Bildverarbeitung bei endoskopischen Hochgeschwindigkeitsaufnahmen der Stimmlippenbewegungen

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

DOKTOR-INGENIEUR

vorgelegt von
Dipl.-Technomath. Denis Dubrovskiy
aus Omsk (Russland)

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

Tag der mündlichen Prüfung:	20.03.2017
Vorsitzender des Promotionsorgans:	Prof. Dr.-Ing. Reinhard Lerch
Gutachter:	Prof. Dr.-Ing. Michael Döllinger Prof. Dr.-Ing. habil. Andreas Maier

Kommunikationsstörungen - Berichte aus Phoniatrie und
Pädaudiologie

Band 24

Denis Dubrovskiy

**Bildverarbeitung bei endoskopischen
Hochgeschwindigkeitsaufnahmen
der Stimmlippenbewegungen**

D 29 (Diss. Universität Erlangen-Nürnberg)

Shaker Verlag
Aachen 2017

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Zugl.: Erlangen-Nürnberg, Univ., Diss., 2017

Copyright Shaker Verlag 2017

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe, der Speicherung in Datenverarbeitungsanlagen und der Übersetzung, vorbehalten.

Printed in Germany.

ISBN 978-3-8440-5351-7

ISSN 1436-1175

Shaker Verlag GmbH • Postfach 101818 • 52018 Aachen

Telefon: 02407 / 95 96 - 0 • Telefax: 02407 / 95 96 - 9

Internet: www.shaker.de • E-Mail: info@shaker.de

Danksagung

Meinem Doktorvater und Betreuer Prof. Dr. Michel Döllinger danke ich herzlichst für die Aufnahme in seine interdisziplinäre Arbeitsgruppe, zuerst als wissenschaftliche Hilfskraft, später mit einem Thema für die Diplomarbeit und weiterführend dann für die Promotion. Ich danke für die kostbaren Erfahrungen, die ich in wissenschaftlicher und klinischer Forschung machen durfte, sowie für die exzellente Möglichkeit, während unserer langjährigen Zusammenarbeit neue Ideen zu verfolgen und umsetzen zu können. Außerdem einen besonders herzlichen Dank für außergewöhnliche und vielseitige Unterstützung, sowie entgegengebrachtes Vertrauen und konstruktive Kritik.

Bei Herrn Prof. Dr. habil. Andreas Maier bedanke ich mich für die Übernahme des Koreferates.

Bei Prof. Dr. Dr. Ulrich Eysholdt, ehemaliger Leiter der Abteilung für Phoniatrie und Pädaudiologie, möchte ich mich aufrichtig für die Aufnahme in die Abteilung, sowie seine Unterstützung bedanken.

Meinen Kolleginnen und Kollegen Dr. Georg Luegmair, Dr. Björn Hüttner, Dr. Anke Ziethe, Dr. Stefan Kniesburges, Marion Semmler, Veronika Birk und Hossein Sadeghi danke ich vielmals für die hinreißende und fruchtbare Zusammenarbeit, die angenehme und entspannte Arbeitsatmosphäre, sowie die nette Unterhaltung nebenher, welche sowohl eine Abwechslung im Arbeitsalltag gebracht als auch meine Deutschkenntnisse positiv beeinflusst hat. Einen besonderen Dank möchte ich an meinen Kollegen Dr. Simon Petermann richten, der den Werdegang dieser Arbeit unmittelbar miterlebt und konstruktiv unterstützt hat. Danke für die zahlreichen vielfältigen Diskussionen und Gespräche, sowie die wertvollen Tipps zu meiner sprachlichen Verbesserung.

Meinen ärztlichen Kolleginnen und Kollegen Prof. Dr. Christopher Bohr, jetziger Leiter der Abteilung für Phoniatrie und Pädaudiologie, Dr. Stephan Dürr und Dr. Anne Schützenberger danke ich herzlich für die gute Zusammenarbeit, Unterstützung und begleitende Hilfe jeglicher Art.

Nicht zuletzt möchte ich mich recht herzlich bei Anna Lottner für ihre unschätzbare Hilfe bedanken.

Ganz besonders möchte ich mich herzlichst bei meiner ganzen Familie dafür bedanken, dass sie mir stets mit Rat und Tat zur Seite gestanden und an mich geglaubt hat. Speziell danke ich meinen Eltern sowie Großeltern, die mir beigebracht haben jede begonnene Aufgabe mit vollem Herzblut zu ihrem logischen Abschluss zu bringen.

Zur guter Letzt möchte ich mich bei allen nicht namentlich genannten Mitarbeiterinnen und Mitarbeiter der Abteilung für Phoniatrie und Pädaudiologie der HNO-Klinik des Universitätsklinikum Erlangen bedanken, die auf unterschiedliche Weise einen Beitrag zum erfolgreichen Abschluss dieser Arbeit beigetragen haben.

„Ich behaupte aber, dass in jeder besonderen Naturlehre nur so viel eigentliche Wissenschaft angetroffen werden könne, als darin Mathematik anzutreffen ist.“

- I. Kant, *Metaphysische Anfangsgründe der Naturwissenschaft* (1786)

Kurzfassung

Stimme und Sprache repräsentieren das wichtigste Ausdrucksmittel in der menschlichen Kommunikation. Eine Stimmstörung (Dysphonie) beeinträchtigt die Fähigkeit des Betroffenen zur Kommunikation, und wirkt sich negativ auf sein psychosoziales und berufliches Wohlbefinden aus, welche wichtige Charakteristika der Lebensqualität sind.

Phonation stellt den Prozess der Erzeugung des primären akustischen Signals im menschlichen Kehlkopf dar, und ist somit ein wichtiger Bestandteil der Stimmgebung. Eine gesunde Stimme zeichnet sich durch die Regularität, Periodizität sowie Symmetrie der Stimmlippenschwingungen während der Phonation aus. Daher ist die visuelle Inspektion des Kehlkopfes sowie die darauf basierende Schwingungsanalyse der Stimmlippen für die klinische Stimmdiagnostik von zentraler Bedeutung. Hierfür bietet die digitale Hochgeschwindigkeitsendoskopie (HSE) ein enormes Potential zur Quantifizierung der 2D-Stimmlippenbewegung, und stellt somit ein wirksames Instrument zur Beurteilung von Anatomie und Funktionalität des Kehlkopfes dar. Die Auswertung von HSE-Aufnahmen, wodurch die Daten über die Gestalt und Regularität der Stimmlippenschwingungen extrahiert werden, erfolgt unter Anwendung von Bildverarbeitungsverfahren. Da die Objektivität und Reliabilität der extrahierten HSE-Merkmale direkt mit der Funktionalität und Verlässlichkeit der eingesetzten Methoden verbunden sind, spielt die Bildverarbeitung eine entscheidende Rolle für die klinische Anwendbarkeit der objektiven Stimmdiagnostik.

In der vorliegenden Arbeit wird ein Bildverarbeitungsverfahren mit hohem Automatisierungsgrad erarbeitet, das dann eine objektive und präzise Auswertung von HSE-Aufnahmen der Stimmlippenschwingungen während der Phonation ermöglicht.

Es wird ein vollautomatisches zweistufigen Verfahren zur Segmentierung von HSE-Aufnahmen entwickelt, um die Daten über die 2D-Bewegung und Regularität der aufgenommenen Stimmlippenschwingungen in Form der festgestellten Stimmlippenkanten sowie der davon abgeleiteten glottalen Flächenfunktion (GAW) zu extrahieren. Hierfür wird anfangs die Position der Stimmlippen auf der Bildebene innerhalb einer HSE-Aufnahme durch die automatische Bestimmung einer Bildregion (ROI) abgeschätzt. Dies erhöht die Genauigkeit der Segmentierungsergebnisse, andererseits führt dies zur Reduzierung des Rechenaufwandes, der für die nachfolgende Stimmlippenkantendetektierung benötigt wird. Der ROI wird dabei anhand von Grauwerten der Bildpunkte, deren Veränderung über die Aufnahmezeit und deren Verteilungen innerhalb der HSE-Einzelbilder unter Berücksichtigung anatomischer und physiologischer Eigenschaften der Stimmlippen ermittelt. Dies erfolgt mit Hilfe der Hauptkomponentenanalyse (PCA) sowie der Approximation mittels verallgemeinerter Ricker-Wavelets. Die eigentliche Segmentierung der HSE-Aufnahme, was die Detektierung der Stimmlippenkanten innerhalb der festgestellten ROI darstellt, erfolgt im zweiten Schritt des Verfahrens. Hierfür wird das Verfahren der Aktiven-Konturen (AKV), das die prinzipielle Ausgangsbasis der Stimmlippenkantendetektierung bildet, unter Berücksichtigung der Spezifika der HSE-Aufnahmen modifiziert und erweitert. Schließlich wird eine Validierung der Funktionalität des entwickelten Segmentierungsverfahren durch-

geführt.

Es wird eine semi-automatische Methode zur Bestimmung der glottalen Mittellinie bei segmentierten HSE-Aufnahmen erarbeitet, um die Informationen über die Gestalt sowie Regularität der einzelnen lateralen Komponenten der Stimmlippenschwingungen und über deren Symmetrie zu gewinnen. Diese Methode ermöglicht die präzise Trennung der detektierten Stimmlippenkanten in die entsprechenden linke und rechte Seite unabhängig von Aufnahmesituation sowie anatomisch bedingten Verformungen der aufzuteilenden Konturen. Hierfür liegt der grundlegende Ansatz darin, die glottale Mittellinie direkt anhand der räumlichen Orientierung der gegebenen Konturen zu bestimmen, was durch Anwendung der PCA auf die Kantenpunkte erfolgt. Das entwickelte Verfahren wird anschließend anhand von realen HSE-Aufnahmen getestet, und bildet die Grundlage für eine zukünftige Automatisierung.

Die praktische Umsetzung der Algorithmen, die in Rahmen der vorliegenden Arbeit entstanden sind, wird anhand einer umfassenden Forschungssoftware „*Glottis Analysis Tools*“ (GAT) demonstriert.

Abstract

Voice and speech are the most important tools for human communication. A voice disorder (dysphonia) impairs the ability to communicate of the person concerned. It has consequently a negative impact on both psycho-social and occupational well-being and thus, on the quality of life of the person affected.

Phonation represents the process of generating the primary acoustic signal in the human larynx and therefore is a prerequisite for voice production and vocalization. A healthy or normal voice is characterized by almost left-right symmetric, periodic oscillations of the vocal folds during phonation. Hence, the visual inspection of the vocal folds during phonation as well as an associated analysis of the vibratory behaviour are of central importance for clinical voice diagnostics. For this purpose, digital high-speed endoscopy (HSE) provides an enormous potential to quantify vocal fold dynamics and thus, represents an effective instrument to assess the anatomy and functionality of the larynx. The evaluation of HSE recordings using image processing techniques allows the extraction of the primary information on the shape and regularity of the vocal fold vibrations. Since the functionality as well as the robustness of applied image processing methods directly influence the objectivity and reliability of obtained HSE features, the image processing plays a crucial role in the clinical applicability of objective voice diagnostics based on HSE.

In this thesis, the author presents an image processing approach with high degree of automation, which allows an objective and precise evaluation of HSE recordings of vocal fold vibrations during phonation.

In order to extract the information of the movements of the vocal folds and their regularity in form of the detected vocal fold edges as well as the glottal area waveform (GAW) derived therefrom, a fully automatic, two stage approach to segment the HSE recordings is presented. At first, an image region, namely the region of interest (ROI), is automatically determined in order to estimate the position of the vocal folds and thus of the glottis gap between on the image plane within the HSE recording. This increases the accuracy of the segmentation results and reduces the computational effort required for the subsequent vocal folds edge detection. The ROI is thereby calculated with the aid of pixel intensities, their changes over the recording time as well as their spatial distribution within the single HSE images. This is done using principal components analysis (PCA) as well as approximations by means of generalized Ricker wavelets with regard to anatomical and physiological characteristics of the vocal folds. The segmentation of the HSE recordings represents the detection of the vocal fold edges inside the evaluated ROI for the single HSE images and takes place in the second step of the developed algorithm. For this purpose, an active contour model without edges, which is the basic starting point for the suggested edge detection approach, is modified and extended taking into account the specifics of HSE recordings. Finally, the functionality of the developed segmentation method is validated based on clinical HSE recordings.

In order to obtain information on shape and regularity of dynamic lateral components of the vocal fold oscillations and on their symmetry, a semiautomatic method to determine the glottal midline in the segmented HSE recordings is developed. This method enables the precise separation of the detected vocal fold edges into the corresponding left and right sides independently of the recording situation and anatomically induced deformations. For this purpose, the basic approach is the definition of the glottal midline based on the spatial orientation of the given contour. This is done by applying PCA to the contour points. The suggested image processing algorithm is then tested using clinical HSE recordings. This approach forms the basis for future automation of the proceeding to determine the glottal midline.

The practical implementation of the developed algorithms in this thesis is demonstrated by a comprehensive research software „*Glottis Analysis Tools*“ (GAT).

Inhaltsverzeichnis

1 Einleitung	1
1.1 Der menschliche Kehlkopf: Anatomische und physiologische Grundlagen	1
1.2 Stimmstörungen: Klinisches Bild	3
1.3 Problemstellung und Stand der Technik	4
1.4 Überblick über den Aufbau der Arbeit	8
2 Segmentierung der HSE-Aufnahmen	11
2.1 Anforderungen	11
2.2 Prinzipieller Aufbau	12
2.3 <i>Region of interest</i>	14
2.3.1 Zugrundeliegendes Konzept / Allgemeine Vorgehensweise	14
2.3.2 <i>Initial-Punkt</i>	18
2.3.3 <i>ROI-Berechnung</i>	34
2.4 Detektierung der Stimmlippenkanten	82
2.4.1 Allgemeine Vorgehensweise	82
2.4.2 Abschätzung der <i>Zeitblocks</i>	88
2.4.3 Detektierung der <i>Glottalen-Region</i>	94
2.4.4 <i>Blocksegmentierung</i>	99
2.5 Forschungsstudie	113
2.5.1 Setup	113
2.5.2 Ergebnisse	114
3 Glottale Mittellinie	121
3.1 Anforderungen	121
3.2 Allgemeine Vorgehensweise	123
3.2.1 Räumliche Orientierung einer Bildregion	125
3.3 Detektierung der <i>glottalen Mittelachse</i>	128
3.4 Forschungsstudie	136
3.4.1 Setup	137
3.4.2 Ergebnisse	137
4 Zusammenfassung und Ausblick	141
A Anhang	145
A.1 Forschungssoftware „ <i>Glottis Analysis Tools</i> “	145
Nomenklatur	153

Inhaltsverzeichnis

Literaturverzeichnis	157
Abbildungsverzeichnis	166
Tabellenverzeichnis	169
Lebenslauf	171