	Results . . . . .	38
3.8	Application to Clinical Data . . . . .	40
3.8.1	Adaptation results . . . . .	42
3.8.2	Symmetry plane . . . . .	45
3.8.3	Classification results . . . . .	49
<b>4</b>	<b>Spatio-temporal Model-based Classification</b>	<b>51</b>
4.1	Corpus of Clinical Data . . . . .	52
4.2	Time-dependent Multi-Mass Model: TMM . . . . .	53
4.3	Optimization Parameter Set . . . . .	55
4.3.1	Time-dependent scaling . . . . .	57
4.3.2	Initialization . . . . .	57
4.4	Objective Function . . . . .	58
4.5	Optimization . . . . .	60
4.5.1	Block-based optimization . . . . .	60
4.6	Classification Criteria . . . . .	61
4.7	Validation . . . . .	63
4.7.1	Methodology . . . . .	63
4.7.2	Results . . . . .	63
4.8	Application to Clinical Data . . . . .	66
4.8.1	Adaptation results . . . . .	67
4.8.2	Symmetry plane . . . . .	68
4.8.3	Classification results . . . . .	71
<b>5</b>	<b>Calibration of High-Speed Recordings</b>	<b>75</b>
5.1	Using Laser Devices . . . . .	76
5.2	Laser Calibration Hardware . . . . .	77
5.2.1	Laser line projection system (LLPS) . . . . .	77
5.2.2	LLPS accuracy measurements . . . . .	77
5.2.3	Results LLPS accuracy . . . . .	78
5.3	Detection of Parallel Laser Lines . . . . .	79
5.3.1	Data reduction: region of interest (ROI) . . . . .	81
5.3.2	Data representation: kymogram construction . . . . .	82
5.3.3	Line detection step 1: kymogram analysis . . . . .	83
5.3.4	Line detection step 2: determining valid laser lines . . . . .	89
5.4	Application to Clinical Data . . . . .	92
5.4.1	Clinical recordings . . . . .	92
5.4.2	LLPS calibrated recordings . . . . .	92

5.4.3	Metrical units of laryngeal magnitudes . . . . .	96
5.4.4	Discussion . . . . .	98
<b>6</b>	<b>Summary and Outlook</b>	<b>101</b>
	<b>Abbreviations and Notations</b>	<b>105</b>
	<b>Bibliography</b>	<b>111</b>
	<b>List of Figures</b>	<b>123</b>
	<b>List of Tables</b>	<b>125</b>
	<b>Curriculum Vitae</b>	<b>127</b>



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Grundlagen</b>	<b>5</b>
2.1	Anatomie und Physiologie	5
2.2	Stimme	8
2.2.1	Produktion	8
2.2.2	Pathologie	9
2.3	Untersuchungsmethoden	10
2.3.1	Akustische Analyse	10
2.4	Hochgeschwindigkeitsaufnahmen	12
2.4.1	Entstehungsgeschichte	12
2.4.2	HS Kamerasystem	12
2.4.3	HS Bilddaten	13
2.4.4	Bildverarbeitung	14
2.5	Phonationsparadigmen	19
2.6	Grundlage der Stimmlippenmodellierung	21
2.6.1	Zwei-Massen-Modell: 2MM	22
<b>3</b>	<b>Modellbasierte Klassifikation</b>	<b>27</b>
3.1	Klinische Daten	27
3.2	Zeitabhängiges Zwei-Massen-Model: TTM	29
3.3	Optimierungsparameter	30
3.3.1	Zeitabhängige Skalierung	30
3.3.2	Initialisierung	31
3.4	Zielfunktion	32
3.5	Optimierung	33
3.5.1	Blockbasierte Optimierung	33
3.6	Klassifizierungskriterien	35
3.7	Validierung	37
3.7.1	Methode	37

3.7.2	Ergebnisse	38
3.8	Anwendung auf klinische Daten	40
3.8.1	Anpassungsergebnisse	42
3.8.2	Symmetrieebene	45
3.8.3	Ergebnisse der Klassifizierung	49
<b>4</b>	<b>Räumlich-Zeitliche modellbasierte Klassifizierung</b>	<b>51</b>
4.1	Klinische Daten	52
4.2	Zeitabhängiges Mehr-Massen-Modell: TMM	53
4.3	Optimierungsparameter	55
4.3.1	Zeitabhängige Skalierung	57
4.3.2	Initialisierung	57
4.4	Zielfunktion	58
4.5	Optimierung	60
4.5.1	Blockbasierte Optimierung	60
4.6	Klassifizierungskriterien	61
4.7	Validierung	63
4.7.1	Methode	63
4.7.2	Ergebnisse	63
4.8	Anwendung auf klinische Daten	66
4.8.1	Anpassungsergebnisse	67
4.8.2	Symmetrieebene	68
4.8.3	Ergebnisse der Klassifizierung	71
<b>5</b>	<b>Kalibrierung von Hochgeschwindigkeitsaufnahmen</b>	<b>75</b>
5.1	Verwendung von Laserprojektoren	76
5.2	Laserkalibrierungsgerät	77
5.2.1	Laserlinien-Projektionssystem (LLPS)	77
5.2.2	Messungen der Genauigkeit des LLPS	77
5.3	Detektion von parallelen Laserlinien	79
5.3.1	Datenreduktion: Region of Interest (ROI)	81
5.3.2	Datendarstellung: Kymogramkonstruktion	82
5.3.3	Schritt 1 der Detektion: Kymogramanalyse	83
5.3.4	Schritt 2 der Detektion: Zulässige Laserlinien	89
5.4	Anwendung auf klinische Daten	92
5.4.1	Klinische Aufnahmen	92
5.4.2	LLPS kalibrierte klinische Aufnahmen	92

5.4.3	Metrische Einheiten laryngealer Größen .....	96
5.4.4	Diskussion .....	98
<b>6</b>	<b>Zusammenfassung und Ausblick</b>	<b>101</b>
<b>7</b>	<b>Bezeichnungen und Abkürzungen</b>	<b>105</b>
<b>8</b>	<b>Literaturverzeichnis</b>	<b>111</b>
<b>9</b>	<b>Abbildungsverzeichnis</b>	<b>123</b>
<b>10</b>	<b>Tabellenverzeichnis</b>	<b>125</b>
<b>11</b>	<b>Lebenslauf</b>	<b>127</b>