

Frühdiagnostik von Larynxkarzinomen mittels phonetischer Analyse

Yasmina Höher

Heinrich-Heine-Schule, 24226 Heikendorf, E-Mail: yh_projects@web.de

Abstract

Jährlich erkranken mehrere Tausend neue Personen allein in Deutschland an einem Larynxkarzinom, d.h. an Kehlkopfkrebs – Tendenz steigend. Das Ziel dieses Beitrags ist es herauszufinden, ob die Stimmveränderung bei Kehlkopfkrebs auf Basis von wenigen Parametern per Sprachanalyse detektierbar und in einem frühen Stadium erkennbar ist. Da nach bestem Wissen keine offene Sprachdatenbank einschlägig erkrankter Patienten existiert, wurde zu einem Trick gegriffen: Es wurden Sprachproben von Personen des öffentlichen Lebens analysiert, die bekanntermaßen an Kehlkopfkrebs erkrankt sind. Dadurch konnte die zeitliche Entwicklung über einen 10-15-jährigen Zeitraum ausgewertet werden. Drei prosodische Merkmale erwiesen sich als aussagefähig: Jitter, Shimmer und Harmonicity. Zwecks Frühdiagnose wird vorgeschlagen, dass jede Person bei Arztbesuchen oder auf einem Internetportal turnusgemäß eine kurze Sprachprobe hinterlegt.

1. Einleitung

Das Gebiet der Sprachanalyse ist sowohl in der Forschung als auch in der Anwendung sehr breit gefächert. Zu den Anwendungen gehört beispielsweise die Optimierung des Kundenservices am Telefon durch Stimmerkennung, die sprachgesteuerte Unterstützung von technischen Geräten, die Komprimierung von Sprachsignalen, die Sprachrekonstruktion nach operativen Eingriffen, sowie die Analyse von krankhaften Stimmen als diagnostische Methode in der Medizintechnik. Letzteres findet beispielsweise bei der Prognose von Morbus Parkinson Anwendung [1]. Wenig erforscht ist die sprachgestützte Diagnose von Kehlkopfkrebs [2, 3]. Dies motiviert das im Abstract definierte Ziel.

Etwa 4000 neue Personen erkranken jährlich in Deutschland an Kehlkopfkrebs, davon sind 87,5 Prozent Männer [4]. Hauptursache sind Nikotin- und Alkoholkonsum sowie zunehmend auch Papillomaviren, da hierdurch die Entstehung maligner Zellen aktiv gefördert wird. Weitere Risikofaktoren bilden Asbest und Schwermetalle wie Nickel oder Chrom. Zudem besteht die Hypothese, dass erbliche Veranlagungen ein Erkrankungsrisiko steigern können. Zu differenzieren sind drei Arten von Kehlkopfkrebs, die sich jeweils in ihrer Lage unterscheiden:

- Supraglottisches Karzinom (oberhalb der Glottis)
- Glottisches Karzinom (auf Höhe der Glottis)
- Subglottisches Karzinom (unterhalb der Glottis)

Das Glottische Karzinom ist mit 65 Prozent aller Fälle die am häufigsten auftretende Art. Da der Tumor direkt an den Stimmlippen liegt, sind begleitende Symptome meist anhaltende Heiserkeit, Halskratzen und vermehrtes Räuspern. In fortgeschrittenem Stadium kommen oftmals

Atembeschwerden hinzu, da die Luftwege durch das wachsende Karzinom verengt werden. Supraglottische Karzinome treten in 30 Prozent aller Fälle auf und werden meist durch ein Fremdkörpergefühl im Hals bemerkt. Erst in einer späteren Entwicklungsphase, wenn sich Metastasen an den Stimmlippen gebildet haben, tritt auch eine Heiserkeit der Stimme auf. Ähnliches gilt für das Subglottische Karzinom, welches nur in den seltensten Fällen auftritt.

2. Phonation im menschlichen Körper

Um die Auswirkungen eines Kehlkopftumors auf die Stimme zu analysieren, muss zunächst verstanden werden, wie ein hörbarer Laut im menschlichen Körper gebildet wird. Der Prozess der Lautbildung nennt sich Phonation und beruht auf drei maßgeblichen Faktoren [5]:

1. Dem Luftstrom, welcher beim Ausatmen entsteht,
2. den Stimmlippen, welche sich im Kehlkopf befinden,
3. und dem sog. Ansatzrohr, welches aus Rachen-, Mund- und Nasenraum besteht und für die Klangformung ausschlaggebend ist.

In der Respirationslage, d.h. in der Atemphase, liegen die Stimmlippen weit auseinander und berühren sich nicht. Die Glottis, d.h. der Bereich zwischen den Stimmlippen, ist weit geöffnet. So kann die Atemluft problemlos nacheinander in beide Richtungen entweichen. Spricht man einen Laut, verschließen sich zunächst die Stimmlippen, so dass sie sich beinahe auf ganzer Länge berühren, man spricht von der sog. Phonationslage. Der Atemstrom staut sich und es entsteht ein subglottaler Druck. Erreicht dieser Druck einen bestimmten Schwellwert, werden die Stimmlippen auseinander gedrückt und die Atemluft entweicht mit hoher Geschwindigkeit in den Rachen. Dabei entsteht eine Sogwirkung, die durch den Bernoulli-Effekt zu erklären ist. Luft bewegt sich in engeren Räumen schneller als in erweiterten, weswegen aufgrund einer Abnahme der Dichte an Molekülen in dem engeren Raum ein Unterdruck entsteht.

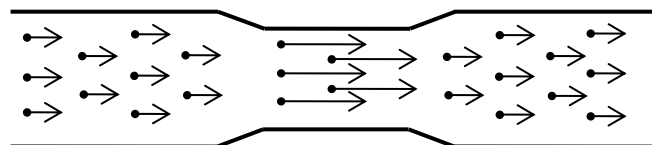


Abb. 1 Bernoulli-Effekt

Der Bernoulli-Effekt ist auch auf die Atemluft in der Luftröhre anwendbar. Die Geschwindigkeit der Ausatemluft nimmt in der Glottis, deren Anatomie schmaler ist als die der Luftröhre, zu. Durch die Beschleunigung der Luftmoleküle entsteht in der Glottis ein Unterdruck, der dazu führt, dass sich die elastischen

Stimmklappen zusammen ziehen. Die Glottis wird verschlossen, es bildet sich erneut ein subglottaler Druck und der gesamte Vorgang, auch Phonationszyklus genannt, wiederholt sich. Es entsteht eine periodische Vibration.

Die Lautstärke des durch die Vibration hervorgerufenen Lautes kann durch die Kraft des Ausatemstroms variiert werden. Die Frequenz dieses Tones hängt von der Spannung der Stimmklappen ab. Je höher die Spannung, desto schneller die Vibration und desto höher erklingt der erzeugte Ton [5]. Die Spannung der Stimmklappen wird vom Kehlkopf reguliert.

Im Ansatzrohr wird der zuvor erzeugte Ton weiterverarbeitet. Durch bestimmte Bewegungen des Gaumens, der Zunge und des Mundes werden die verschiedenen Töne zu Lauten geformt.

Die Laute, welche wie beschrieben mit bewegten Stimmbändern produziert werden, werden als stimmhaft bezeichnet. Dazu zählen alle Vokale (z.B. a; e; i) sowie viele Konsonanten (z.B. b; d; g). Stimmlose Laute werden bei geöffneten Stimmklappen gebildet (z.B. h; s; f), die Stimmklappen bewegen sich nicht und erzeugen folglich auch keine Vibration, alleine der Ausatemstrom sowie das Ansatzrohr sind für die Entstehung eines stimmlosen Lautes grundlegend [6].

3. Methode

3.1 Analysetool und Parameterauswahl

Zur Sprachsignalanalyse wurde Praat [7], ein freies Softwarepaket für phonetische Analysen auf Signalebasis verwendet. Dieses bietet eine Vielzahl an Optionen. Ich arbeitete hauptsächlich mit der Funktion Voice Report. Im Rahmen der eigenen Untersuchungen hat sich herausgestellt, dass folgende drei prosodische Merkmale besonders aussagefähig sind:

(1) Jitter

Bei gesunder Phonation sind stimmhafte Sprachsignale beinahe periodisch. Bei krankhaften Stimmen kann es aber zu einer Perturbation, d.h. einer Störung, der Grundfrequenzen kommen. Die Rauigkeit einer Stimme wird anhand des Parameters Jitter gemessen. Je höher die Angabe in Prozent, desto höher ist die Perturbation.

(2) Shimmer

Der Shimmer ist ähnlich wie der Jitter definiert. Nur wird hierbei die Störung der Amplitude gemessen. Die Amplitude bestimmt die Lautstärke des Tones. Bei einer Perturbation klingt die Stimme also zittrig. Auch hier gilt: Je höher die Angabe in Prozent, desto höher ist die Perturbation.

(3) Harmonicity

Für die Messung der Harmonizität werden Signale im Frequenzbereich korreliert. Je höher die Ähnlichkeit ist, desto näher liegt der Wert bei eins. Da Rauschen auch im Frequenzbereich unkorreliert ist, gibt dieses Merkmal Aufschluss über den Rauschanteil in der Stimme.

Manchmal liefert auch die Grundfrequenz („Pitch“) Hinweise für eine Veränderung am Sprachapparat. Zusätzlich könnte betrachtet werden, wie oft die Phonation abgebrochen bzw. unterbrochen wird. Dabei ist wichtig zu beachten, dass die Sprechpausen, welche zwischen den Wörtern eingelegt werden, außen vorgelesen werden.

Um die einzelnen Praat-Befehle besser kennenzulernen und ein Gefühl für das Aussehen einer gesunden Stimme zu entwickeln, untersuchte ich zunächst das Sprachsignal gesunder Menschen. Dafür verglich ich die Stimmen von vier gesunden Sprechern, die alle das Wort „Hallo“ phonierten. Aus den geplotteten Signalen markierte ich zunächst den ersten Vokal, in diesem Fall das „a“ und analysierte es in Bezug auf die drei genannten und weitere Parameter. Diese Vorgehensweise wurde mir von Spezialisten der Audiologie des Universitätsklinikums Schleswig-Holstein (UKSH Kiel) bestätigt. Ich merkte sofort, dass es wichtig ist, Hintergrundgeräusche so gut es geht zu vermeiden, da schon kleinste Geräusche das Ergebnis beeinflussen. Für die weitere Arbeit musste ich also möglichst „klinisch reine“ Stimmproben verwenden, damit eine zuverlässige Aussage über die Stimmveränderung bei einem Larynx tumor getroffen werden kann.

3.2 Suche von Kehlkopfkrebspatienten

Die Suche nach Stimmproben von Kehlkopfkrebspatienten gestaltete sich als äußerst schwierig und zeitaufwändig. Es gibt zwar öffentlich zugängliche Sprachkorpora, allerdings handelt es sich bei dem Inhalt dieser Datenbanken nicht um langzeitige Stimmproben von Probanden, die zwischenzeitlich an Kehlkopfkrebs erkrankt sind. Deswegen recherchierte ich nach prominenten Personen, von denen bekannt ist, dass sie an Kehlkopfkrebs erkrankt sind. Im deutschsprachigen Raum sind mir keine Namen bekannt, allerdings im englischsprachigen Raum vor allem in den USA. Bei der Suche habe ich vorzugsweise Schauspieler berücksichtigt, bei denen die Stimme aus der (gesunden) Vergangenheit zugänglich ist, sodass eine Entwicklung der Sprache analysiert werden kann. Wenn nur die krankhafte Stimme mit einer x-beliebigen gesunden Stimme verglichen wird, kann keine absolut sichere Aussage getroffen werden. Wird die krankhafte Stimme hingegen mit der gesunden desselben Sprechers aus der Vergangenheit verglichen, können genauere Rückschlüsse auf eine Veränderung gezogen werden.

Ein weiteres Problem war die Suche nach „klinisch reinen“ Stimmproben, d.h. Hintergrundgeräusche sollten so gering wie möglich sein. Theoretisch könnte man an ein Bandpassfilter denken, das Frequenzen außerhalb des Sprachfrequenzbereiches wegfiltert. Damit ist allerdings nicht gesichert, dass sämtliche Störgeräusche innerhalb dieser Bandbreite weggefiltert werden oder noch schlimmer, dass jene Signalanteile weggefiltert werden, welche ein Indiz für das Larynxkarzinom sind. Dies würde zu einer Verfälschung des Nutzsignals führen. Folglich musste ich mich bei der Suche darauf konzentrieren, dass möglichst kein Geräusch im Hintergrund zu hören ist. Dies

schließt Stimmproben beispielsweise von Sängern aus, da hierbei oftmals im Hintergrund begleitende Musik läuft.

3.3 Stimmproben

Für die Stimmanalyse konzentrierte ich mich auf oft benutzte Wörter, welche stimmhaft sind, wie „you“ und „I“. Ich nahm zehn unkorrelierte Stimmproben pro Film (Jahr), bestimmte und dokumentierte die genannten Parameter und wertete sie in Boxplots aus. Die roten Linien präsentieren den Median, die Whisker die Abweichungen.

4. Analyseergebnisse

Exemplarisch werden im Folgenden Analyseergebnisse für einen bekannten US-Schauspieler gezeigt. Aus ethisch/rechtlichen Gründen wird auf eine Namensnennung verzichtet. Weitere Untersuchungen bei weiblichen und männlichen Patienten wiesen ähnliche Tendenzen auf.

Die Abbildungen 2 und 3 zeigen Boxplots einer Jitter- und Shimmeranalyse. Die Jitter- und Shimmerwerte steigen im untersuchten Zeitraum kontinuierlich an. Dies ist darauf zurückzuführen, dass die Periodizität und die Lautstärke der Signale durch den Kehlkopftumor gestört werden. Für das menschliche Gehör erklingen solche Laute rau.

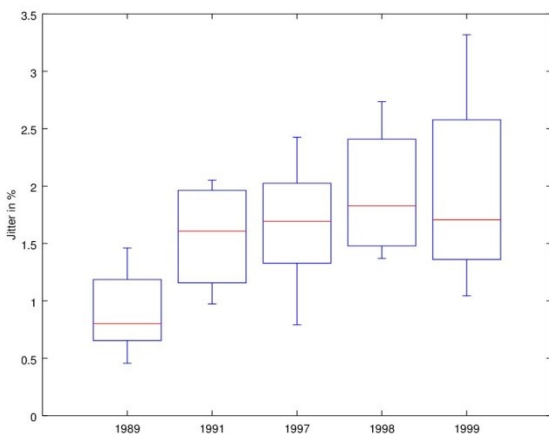


Abb. 2 Stimmveränderung gemessen am Parameter Jitter

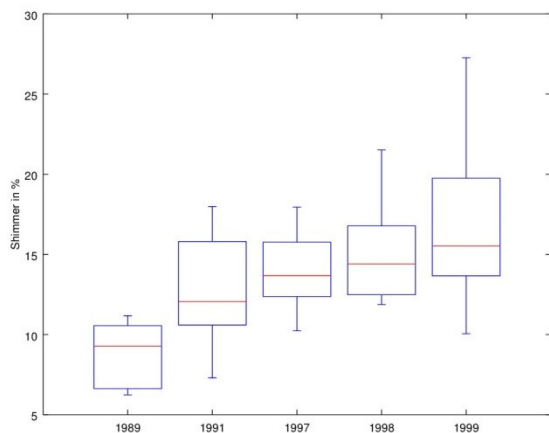


Abb. 3 Stimmveränderung gemessen am Parameter Shimmer

Der Tumor an der Glottis verursacht ebenfalls eine Perturbation der Harmonizität der Stimme, siehe Abbildung 4. Wenn die Glottis aufgrund einer Schwellung der Stimmlippen, die durch den Tumor bedingt ist, während der Phonationsstellung nicht mehr richtig geschlossen werden kann, gelangt mehr Ausatemluft durch als normal [8]. Es entstehen zusätzliche Geräusche und der daraus resultierende Klang der Stimme ist behaucht [9].

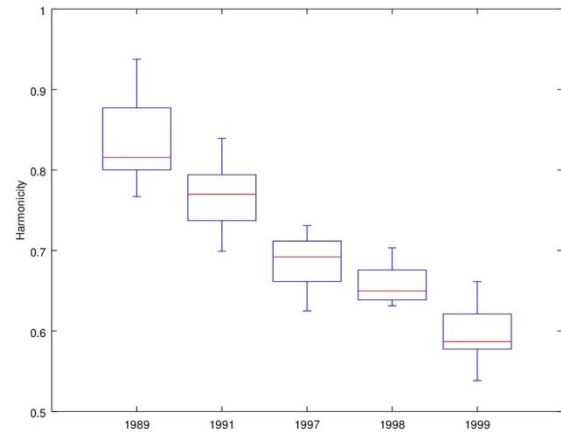


Abb. 4 Stimmveränderung gemessen am Parameter Harmonicity

5. Erkenntnisgewinn

Die Auswertungen liefern wichtige Hinweise in Bezug auf die Veränderungen am Sprachapparat, die durch einen Kehlkopftumor hervorgerufen werden. Am Sichersten kann eine Diagnose gestellt werden, wenn ein Laut über einen Zeitraum von zwei bis vier Sekunden angehalten wird. Am besten eignet sich hierfür ein Vokal, wie z.B. das „a“, da hierbei die Benutzung der Stimmlippen garantiert ist.

Die Messwerte habe ich manuell ausgewertet und bewertet. Zukünftig könnte dieser Vorgang computer-gesteuert ablaufen.

6. Kritische Auseinandersetzung

Bei den Stimmproben ist es sehr wichtig zu beachten, dass das Hintergrundrauschen so gering wie möglich ist. Aus diesem Grund sind öffentlich zugängliche Quellen für verlässliche Aussagen oft nicht nutzbar.

Aufgrund des geringen Umfangs der mir zur Verfügung stehenden Sprachproben darf die statistische Aussagekraft zum jetzigen Zeitpunkt nicht überbewertet werden.

Die Sprachproben der Schauspieler habe ich YouTube-Videos entnommen. In YouTube-Videos ist keine getrennte Sprachspur vorhanden. Deshalb habe ich Sequenzen aus YouTube-Videos über mein Smartphone abgespielt und mit einem Mikrophon aufgenommen, welches ich an die Soundkarte meines Laptops angeschlossen habe. Es ist wichtig während der Aufnahme keine elektronischen Geräte in der Umgebung zu laden, da dies Auswirkung auf das Sprachsignal haben kann. Die Geräte wurden im Batteriebetrieb verwendet um 50 Hz Netzbrumm ausschließen zu können.

7. Zusammenfassung und Ausblick

Im Rahmen dieses Beitrags wurde die Möglichkeit untersucht, Kehlkopfkrebs mittels einer Sprachanalyse zu diagnostizieren. Hierfür wurden Stimmproben von Schauspielern in Bezug auf relevante prosodische Merkmale analysiert und sowohl mit ihren gesunden Stimmen aus der Vergangenheit als auch mit Stimmen gesunder Probanden verglichen. Die Stimme von Kehlkopfkrebspatienten klingt rauher und behauchter. Diese zeitliche Stimmentwicklung findet sich in den Analyseparametern Jitter, Shimmer und Harmonicity wieder.

Mithilfe dieser Erkenntnisse könnte zukünftig eine Frühdiagnose von Erkrankungen an den Stimmlippen möglich sein. Im Rahmen der regelmäßigen Vorsorgeuntersuchungen (z.B. beim Hausarzt) oder auf einem Internetportal könnte jeder von uns eine kurze Sprachprobe hinterlegen. Da die untersuchte Messmethode nicht-invasiv, kosten- und aufwandsgünstig ist, kann sie weltweit diagnostisch eingesetzt werden. Zu den wissenschaftlichen Herausforderungen zählt die Berücksichtigung von Nikotinkonsum auf die Stimme sowie altersbedingte Stimmveränderungen [10].

8. Danksagung

Die Arbeiten wurde im Rahmen von Jugend forscht 2018 durchgeführt. Ich bedanke mich herzlichst bei Herrn Prof. Dr.-Ing. U. Heute für seine fachliche Unterstützung auf dem Gebiet der Phonetik und die Kontaktherstellung zum UKSH. Herrn Dr. rer. nat. M. Hey und Frau A. Heinhold von der Klinik für Hals-, Nasen- Ohrenheilkunde, UKSH danke ich für die nette Diskussion zu meinen Arbeiten. Bei M.Sc. Christin Baasch von der Technischen Fakultät der CAU Kiel möchte ich mich für die Kurzeinweisung in das Analyse-Programm Praat bedanken. Herrn Dr. med. J. Orthmann danke ich für seinen Hinweis auf Papillomaviren, die durch veränderte gesellschaftliche Verhaltensweisen zu einer wachsenden Gefahr werden. Zuguterletzt bedanke ich mich für das inspirierende Gespräch mit Prof. Dr. E. Nöth auf der DAGA 2019 in Rostock.

9. Literatur- und Quellenverzeichnis

- [1] U. Heute, „Observations from a Simple Vocal-Tract Model's Behaviour for PD-Dysarthric Speech: Evaluations,“ in Proc. *DAGA 2019*, Rostock, März 2019.
- [2] B. Bartke, T. Haderlein, M. Döllinger, E. Nöth, S. Graf, U. Eysholdt, A. Ziethe, „Perzeptive und maschinelle Stimm- und Sprechanalyse bei chronischer Laryngitis und T1-Stimmlippenkarzinom,“ *HNO*, Vol. 61, Issue 8, pp. 672-677, Aug. 2013.
- [3] R. Clapham, C. Middag, F. Hilgers, J.-P. Martens, M. Van Den Brekel, R. Van Son, „Developing Automatic Assessment Techniques for Speakers Treated for Advanced Head and Neck Cancer,“ *Speech Communication*, Vol. 59, pp. 44-54, April 2014.
- [4] Robert Koch-Institut (Hrsg.), *Krebs in Deutschland 2007/2008*. Berlin, 8. Ausg., 2012.
- [5] J. Wendler, W. Seidner, U. Eysholdt, *Lehrbuch der Phonatrie und Pädaudiologie*. Stuttgart: Georg Thieme Verlag, 4. Aufl., 2005.
- [6] T. Haderlein, C. Schwemmler, M. Döllinger, V. Matoušek, M. Ptok, E. Nöth, „Automatic Evaluation of Voice Quality Using Text-Based Laryngograph Measurements and Prosodic Analysis,“ *Computational and Mathematical Methods in Medicine*, Vol. 2015, Article ID 316325, 2015.
- [7] Praat Homepage: <http://www.fon.hum.uva.nl/praat/>.
- [8] P. Vary, U. Heute, W. Hess, *Digitale Sprachsignalverarbeitung*. Stuttgart: B.G. Teubner, 1998.
- [9] S. Euler, *Grundkurs Spracherkennung*. Wiesbaden: Vieweg, 2006.
- [10] L. Schwinn „Impact of Pathologies on Age Estimation,“ in Proc. *DAGA 2019*, Rostock, März 2019.