

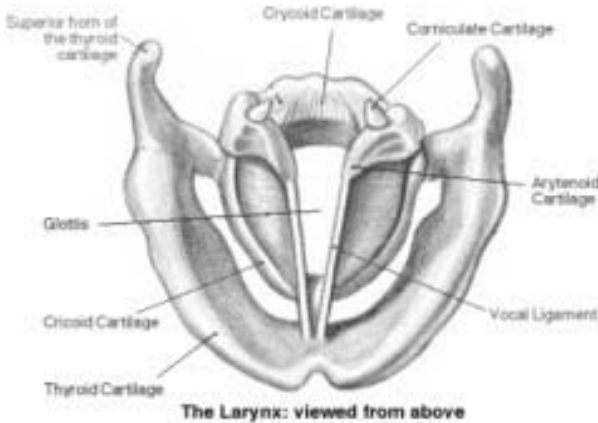
Vollautomatische Glottisdetektion bei Hochgeschwindigkeitsaufnahmen

Sven Anderson, Dirk Michaelis, H. W. Strube

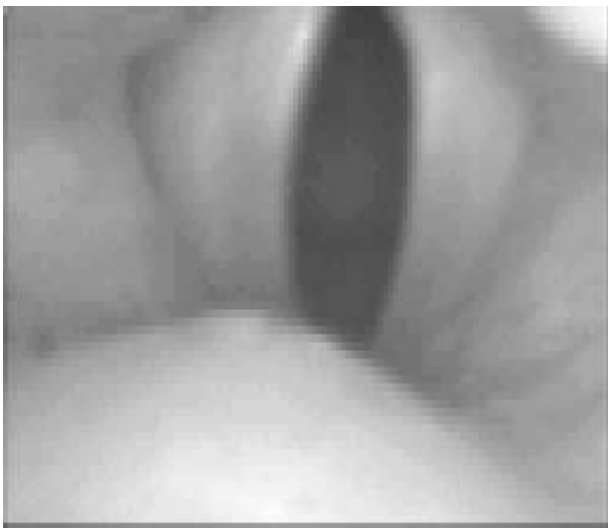
Drittes Physikalisches Institut /
 Forschung Stimme und Sinnesentwicklung - Abteilung für Phoniatrie und Pädaudiologie
 Universität Göttingen

Einleitung

Unter dem aus dem griechischen stammenden Wort „Glottis“ versteht man die Stimmritze im Kehlkopf, also das durch die beiden Stimmlippen begrenzte Gebiet. Neben ihrer Hauptfunktion – dem Verschluss der Luftröhre beim Schlucken – dient sie der Laut- und Spracherzeugung, indem sie zur Schwingung angeregt wird.



Die Glottis wird nun während dieser Schwingung durch ein so genanntes Laryngoskop mit einer Hochgeschwindigkeitskamera aufgenommen. Hierbei entsteht eine monochrome Filmaufnahme in 256 Graustufen, bei einer maximalen Auflösung von 256^2 Pixeln und einer maximalen Bildfrequenz von 4000 Hz. Eine beispielhafte Momentaufnahme sieht man hier:



Mit dem späteren Ziel die Korrelation zwischen Glottisfläche und dem Rauschanteil der Stimme zu untersuchen, und auch unter dem Aspekt ein neues Diagnoseinstrument für Ärzte zu ermöglichen, gilt es nun, in einem solchen Film das Gebiet der Glottis vollautomatisch zu finden und zur späteren Weiterverarbeitung auszugeben. Hierbei hat man leider mit einigen Problemen zu kämpfen. Die Ausleuchtung, der Bildausschnitt und auch die Anatomie an sich – vor allem bei pathologischen Fällen – des aufgenommenen Gebietes unterliegen starken Schwankungen. Des weiteren entstehen durch die notwendige starke Beleuchtung sehr helle Reflexionen auf dem feuchten Gewebe. Das Verfahren sollte also damit möglichst gut zurecht kommen.

Zeitpunkt der geöffneten Glottis

Zunächst ist es hilfreich zu wissen, wann die Glottis möglichst weit geöffnet ist, um später einen Ausgangspunkt zu haben. Da die Glottis bei den Aufnahmen immer als dunklere Fläche als ihre Umgebung erscheint, kann man annehmen, dass bei maximal geöffneter Glottis die Gesamthelligkeit des Bildes minimal wird:

$$b(t) = \sum_{x,y} p(x,y,t).$$

Hierbei ist $p(x,y,t)$ der Helligkeitswert des Pixels (x,y) zum Zeitpunkt t .

Um die Minima zu finden, wird $b(t)$ mittelwertgefiltert und dann ein Schwellwert $s(t)$ bestimmt:

$$s(t) = \bar{b}_W(t) - \sqrt{\frac{\sum_{\text{alle } T} (b(T) - \bar{b}_W(T))^2}{N}}.$$

$\bar{b}_W(t)$ ist hierbei der Mittelwert eines gleitenden Fensters der Breite $2W + 1$ um t herum, und nicht, wie bei der Standardabweichung, der Mittelwert des ganzen Signals. Hierdurch wird ein Einfluss von weitläufigeren Helligkeitsänderungen, z.B. durch veränderte Lichtverhältnisse, vermieden. N ist die Anzahl der Einzelbilder.

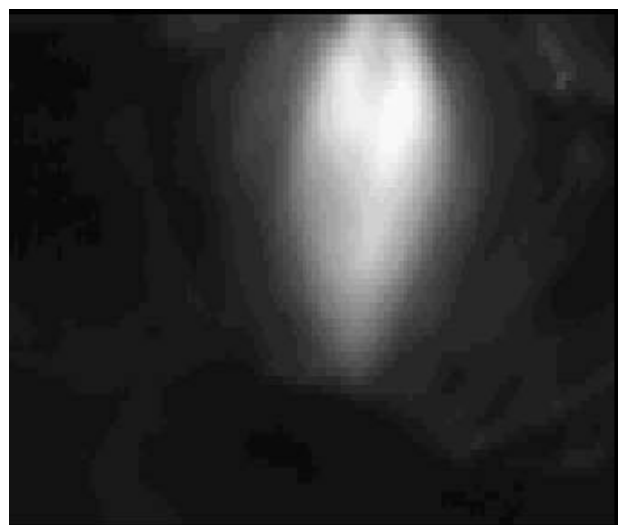
Zu Zeitpunkten, bei denen $b(t) < s(t)$ ist, ist die Glottis geöffnet. Ist $b(t)$ zusätzlich ein lokales Minimum, ist die Glottis in dieser Periode maximal geöffnet.

Aktivität

Ebenfalls hilfreich ist es zu wissen, an welchen Stellen des Bildes sich die Helligkeit im Zeitverlauf wie stark ändert; mit der Erwartung, dass im Bereich der schwingenden Glottis eine hohe solche „Aktivität“ vorhanden ist:

$$A(x,y) = \sqrt{\frac{\sum_T (p(x,y,T) - \bar{p}_W(x,y,T))^2}{N}}.$$

Auch hier ist \bar{p}_W wieder der Mittelwert eines gleitenden Fensters. Die resultierenden Aktivitätsbilder bestätigen die Annahme. Hier ein Beispiel, bei dem man deutlich das helle (aktive) Gebiet im Bereich der Glottis sieht:



Gewichtetes Histogramm

Man kann das beobachtete Gebiet grob in zwei Bereiche teilen: die Glottis (dunkel) und das Gewebe (hell). Bildet man nun ein Histogramm über die 256 Graustufen des Bildes, müssten sich zwei den beiden Gebieten entsprechende Häufungen ausbilden, deren dazwischen befindliches Minimum genau dem Grauwert der Glottiskante entspricht. Leider ist dem aber oft nicht so, da sich z.B. auch im Bereich des Gewebes oft sehr dunkle Stellen befinden, durch Schatten oder schlechte Ausleuchtung.

Um dies zu kompensieren, werden die Punkte zuvor *gewichtet*:

1. Es erhalten nur die Punkte aus Bildern mit *geöffneter Glottis* ein Gewicht, denn nur dann lässt sich das Bild klar in hell und dunkel unterteilen.
2. Punkte in *homogenen Gebieten* bekommen ein stärkeres Gewicht:

$$w_1(x, y, t) = \frac{1}{1 + \sum_{\text{Ser-Umgebung}} (p_{\Delta} - p(x, y, t))^2},$$

wobei p_{Δ} einer der 8 Nachbarpunkte von $p(x, y, t)$ ist. Dadurch wiegen Punkte an Kanten weniger, wodurch das gesuchte Minimum stärker ausgebildet wird.

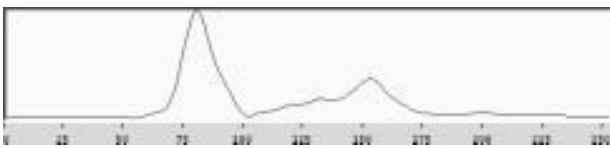
3. Punkte in *aktiven Gebieten* wiegen mehr:

$$w_2(x, y) = A(x, y).$$

Dadurch werden dunkle Gebiete am Rand ausmaskiert, die ansonsten das gesuchte Minimum überdecken.

Das Gewicht eines Punktes beim Bilden des Grauwert-Histogramms ist also $w_1 \cdot w_2$.

Anschließend wird das Histogramm tiefpassgefiltert. Hier ein Beispiel eines solchen Histogramms:



Zur Bestimmung des Minimums werden zunächst die beiden ersten lokalen Maxima bestimmt, wobei hier „lokal“ heisst, innerhalb einer Mindestumgebung, die die jeweils andere nicht überlappt. Danach wird das Minimum zwischen diesen beiden Maxima bestimmt. Der so ermittelte Grauwert ist der Schwellwert S , der dem des Glottisrandes entspricht, also die Glottis begrenzt.

3D-Flood-Fill

Manchmal ist das Gebiet der Glottis (z.B. durch Schleimfäden) zerteilt. Würde man nun auf Einzelbildbasis ein Flächenwachstumsverfahren anwenden, würde bei diesen Bildern nur ein Teil der Glottis erschlossen werden. Verwendet man stattdessen ein 3-dimensionales Volumenwachstumsverfahren, bei dem die Zeit die dritte Dimension darstellt, sind solche getrennten Gebiete teil desselben Körpers, werden also konsistent erschlossen.

Dazu wird folgender *Flood-Fill-Algorithmus* angewandt:

1. Samenpunkt in FIFO-Speicher ablegen und markieren.
2. Punkt aus FIFO holen.
3. Alle Nachbarpunkte in der 6er- oder 26er-Umgebung im FIFO ablegen und markieren, die einem Kohärenzkriterium genügen und noch nicht markiert sind.
4. Optional: \exists unmarkierte Nachbarpunkte
 → Punkt als Körper-Randpunkt markieren.
 \exists unmarkierte Nachbarpunkte mit selbem t
 → Markierung als Einzelbild-Randpunkt.

5. Weiter bei 2., solange FIFO nicht leer.

6. Gesamtes zusammenhängendes Gebiet ist markiert.

Im Vergleich zu einer rekursiven Lösung „merkt“ sich dieser Algorithmus lediglich die Oberfläche eines Körpers, der sich so weit wie möglich ausdehnt, was bei großen Datenmengen eine erhebliche Speicherauslastung ausmacht.

Markierung der Glottis

Wenn man also den Film als dreidimensionales Feld betrachtet, bildet sich die schwingende Glottis dort als Blasen oder pulsierender Schlauch ab, je nachdem ob die Glottis sich während einer Periode vollständig verschließt oder nicht. Diese Objekte gilt es nun zu finden und zu erschließen.

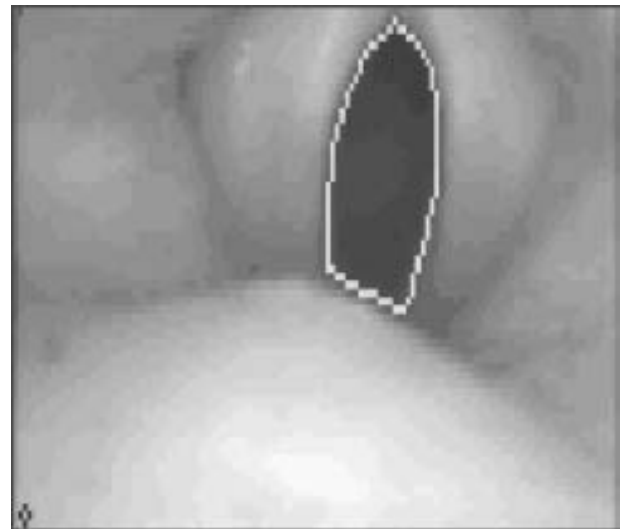
Reflexionen aus der Luftröhre bilden gelegentlich helle Stellen innerhalb der Glottis, die heller sind als der Schwellwert S . Würde man nun den „Glottis-Körper“ direkt anhand von S ausfüllen, würden diese Stellen unerwünschte Lücken im Körper bilden.

Daher wird zunächst ein Art Hohlform erzeugt, indem vom Bildrand beginnend alle Punkte gefüllt werden, deren Helligkeit $\geq S$ ist. Nun werden die für den Flood-Fill notwendigen Samenpunkte wie folgt bestimmt:

1. Zu jedem Zeitpunkt t wird maximal ein Samenpunkt gesetzt.
2. Wenn zur Zeit t die Glottis geöffnet ist, wird ein Samenpunkt innerhalb des Hohlraums dort gesetzt, wo $(255 - p(x, y, t)) \cdot A(x, y)$ maximal ist. So wird verhindert, dass der Samenpunkt in Hohlraumblasen gesetzt wird, die durch andere dunkle Stellen und nicht durch die Glottis erzeugt wurden.
3. Zu sonstigen Zeiten t wird der Samenpunkt von $t - 1$ übernommen, sofern diese immer noch im Hohlraum liegt.

Nun werden anhand aller Samenpunkte mit demselben Flood-Fill-Algorithmus die gesuchten Körper bis zur Hohlraumgrenze ausgefüllt. Dabei werden sogleich die Randpunkte des Körpers und die Randpunkte der Glottisflächen der Einzelbilder gesondert markiert.

Zur Illustration hier ein Bild mit eingezeichnetem Einzelbild-Rand:



Erweiterungen

Um eine noch genauere Lokalisation des Randes zu erreichen, kann man dieses Verfahren nun zur Initialisierung von so genannten *Aktiven Konturen (Adaptive Snakes)* verwenden, bei denen der Gradient der Bildpunkte und physikalische Eigenschaften, die man den Glottiskanten zuweist, zum Tragen kommen. Solche Verfahren benötigen allerdings erheblich mehr Rechenaufwand.