Analyse von digitalen Hochgeschwindigkeitsvideos der Ersatzstimmgebung

Jörg Lohscheller, Maria Schuster, Michael Döllinger, Ulrich Hoppe, Ulrich Eysholdt

Abteilung für Phoniatrie und Pädaudiologie
Friedrich-Alexander Universität Erlangen-Nürnberg, 91054 Erlangen
Email: joerg.lohscheller@phoni.imed.uni-erlangen.de

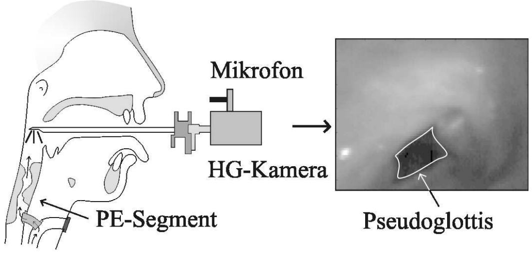
Zusammenfassung. Bei Verlust des Kehlkopfes im Rahmen einer Tumorbehandlung (Laryngektomie) kann eine Rehabilitation der Stimme ermöglicht werden, indem die Speiseröhre durch ein Einweg-Ventil aus Silikon (Stimmventilprothese) mit der Luftröhre verbunden wird. Mithilfe der Stimmventilprothese können beim Ausatmen Schleimhäute in der Speiseröhre durch den Luftstrom in Schwingungen versetzt und so als Quelle einer Ersatzstimme genutzt werden. Die Qualität der Ersatzstimme ist wesentlich durch die Schwingungscharakteristik der Schleimhäute bestimmt. Zur Visualisierung des Stimmgebungsprozesses wurden simultan digitale Hochgeschwindigkeitsvideos der Schleimhautbewegungen und das akustische Signal während der Phonation aufgezeichnet. Zur Bestimmung der für den Stimmgebungsprozess relevanten Bildanteile wurde eine Korrelationsanalyse von Bildsequenzen und den dazugehörigen akustischem Signalen vorgenommen.

1 Einleitung

Die operative Entfernung des Kehlkopfes infolge von Kehlkopfkrebs erfordert die Trennung des Luft- und Speiseweges. Die weitere Atmung des Patienten erfolgt über eine hergestellte Öffnung der Luftröhre (Trachea) am Hals, indem der Stumpf der Luftröhre in die Halshaut eingenäht wird (Tracheostoma). Als weitere Folge der Operation geht die Funktion des Kehlkopfes als Stimmgenerator verloren. Eine Möglichkeit zur Ausbildung einer Ersatzstimme besteht darin, die voneinander getrennte Luft- und Speiseröhre (Ösophagus) durch ein Einwegventil aus Silikon erneut miteinander zu verbinden [1]. Während des Ausatmens ermöglicht der manuelle Verschluss des Tracheostomas die Umleitung des Luftstromes durch die Stimmventilprothese in die Speiseröhre. Die sich am oberen Ende der Speiseröhre befindlichen Schleimhäute ('pharyngeal-esophageal Segment', PE-Segment) lassen sich durch den Luftstrom in Schwingung versetzen. Die Qualität dieser Ersatzstimme ist im wesentlichen durch die Schwingungscharakteristik des PE-Segmentes bestimmt [2].

Zur Visualisierung des Ersatzstimmgebungsprozesses bei laryngektomierten Patienten wurden digitale Hochgeschwindigkeitsvideos (HG-Videoaufnahmen)

Abb. 1. Aufnahme der Bewegungen des PE-Segmentes während der Phonation mit einer Hochgeschwindigkeitskamera.



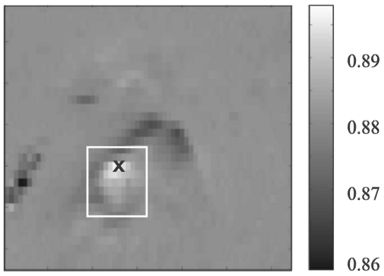
der Schleimhautbewegungen aufgezeichnet. Der Schwingungsverlauf der Schleimhäute kann dabei mit ausreichender Zeitauflösung dargestellt werden. Auf den Hochgeschwindigkeitsvideos lässt sich der Schwingungsvorgang der Schleimhäute durch die zeitabhängige Öffnungsfläche des PE-Segmentes identifizieren. In Anlehnung an die normale Stimmgebung wird die sich ausbildende Öffnungsfläche als Pseudoglottis bezeichnet. Die zeitliche Veränderung der Pseudoglottis wird als verantwortliche Größe für das Stimmsignal angesehen.

An einer Gruppe von sieben Laryngektomierten mit einer Stimmventilprothese (Typ Provox) wurden endoskopische Hochgeschwindigkeitsvideos (3704 Bilder/s, 128x64 Bildpunkte) des PE-Segmentes während der Stimmgebung unter klinischen Bedingungen aufgenommen (Abb.1). Gleichzeitig wurden die akustischen Stimmsignale aufgezeichnet (Mikrophon: Brüel & Kjaer 4129, Abtastrate: 44.4 kHz, 16 bit Auflösung). Bei geöffneter Pseudoglottis lassen sich aufgrund der Kameraposition auch tieferliegende Abschnitte des Ösophagus aufnehmen. So konnten auf den HG-Aufnahmen von kaudal nach kranial verlaufende Oberflächenwellen auf dem PE-Segment beobachtet werden. Diese zusätzlichen zeitlichen Strukturveränderungen des Ösophagus erschweren die Segmentierung der Pseudoglottis. Bedingt durch die Untersuchungssituation treten während der HG-Aufnahme zudem Ungenauigkeiten bei der Fokussierung als auch Verdeckung durch starke Schleimbildung auf.

2 Methode

Die Extraktion der Pseudoglottiskontour ist erschwert durch die geringe Bildqualität der Aufnahmen und stellt daher hohe Ansprüche an die verwendete Bildverarbeitung. Es ist daher erforderlich, die Bildbereiche, die für die Stimmproduktion relevant sind (region of interest, ROI), zu identifizieren. Zur Bestimmung dieser Bildbereiche wird die Kreuzkorrelation der zeitlichen Änderung der Grauwerte eines Pixels und des akustischen Signals berechnet. Daraus lassen sich die Bildbereiche visualisieren, deren zeitliche Grauwertvariation in einem engen Zusammenhang mit dem Stimmsignal stehen. Der Bildpunkt maximaler

Abb. 2. Kreuzkorrelationsanalyse aus der zeitlichen Änderung der Grauwerte mit dem akustischen Signal. Markiert sind die Region of Interest (ROI, weißes Rechteck) und der berechnete Pseudoglottisursprung (Kreuz).



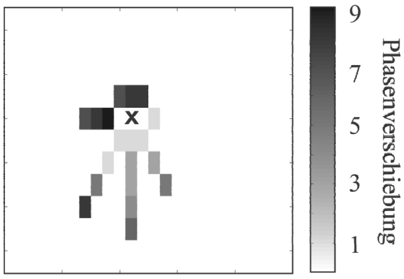
Kreuzkorrelation wird für die betrachtete Bilderfolge als Ursprung der Pseudoglottisöffnung definiert. Mittels eines Schwellwertverfahrens wird der Bildbereich segmentiert, dessen Schwingungscharakteristik mit dem des akustischen Signals am besten übereinstimmt (Abb. 2). Lassen sich mit diesem Verfahren Verfahren keine Übereinstimmungen mit dem akustischen Signal finden, so wird dies als Kriterium für eine zu geringe Bildqualität gesehen werden.

Des weiteren kann die zeitliche Änderung der Pseudoglottisgeometrie abgeschätzt werden, indem die Kreuzkorrelation der zeitlichen Änderung der Grauwerte des berechneten Ursprungpunktes und der zeitlichen Änderung der Grauwerte eines jeden Pixels bestimmt wird. Diese Vorgehensweise begründet sich wie folgt: Die Pseudoglottisöffnung breitet sich beginnend vom Ursprung nach außen aus. Die Analyse der Verschiebung des Maximalwertes der jeweilige zugehörigen Kreuzkorrelation von Bildpixeln innerhalb der ROI und dem Ursprung der Pseudoglottis liefert ein Maß für die Phasenverschiebung und damit der Richtung und Geschwindigkeit der Ausbreitung der Pseudoglottis. Die Phasenverschiebung der zeitlichen Grauwertverläufe von Bildpixeln entlang von vier Geraden ist exemplarisch in Abbildung 3 dargestellt.

3 Ergebnisse und Diskussion

In übereinstimmung mit der männlichen Normalstimme bestehen die Fourierspektren der akustischen Signale der Ersatzstimme hauptsächlich aus einer Komponente zwischen 60 Hz und 185 Hz (Grundfrequenz) und weiteren Maxima bei ganzzahligen Vielfachen (Harmonische). Im Unterschied zur normalen Stimme lassen sich Verbreiterungen der lokalen Maxima im Spektrum und ein erhöhter Rauschanteil identifizieren. Das vorgestellte Analyseverfahren konnte bei hinreichender Bildqualität die für das akustische Signal relevanten Bildanteile (ROI) identifizieren. Die Darstellung der Korrelationskoeffizienten und der Phasenverschiebungen erlauben Aussagen über die Veränderung der Geometrie der Pseudoglottis während des Schwingungsprozesses. Bei allen Probanden stimmen die Grundfrequenzen der Stimmsignale mit den aus den Grauwertkurven

Abb. 3. Hauptausdehnungsrichtungen und Phasenverschiebung in der ROI entlang von Geraden durch den Pseudoglottisursprung (Kreuz).



der Bildpunkte innerhalb der ROI bestimmten Grundfrequenzen überein. Unter Berücksichtigung des Abstandes der Pixel vom ermittelten Ursprung und der Phasenverschiebung zum Ursprung der Pseudoglottis kann auf die Ausbreitungsgeschwindigkeit der Pseudoglottis und damit auf elastische Eigenschaften des ösophagealen Gewebes geschlossen werden.

4 Schlussfolgerung

Obwohl die endoskopischen Hochgeschwindigkeitssequenzen eine zweidimensionale Projektion eines dreidimensionalen Bewegungsvorganges darstellen, werden in den Aufnahmen die für die Stimme wichtigen Schwingungsvorgänge ausreichend erfasst. Der in dieser Arbeit vorgestellten Visualisierungsansatz besteht aus einer Identifikation und Analyse der für das akustische Signal relevanten Bildbereiche. Mögliche Einsatzgebiete des Verfahrens bestehen in der Beurteilung der Bildqualität endoskopischer Aufnahmen und in der Vorverarbeitung für Segmentierungsverfahren zur Bestimmung der zeitabhängigen Pseudglottisgeometrie

Literatur

1. C. van As, M. Tigges, F. Hilgers, and U. Eysholdt, “Oesophageal Vibration in Voice Rehabilitation after Laryngectomy,” in *Advances in Quantitative Laryngoscopy* (T. Wittenberg, P. Mergell, M. Tigges, and U. Eysholdt, eds.), (Göttingen), pp. 95–102, Verlag Abteilung Phoniatrie, 1997.
2. M. de Vries, *A new voice for the voiceless, Design and in-vitro testing of a voice-producing element.* 2000.