Quantitative examination of vocal folds, perspectives for image analysis and OCT with ultra-high resolution

Summary

To directly relate tissue abnormalities to dysfunctional voicing, it is decisive to temporally resolve the vocal fold movement during phonation; and that on the microscopic level. High-speed video (HSV) can record the vocal folds with 2-4.000 fps. Ultrahigh resolution optical coherence tomography (UHR-OCT) can distinguish cellular layers with a resolution better than 5 µm within a tissue depth of 1 mm. We propose combining the two technologies and apply deep learning-based image segmentation to establish statistical evident and reproducible documentation for voice related diseases.

Introduction

The human voice production takes place in the larynx. We have attempted to collect the latest research in imaging and analysis of the vocal folds, in which we will review options for quantification of vocal fold measurements [1, 2], We will describe the application of Optical Coherence Tomography with ultra-high resolution (UHR-OCT) and deep learning (subgenre of artificial intelligence) for quantitative analysis of the vocal folds based on High-Speed Video (HSV).

The perspective is to combine the methods, partially in cooperation with DTU Photonics (Technical University of Denmark). A review was made accompanying the project “Shape OCT” [3]. DTU Photonics has offered to develop a new probe for UHR-OCT to analyse the movement in real time both in the tissue, and on a microscopical level.

These results can be combined with High-Speed Video (HSV) of the vocal folds, and artificial intelligence. We can obtain possibilities for diagnostics similar to the practices for eye- and skin- conditions, where there is a great interest for tissue understanding [5, 6, 7].

The general practitioner has an interest in knowing the possibilities for diagnosing voice disorders, considering that artificial intelligence is being leveraged on multiple areas in our daily lives. The second perspective concerning the general practitioner is the possibility for understanding the tissue function. There is off course a difference between the focus of ophthalmologists and dermatologists, and the laryngological aspects, but until now we haven’t had the possibility for non-invasive analysis of tissue during phonation for differential diagnostics. We have only had the option to evaluate the tissue in immobile vocal folds [8, 9].

We will attempt to describe the technology for UHR-OCT and provide examples for clinical use. We wish to present the new technologies that provides solutions for an exact diagnosis of e.g. benign and malignant tumors in the larynx during phonation. Combining HSV with UHR-OCT, for analysing vocal fold movement and the mucosa, is suitable because of the higher frequencies that UHR-OCT provides are a better match to the 4.000 frames per second of HSV [10]. Large amounts of data can be collected with artificial intelligence [11, 12], and longer sequences [13].

High-Speed Video (HSV)

The possibility for exact diagnosis of mucosal changes in the larynx with altered phonation, is improved considerably with HSV [1, 2]. This apparatus can e.g. analyse a slowmotion recording of the vocal folds exact movement with 2-4.000 frames per second (Richard Wolf GmbH Endocam 5562). Stroboscopy has become widespread in laryngology, but only records around 25 frames per second (Hz), Therefore HSV was a great improvement since the average frequency for speech in men and women are respectively 110 Hz and 220 Hz (oscillation per second). So far the equipment for HSV is more expensive than for stroboscopy.

HSV visualizes the real time oscillations of the right and left vocal fold, and the area between them is clearly presented by marking the edges of the vocal folds (segmentation).
We often see benign disorders with more or less visible oedema of the vocal folds with HSV [14]. It can be challenging to give a sufficient diagnosis only relying on HSV, especially for those depending on their voice in their professional lives. Another interesting aspect, that HSV cannot assist in either, is the hormonal influence on the tissue during puberty [15].
HSV can be used to optimize differential diagnostic of tumors – benign and malignant, scar tissue, bleeding, traumas, etc. because it visualizes the real times oscillation of the vocal folds [16, 17, 18].
It is necessary to combine HSV with another technology to obtain sufficient statistical evidence for the aforementioned conditions and symptoms [19].

Figure 1 A & B. Shows the normal larynx with manual marking of the area for analysis and the center of glottis for quantitative analysis with HSV, and a marking of the vocal fold edges for quantitative calculations (segmentation). These visualizations have not been sufficient to explain and attain statistical characteristics and evidence [20].
C. Shows an image from a HSV (4.000 frames per second), of a patient with mucus reflux from the stomach irritating the larynx, this happens in less than 0,2 seconds [21].

Figure 1 – Images of the larynx

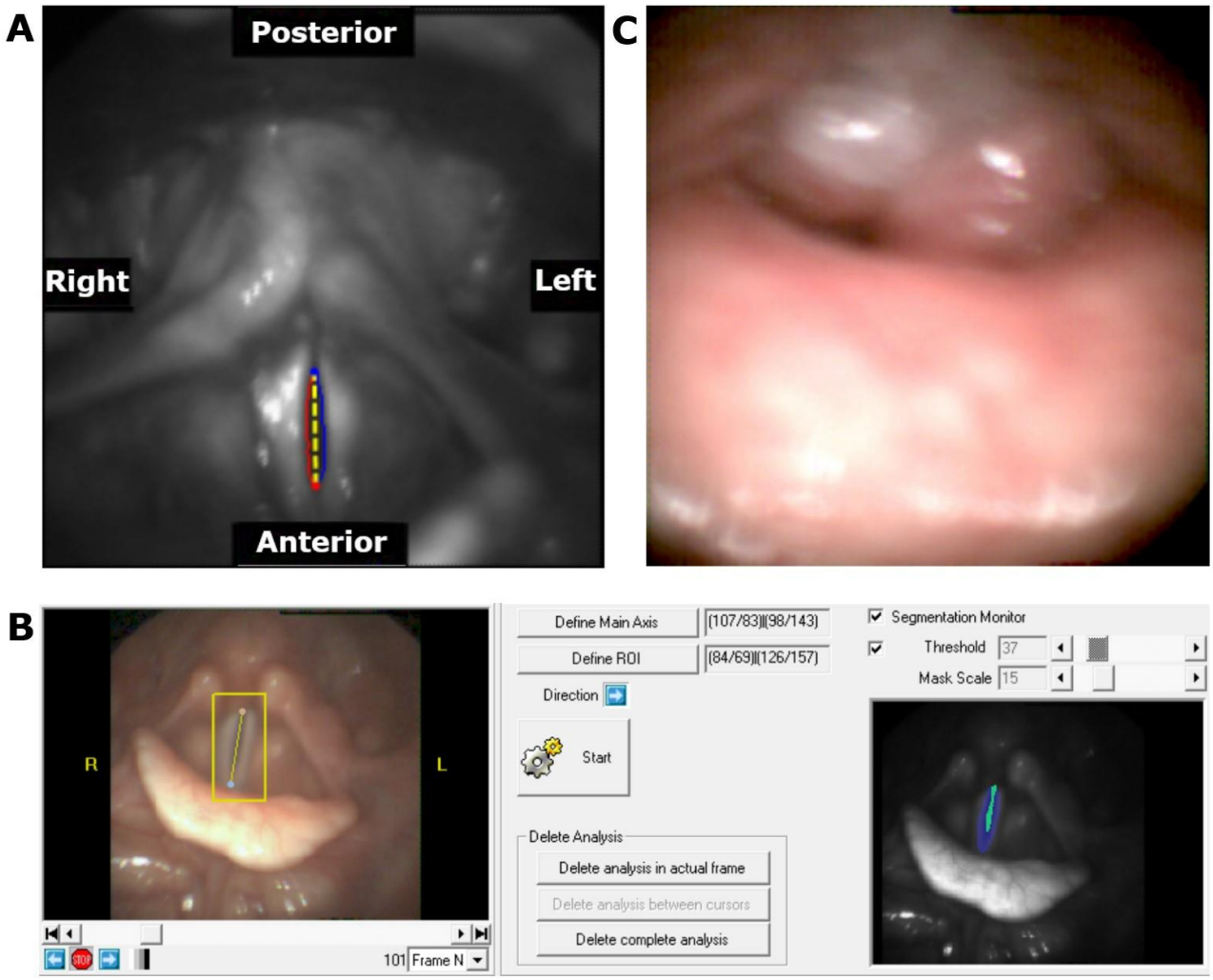


Figure 1 A. A normal larynx with markings of the center of glottis. **B.** Software measurements and markings of the edges of the vocal folds with HSV. **C.** An image from a High-Speed Video with mucus reflux in the larynx, happening in less than 0,2 seconds.

Deep learning (artificial intelligence)

Artificial intelligence (AI) is generally known to be used in cases with large amounts of data such as video sequences. Deep learning is a subgenre of AI and is in its infancy for voice analysis but requires large amounts of processing power. Deep learning has great future potential since it contains a feedback loop that makes it self-learning, thereby increasing the precision [22]. Mona Fehling et al. [11] reviews different possibilities for deep learning and concludes that a U-LSTM-segmentation is best for analysis of the vocal folds (Figure 2).

In Figure 2 we show how deep learning can be used quantitatively for differential diagnosis between vocal folds with normal closure and vocal folds with insufficient closure. It is worth noting that the relative area between the vocal folds is visualized. Deep learning is used for automatic segmentation of the vocal folds. With manual segmentation the clinician has to mark the edges of the vocal folds meticulously before the calculations can be made, but since the images in the video often move and varies in color, errors in the calculations often occur. What often had to be done manually and with limited success can now be done automatically. According to Kist et al. [12] manual segmentation between the vocal fold edges and glottis takes more than 15 minutes for a trained professional to perform accurately, and less than 1 minute for their neural network. This will save large amounts of time in the daily treatment of patients for ear- nose-throat-specialists and increase precision in diagnosing benign and malignant tumors/leukoplakia, sulcus and many other voice disorders.

Figure 2 shows examples on the clinical importance of deep learning, based on HSV analysed by Mona Fehling. **A.** (a) shows images from a high-speed video of a normal subject from our own database, (b) Ground truth is the trained professionals manual segmentation of the vocal folds. Ground truth is used to measure the precision of a neural networks (deep learning) ability to perform segmentation, (c & d) shows segmentation results for 2 neural networks, where U-LSTM is the more precise [11].
B. (a) shows HSV and part of a single oscillation from a patient with insufficient closure in the rear from our database, (b) the chosen neural network (U-LSTM) is used to estimate the right and left vocal fold, and glottis, (c) Layering the estimation of the neural network on top of the original image for comparison, (d) shows the relative area between the vocal folds on a curve given in time (of the 100 frames per analysis). These results are the latest from Trier (University of Applied Sciences in Germany) and is of considerable importance for future evidence of diagnosing disorders in the larynx [11].

Figure 2 – Deep learning

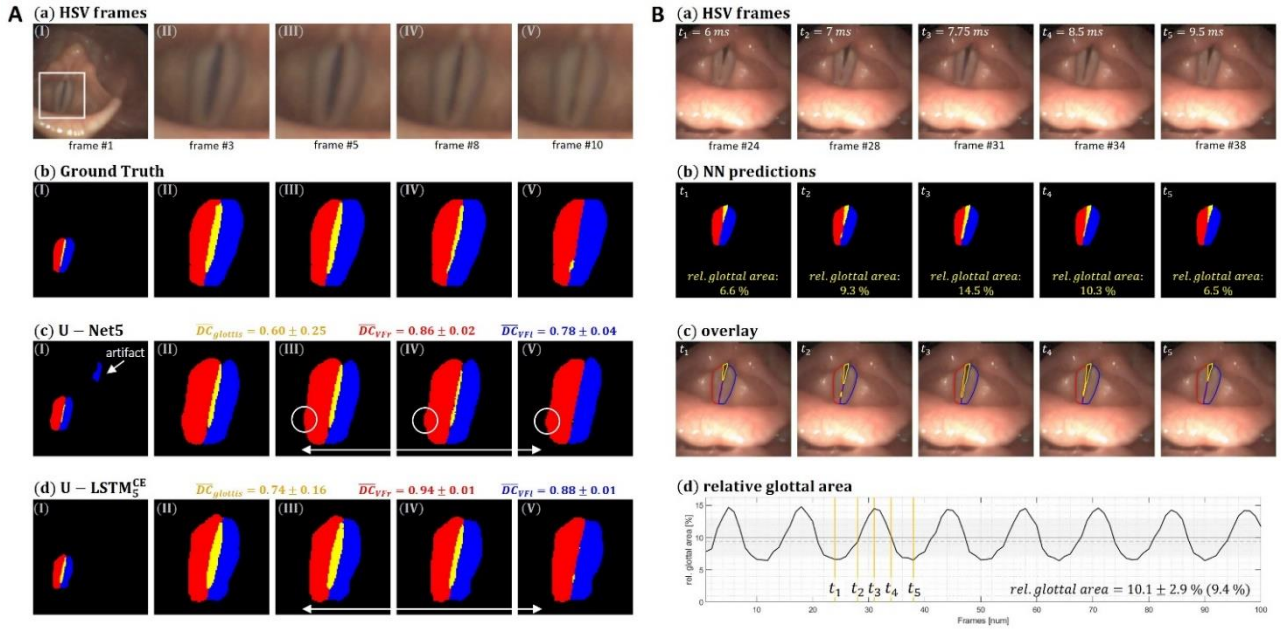


Figure 2 A. Shows a comparison of the segmentation results of individual images of a normal subject from our database, (a) High-Speed Video, (b) Ground Truth, (c) U-Net segmentation, and (d) U-LSTM segmentation. The dice coefficient for each class represents the mean and standard deviation for the whole sequence (100 frames).

B. Shows segmentation results for a HSV of a single oscillation from a patient with insufficient closure in the rear from our database, (a) High-Speed Video, (b) neural network segmentation generated with U-LSTM_5^CE, (c) overlay of segmentation results, and (d) mean and standard deviation for the normalized relative area between the vocal folds for the entire sequence, equal to the sum of the area between the vocal folds + right vocal fold + left vocal fold. The yellow lines indicate from where in the oscillation the above pictures are taken.

(Analysis with deep learning reproduced with permission from Mona Fehling, University of Applied Sciences, Trier).

Optical Coherence Tomography with Ultra-High Resolution (UHR-OCT)

Optical coherence tomography (OCT) is a more recent scanning method than ultrasound. OCT uses light instead of sound and can therefore achieve a more special resolution [23, 24 ,25]. UHR-OCT has a spatial resolution of less than 5 µm, and a can reach depths between 0,4 and 1mm in the tissue. OCT has been used with patients in anaesthesia, with no phonation [8, 9]. OCT improves the ability to differentiate between which tumors to operate and which to treat otherwise. The new aspect is that the tissue can be analysed during phonation with UHR-OCT to avoid the risks related to invasive procedures.

Until now we haven't been able to document evidence-based treatment effect for the subjective complaint: Hoarseness. With OCT it is possible to see the cellular layers of the vocal folds during phonation, including the regularity of the edges, this is great progress. Unfortunately, the speed of OCT is typically between 50-100 cross-section images per second, which causes artifacts [25]. This complicates a precise diagnosis of the vocal fold movement. This was also documented when stroboscopy and electroglottography was combined many years ago [26].

There is now constructed a UHR-OCT setup that can combine HSV during phonation (4.000 frames per second) [10]. The high resolution of OCT provides precise information of the cellular levels, for a better understanding of dysfunction and mucosal changes in the larynx, especially for the vocal folds.
DTU Photonics has produced a handheld probe for imaging of the oral mucosa [10]. A probe for examination of the larynx during phonation can have the shape and length of a laryngoscope. It is therefore possible to combine it with the probe for HSV. It contains a laser pointer for aiming at the area of interest [25, 27]. A probe for imaging of the vocal folds during phonation contains a line scan procedure (in contrast to traditional laser spot scan) combined with a 2D camera spectrometer for achieving sufficient speed (frames per second) and is equipped with a supercontinuum source for achieving sufficient depth resolution. Therefore, creating the possibility for direct diagnostics with UHR-OCT of the larynx and the vocal folds during phonation. Biopsies is an invasive procedure and UHR-OCT is preferable in benign disorders because the examination can be performed without anaesthesia and with minor discomfort. Some initial issues with interpretation are expected until a standardisation has been made.
Figure 3 A. Equipment for OCT on a probe presented with permission from Beckman Institute in California. **B.** A cross-section of closed vocal folds with OCT. **C.** Setup for OCT in Beckman Institute. **D.** An image from a video of OCT with 200 frames per second.
Below an *in vivo* recording made with UHR-OCT setup from DTU Photonics is shown [10]. **E.** From the oral mucosa (Inside the lower lip) with epithelia, glands and blood vessels, **F.** Individual skin papillae and capillaries of the hand, These recordings illustrate the advantage of ultra-high resolution [16].

Figure 3 – OCT-setup

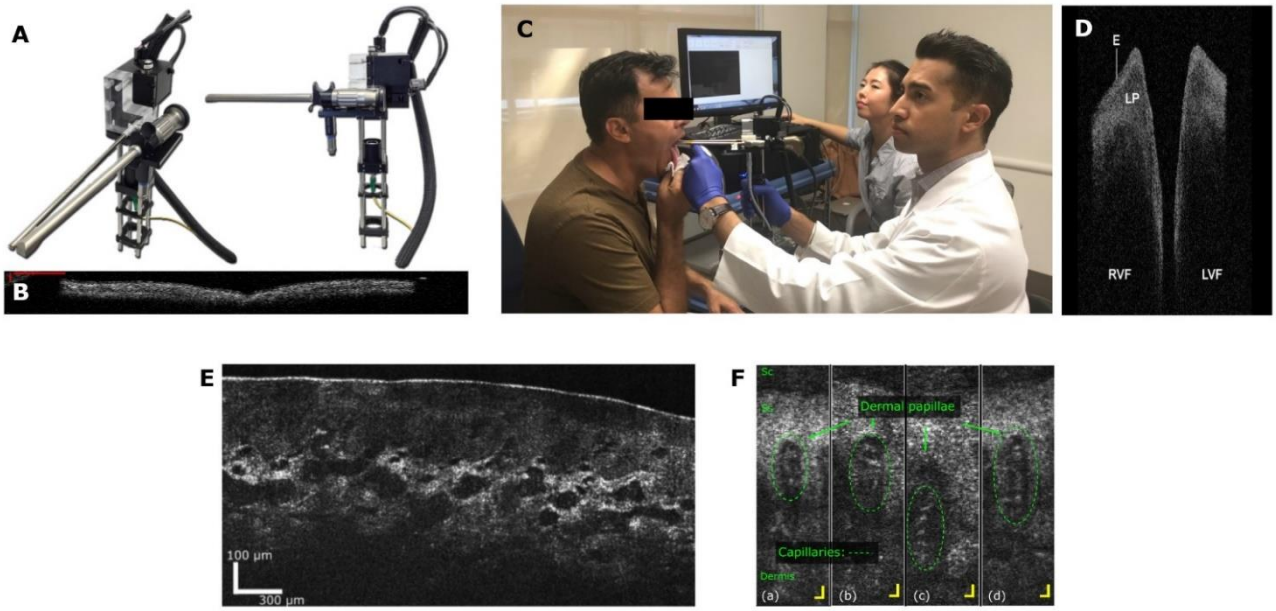


Figure 3 OCT-setup **A.** Equipment for OCT on a probe from Beckman Institute in California. **B.** A cross-section of closed vocal folds with OCT. **C.** Setup for OCT in Beckman Institute. **D.** An image from a video of OCT with 200 frames per second. (A-D with permission from Brian Wong, Beckman Institute) **E.** In vivo recording UHR-OCT from oral mucosa (inside of lower lip) with epithelia, glands, and blood vessels [10, 28]. **F.** Individual skin papillae and capillaries of the hand, scale is 20 µm [6].

Discussion and conclusion

The purpose of this paper is to introduce possibilities for combining HSV with other methods for diagnostic of the upper airways, especially when voice disorders result in decrease function of the vocal folds. It is persistently difficult to differentiate causes for patient complaints of hoarseness sufficiently [17, 18, 19, 20]. Due to the large amounts of data contained in HSV, we presented some new research results with different technologies and their possibilities in the frame of laryngology. We have discussed how the vocal fold edges and glottis can be segmentated automatically, so we can achieve faster and more precise diagnosis for many disorders in laryngology.

We have presented options for a better tissue analysis that provide solutions for exact diagnoses. HSV and UHR-OCT can be combined for analysis of vocal fold movement and mucosa, because UHR-OCT has a higher frequency that matches HSV [10]. This can be supported with deep learning. Stroboscopy causes artefacts due to low amount of frames per second compared to the vocal fold frequency, therefore difficult to quantify [1]. Stroboscopy only presents every 4th vocal fold oscillation with an image, dependant on the frequency. It was a step in the right direction initially, and research is being made into combining

stroboscopy and OCT [29]. The next step for understanding the function of the larynx during phonation is to combine UHR-OCT and deep learning with HSV, in order to increase evidence-based quantitative analysis.

Litterature

1. Woo P. Stroboscopy and High-Speed Imaging of the Vocal Function. Plural publishing 2nd ed. 2021.
2. Pedersen M, Eeg M, Jønsson A et al. Chapter 8, Working with Wolf Ltd. HRES 5562 Analytic system for high-speed recordings, Normal & Abnormal Vocal folds Kinematics: HSDP, OCT & NBI, 2015;1:57-65.
3. DTU fotonik. ShapeOCT (2015-2019), med 25,7 mio. fra Innovationsfonden, Grant/Award Number: 4107-00011A.
4. Ran AR, Tham CC, Chan PP et al. Deep learning in glaucoma with optical coherence tomography: A review. Eye. 2021 Jan;35(1):188-201.
5. Del Amor R, Morales S, Colomer A et al. Automatic segmentation of epidermis and hair follicles in optical coherence tomography images of normal skin by convolutional neural networks. Frontiers in Medicine. 2020 Jun 4;7:220.
6. Israelsen NM, Maria M, Mogensen M et al. The value of ultrahigh resolution OCT in dermatology - delineating the dermo-epidermal junction, capillaries in the dermal papillae and vellus hairs, Biomedical Optics Express, 2018; 9(5). 2240-2265.
7. Israelsen NM, Mogensen M, Jensen M et al. Delineating papillary dermis around basal cell carcinomas by high and ultrahigh resolution optical coherence tomography-A pilot study. J Biophotonics. 2021 Jul 10:e202100083.
8. Brian J Wong. In Vivo Optical Coherence Tomography of the Human Larynx: Normative and Benign Pathology in 82 Patients. Laryngoscope. 2005;115(11):1904-11.
9. Klein AM, Pierce MC, Zeitels SM et al. Imaging the Human Vocal Folds in Vivo with Optical Coherence Tomography: A Preliminary Experience. Annals of Otology, Rhinology & Laryngology. 2006;115(4):277-284.
10. Israelsen NM, Jensen M, Jønsson AO et al. Ultrahigh Resolution Optical Coherence Tomography for Detecting Tissue Abnormalities of the Oral and Laryngeal Mucosa: A Preliminary Study, MAVEBA Proceedings, 2016; 195-197.
11. Fehling MK, Grosch F, Schuster ME et al. Fully automatic segmentation of glottis and vocal folds in endoscopic laryngeal high-speed videos using a deep Convolutional LSTM Network. PLoS One. 2020;15(2):1–29.

12. Kist AM, Gómez P, Dubrovskiy D et al. A Deep Learning Enhanced Novel Software Tool for Laryngeal Dynamics Analysis. J. Speech, Language and Hearing Research. 2021;64(6):1889-1903.
13. Yousef AM, Deliyski DD, Zacharias SRC et al. A Hybrid Machine-Learning-Based Method for Analytic Representation of the Vocal Fold Edges during Connected Speech. Applied Sciences. 2021; 11(3):1179.
14. Watanabe T, Kaneko K, Sakaguchi K, Takahashi H. Vocal-fold vibration of patients with Reinke's edema observed using high-speed digital imaging. Auris Nasus Larynx. 2016;43(6):654-657
15. Garcia JA, Benboujja F, Beaudette K et al. Using attenuation coefficients from optical coherence tomography as markers of vocal fold maturation. Laryngoscope. 2016; 126(6): E218–23.
16. Pedersen M, Agersted A, Akram B et al. Optical coherence tomography in the laryngeal arytenoid mucosa for documentation of pharmacological treatments and genetic aspects: a protocol, Advances in Cellular and Molecular Otolaryngology, 2016; 4:1.
17. Roth DF, Abbott KV, Carroll TL et al. Evidence for primary laryngeal inhalant allergy: a randomized, double-blinded crossover study. International forum of allergy & rhinology. 2013;(1):10-8.
18. am Zehnhoff-Dinnesen A, Wiskirska-Woznica B, Neumann K et al. Phoniatrics I. Springer Berlin Heidelberg 2020.
19. Pedersen M, McGlashan J. Surgical versus non-surgical interventions for vocal cord nodules (Review), The Cochrane Library. 2012; 1-13.
20. Pedersen M. Which Mathematical and Physiological Formulas are Describing Voice Pathology: An Overview, Journal of General Practice, 2016; 4:3.
21. Woisard V: Gastro-esopharyngeal Reflux Influences on Larynx and Voice, page 263-271. In Phoniatrics I. am Zehnhoff-Dinnesen A, Wiskirska-Woznica B, Neumann K, Nawka T, editors Springer Berlin Heidelberg 2020.
22. Pham TT, Chen L, Heidari AE et al. Computational analysis of six optical coherence tomography systems for vocal fold imaging: A comparison study. Lasers in surgery and medicine 2019; 51:412-422.
23. Sergeev AM, Gelikonov GV, Gelikonov FI et al. In vivo endoscopic OCT imaging of precancer and cancer states of human mucosa. Optics express, 1997; 1,13: 434-440.
24. Just T, Guder E, Witt G et al. Confocal Endomicroscopy and Optical Coherence Tomography for Differentiation Between Low-Grade and High-Grade Lesions of the Larynx in Biomedical Optics in Otorhinolaryngology: Head and Neck surgery. Eds. Springer New York, 2016; 479-490.

25. Coughlan CA, Chou L, Jing JC et al. In vivo cross-sectional imaging of the phonating larynx using long-range Doppler optical coherence tomography. Nature Science Reports. 2016; 6: 22792.
26. Pedersen MF. Electroglottography compared with synchronized stroboscopy in normal persons. Folia phoniatr; 1977; 29:191-200.
27. Donner S, Bleeker S, Ripken T et al. Automated working distance adjustment enables optical coherence tomography of the human larynx in awake patients, J Med imaging (Bellingham, Wash.) 2015; 2.2, 026003.
28. Wei W, Choi WJ, Men S et al. Wide-field and long-ranging-depth optical coherence tomography microangiography of human oral mucosa (Conference Presentation), Proceedings SPIE 10473, Lasers in Dentistry XXIV, 2018, 104730H
29. Maguluri GN, Mehta DD, Kobler JB et al. Optical biopsy of vocal folds during phonation using parallel OCT (Conference Presentation). In: Alfano RR, Demos SG, Seddon AB, editors. Optical Biopsy XVII: Toward Real-Time Spectroscopic Imaging and Diagnosis. SPIE; 2019; 13.

DANISH VERSION

Kvantitativ undersøgelse af stemmebånd, perspektiver for billedanalyse og OCT med ultrahøj opløsning

Introduktion

Den menneskelige lyddannelse foregår i struben. Vi har forsøgt at samle den seneste forskning inden for billeddannelse og analyse af stemmebåndene, hvor vi vil gennemgå nogle muligheder til kvantificering af stemmebåndsmålinger [1, 2]. Vil vi beskrive anvendelse af optisk kohærens tomografi med ultrahøj opløsning (på engelsk forkortet: UHR-OCT) og deep learning (en gren inden for kunstig intelligens) til kvantitativ analyse af stemmebånd på basis af High-speed video (HSV).

Perspektivet er at koordinere metoderne, delvist i samarbejde med DTU Fotonik (Danmarks Tekniske Universitet). En statusartikel derfra foreligger efter færdiggørelse af projekt ”Shape OCT” [3]. Med et nyt projekt har DTU Fotonik tilbudt at udvikle en probe til UHR-OCT således at stemmebåndenes bevægelser kan følges direkte, nede i vævet og på mikroskopisk niveau. Disse resultater kan kombineres med high-speed video (HSV) af stemmebåndene og kunstig

intelligens. Hermed kan der opnås store muligheder for diagnostik svarende til lignende forhold for øjenlidelser [4] og hudlidelser, hvor der er stor interesse for vævsforståelse [5, 6, 7].
De praktiserende læger har interesse i at vide hvor vi står med hensyn til diagnose muligheder for stemmelidelser, nu hvor kunstig intelligens har gjort sit indtog i alle aspekter omkring os.
Det andet perspektiv som også vedkommer praktiserende læger er vores nye muligheder for vævsforståelse. Principielt er der naturligvis forskel på hvilke lidelser øjen- og hudlæger fokuserer på, og vores laryngologiske aspekter, men indtil nu har vi ikke haft muligheder for noninvasiv vævsanalyse under fonation til differential diagnostik. Vi har kun haft muligheden for at vurdere vævet med immobile stemmebånd [8, 9].
Vi vil beskrive teknologien til UHR-OCT og give eksempler på brug heraf. Vi ønsker at vise de nye teknologier der giver løsninger til en eksakt diagnose af f.eks. benigne og maligne tumorer i struben under fonation. En kombination af high-speed video (HSV) og UHR-OCT, både mht. stemmebånds bevægelser og slimhindelidelser, som en del af øvre luftveje giver mening fordi UHR-OCT har høje frekvenser som svarer til HSV [10]. Store datamængder kan samles med AI [11, 12], også med længere sekvenser [13].

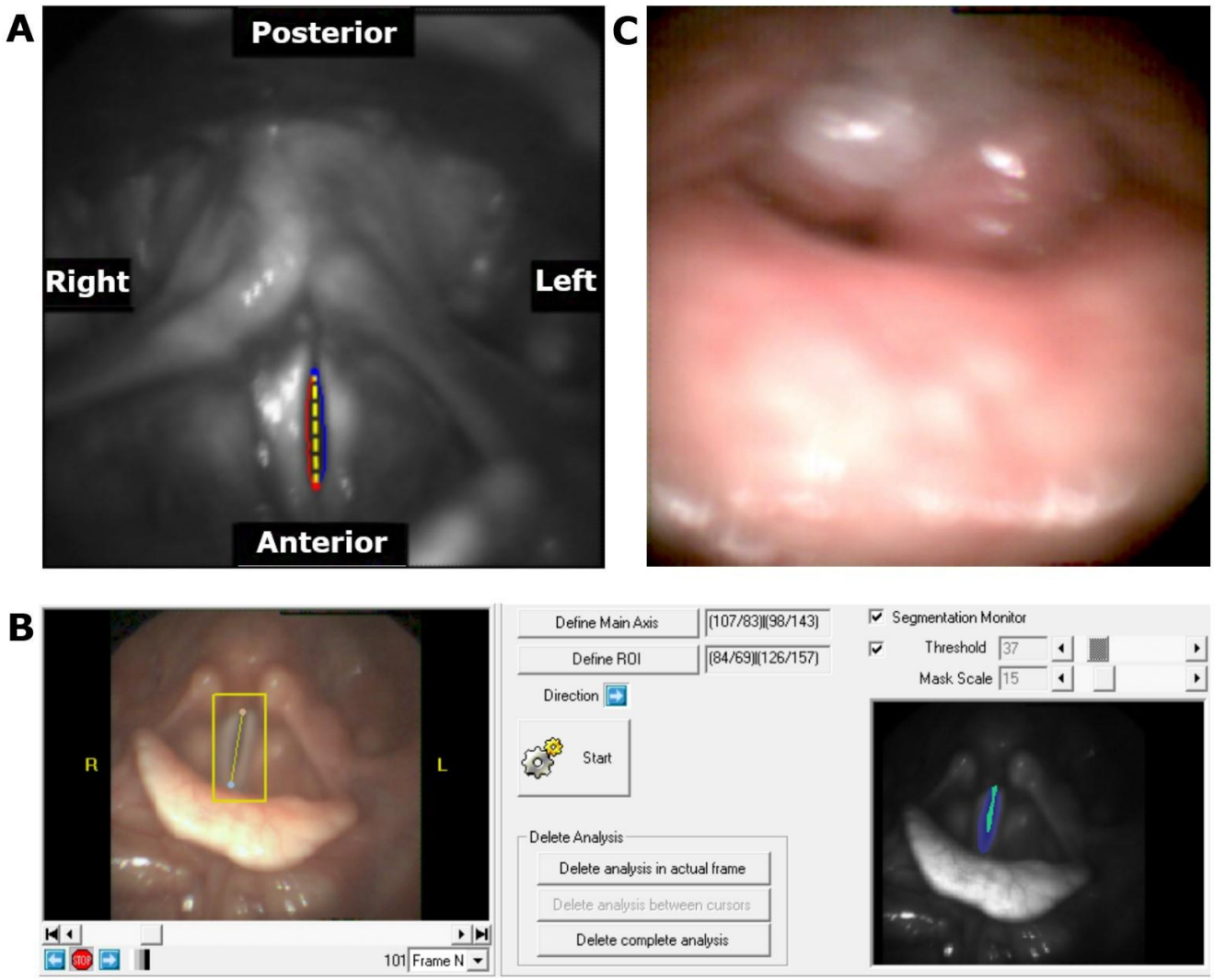
High-speed video (HSV)
Muligheden for en eksakt diagnose for slimhindeforandringer i struben med ændret fonation til følge, er forbedret væsentligt med high-speed video (HSV) [1, 2]. Med dette apparatur kan der bl.a. analyseres en slowmotion video med 2-4.000 billeder af stemmebåndenes eksakte bevægelser per sekund (Richard Wolf GmbH Endocam 5562). Stroboskopi har vundet udbredt anvendelse i laryngologien, men optager kun 25 billeder i sekundet (Hz), derfor er HSV en stor forbedring da gennemsnits tale-frekvenser for hhv. mænd og kvinder er på 110 Hz og 220 Hz (svingninger pr. sek.). Udstyret til HSV er indtil videre dyrere end stroboskopi.
HSV visualiserer de regelrette svingninger af højre og venstre stemmebånd samt arealet imellem dem, der ses tydeligt ved en markering af stemmebåndenes kanter, kaldet segmentering.

Vi ser med HSV ofte benigne lidelser med mere eller mindre tydeligt ødem af stemmebåndene [14]. Det kan være vanskeligt med HSV at give patienterne en sufficient diagnose, især til dem der er afhængige af stemmebrug. Et andet interessant aspekt, der heller ikke kan forklares ved hjælp af HSV, er forståelse af hormoners indflydelse på vævet ved stemmens udvikling i puberteten [15].
HSV kan bruges til at optimere differential diagnostik af tumorer – benigne og maligne, arvæv, blødninger, traumer og lignende, fordi at man har et retvisende billede af stemmebåndenes svingninger [16, 17, 18].

Det er derfor meget aktuelt at koble en anden teknologi sammen med HSV for at opnå den nødvendige statistiske evidens for ovenstående sygdomme og symptomer [19].

Figur 1 A & **B.** Viser en normal strube med manuel markering af området til analyse og midten af stemmeridsen til brug ved HSV til kvantitativ analyse, samt en markering af stemmebåndenes kanter, med henblik på kvantitative beregninger (segmentering). Disse visualiseringer har ikke været tilstrækkelige til f.eks. at forklare og opnå statistiske kendetegn og evidens [20].
C. Viser et billede fra en HSV (4.000 billeder per sekund), af en patient hvor slim kommer op fra mavesækken og irriterer struben, dette forekommer på mindre end 0,2 sekunder [21].

Figur 1 - Billeder af struben



FIGUR 1 A. En normal strube med markeringer af midten af stemmeridsen. **B.** Software målinger der markerer de frie kanter af stemmebåndende med high-speed film. **C.** Et billede fra en high-speed video, hvor slim kommer op i struben og forsvinder igen på 0,2 sekund.

Deep learning (kunstig intelligens)

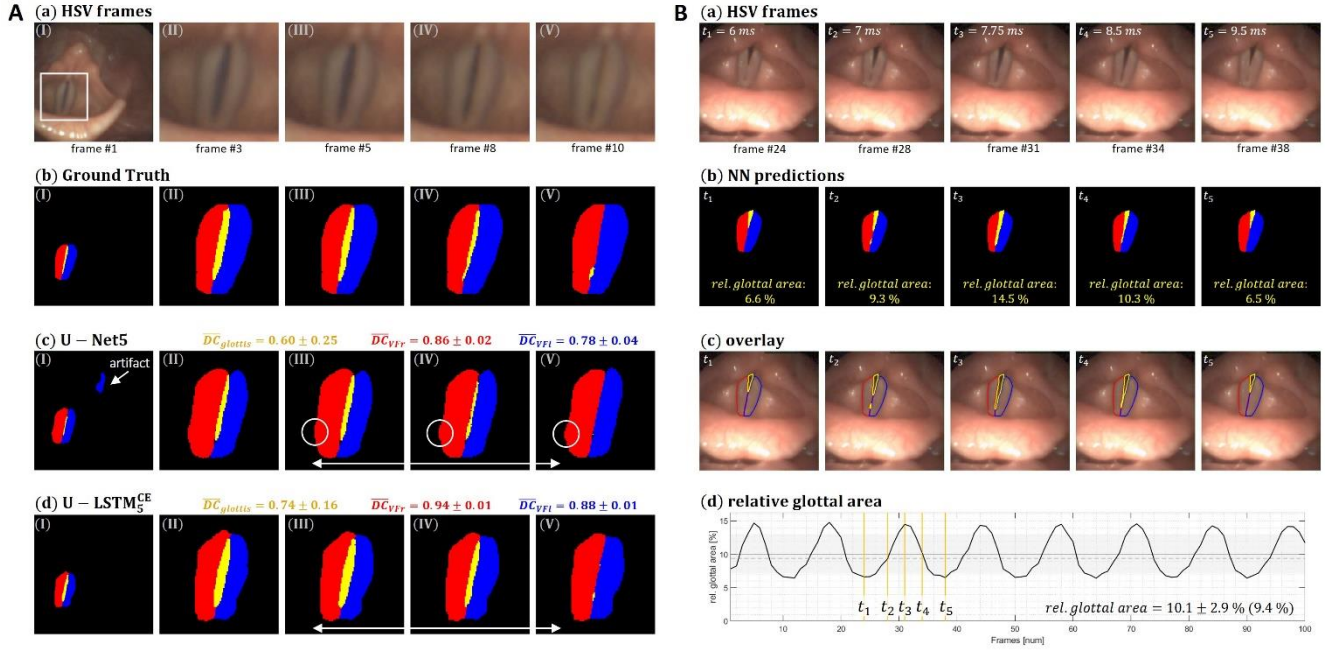
Kunstig intelligens (AI) er generelt kendt for at kunne bruges i tilfælde hvor man har store mængder data fra f.eks. videosekvenser. Deep learning er en gren af AI og er i sin begyndelse blandt værktøjerne til stemmeanalyse, men kræver meget processorkraft. Deep learning har store fremtidsmuligheder da det indeholder et feedback loop som gør det selvlærende, dermed øges præcisionen [22]. Mona Fehling et al [11] gennemgår de forskellige muligheder for deep learning og konkluderer at U-LSTM-segmenteringstypen er bedst til analyse af stemmebånd (Figur 2).

I Figur 2 viser vi hvordan deep learning kan bruges til kvantitativt at differential diagnosticere mellem stemmebånd med normal lukkefunktion og stemmebånd med insufficient lukkefunktion. Det er værd at bemærke at man samtidig kan registrere det relative areal mellem stemmebåndene. Deep learning er benyttet til at automatisere segmenteringen af stemmebåndene. Ved manuel segmentering skulle lægen markere stemmebåndenes kanter på strubebilledet før beregningerne kunne foretages, men da billedet bevæger sig og ofte varierer i farve, resulterer det i fejl ved beregningerne. Dette, der før skulle gøres manuelt og ofte med begrænset succes, kan nu gøres automatisk. Ifølge Kist et al. [12] tager manuel segmentering, differentiering med stemmebåndenes kanter og åbningen imellem dem over 15 min for en specialist at gøre præcist, og under 1 minut for deres neurale netværk. Dette vil spare store mængder tid i den daglige behandling af patienter for Øre- Næse- Halslæger og især øge præcisionen til diagnostik af benigne og maligne tumorer/leukoplakier, sulcus og mange andre stemmelidelser.

Figur 2 viser eksempler på den kliniske betydning af deep learning, som baseres på high-speed video (HSV) analyseret af Mona Fehling. **A.** (a) viser billeder fra en high-speed video fra en normal person fra vores egen database, (b) Ground truth er specialisters manuelle segmentering af stemmebåndene. Ground truth bliver brugt til at måle hvor præcist et givent neural netværk (deep learning) er til at segmentere, (c & d) viser segmenterings resultater for 2 neurale netværk, hvor U-LSTM er mere præcist [11].
B. (a) viser HSV og dele af en enkeltcyklus af en patient med insufficient lukkefunktion bagtil fra vores database, (b) bruger det valgte neural netværk (U-LSTM) til at estimere venstre og højre stemmebånd, og arealet imellem dem, (c) lægger estimeringen af det neurale netværk ovenpå det originale billede til sammenligning, (d) viser det relative areal mellem stemmebåndene på en kurve givet i tid (ud af 100 billeder per analyse).

Disse resultater er de seneste fra Trier (University of Applied Sciences i Tyskland) og er af væsentlig betydning for fremtidigt at opnå evidens til diagnosticering af sygdomme i struben [11]:

Figur 2 – Deep learning



FIGUR 2 A. Viser en sammenligning af segmenterings resultater af individuelle billeder fra en normal person fra vores database, (a) high-speed video, (b) Ground Truth, (c) U-Net segmentering, og (d) U-LSTM segmentering. Dice koefficienten for hver klasse repræsenterer gennemsnit og standarddeviationen for hele sekvensen (100 billeder).
B. Viser segmenteringsresultatet for en HSV af en enkelt svingingscyklus fra en patient med insufficient lukkefunktion bagtil fra vores database, (a) high-speed video, (b) neuralt netværks segmentering genereret med U-LSTM_5^CE, (c) overlay af segmenterings resultater, og (d) gennemsnit og standardafvigelser for det normaliserede relative areal mellem stemmebåndene for hele sekvensen, svarende til summen af arealet mellem stemmebåndene + højre stemmebånd + venstre stemmebånd. De gule linjer indikerer hvorfra i svingningen de ovenstående billeder er taget.
(Analyse med deep learning gengivet med tilladelse fra Mona Fehling, University of Applied Sciences, Trier).

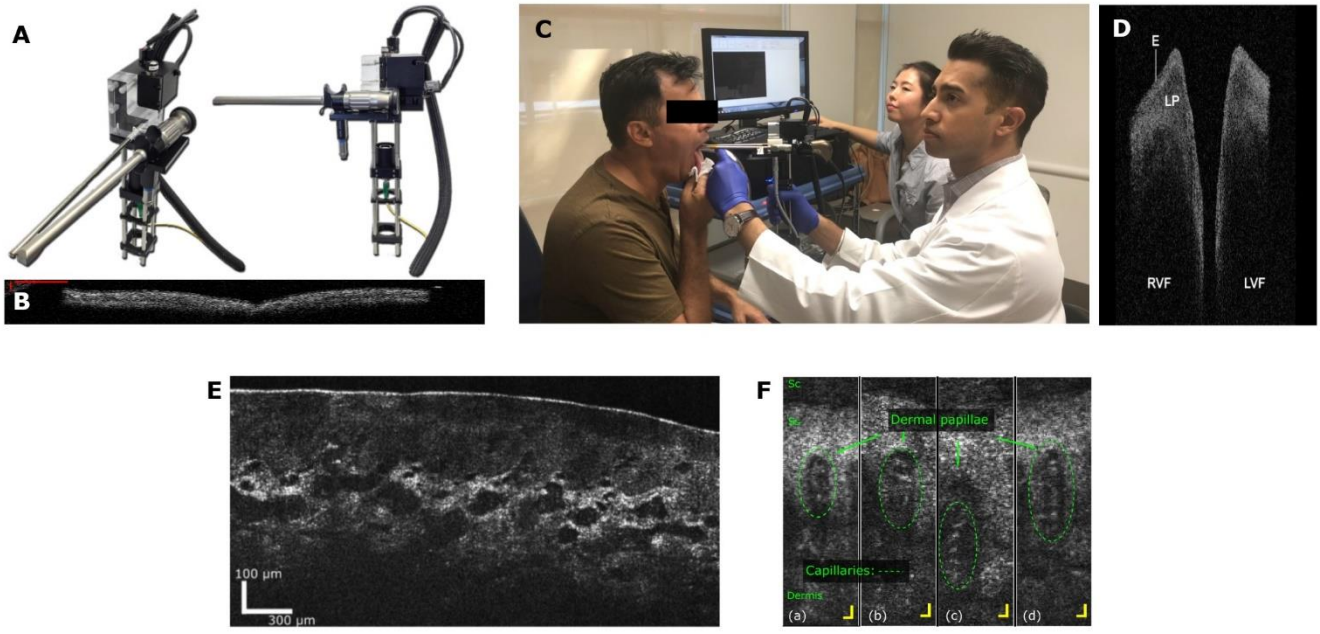
Optisk kohærens tomografi med ultrahøj opløsning (UHR-OCT)

Optisk kohærens tomografi (OCT) er en nyere scanningsmetode end ultralydsscanning. Den bruger lys i stedet for lyd og kan derfor opnå en langt bedre rumlig opløsning [23, 24, 25]. UHR-OCT har en rumlig opløsning på mindre end 5 µm og en dybderækkevidde ned i vævet mellem 0,4 og 1mm. OCT har været anvendt til patienter i narkose, altså uden fonation [8, 9]. Fordelen ved OCT er at man kan differential diagnosticere bedre hvilke tumorer der skal opereres, og hvilke der skal behandles på anden vis. Det nye er at man kan analysere væv under fonation med UHR-OCT således at vi undgår alle risici forbundet ved et invasivt indgreb.

Indtil nu har der ikke kunnet dokumenteres evidensbaseret behandlingseffekt for den subjektive klage: hæshed. På OCT er det muligt at se de cellulære lag i stemmebåndene under fonation inklusive kanternes regelmæssighed, dette er et meget stort fremskridt. Desværre er hastigheden af OCT typisk kun 50-100 tværsnitsbilleder per sekund, hvilket bevirker artefakter [25]. Dette besværliggør en sikker diagnose af stemmebåndenes bevægelser. Præcis det blev også dokumenteret da man begyndte at koordinere stroboskopi med electroglottografi for mange år siden [26].

Der er nu konstrueret en UHR-OCT-opsætning som kan kombineres med HSV under fonation (4.000 billeder pr. sek.) [10]. Den høje OCT-opløsning giver nøjagtig information om cellelag, for væsentlig bedre forståelse af dysfunktioner og slimhindeforandringer i struben og især på stemmebåndene.
DTU Fotonik har indtil nu fremstillet en håndholdt probe der kan afbilde mundslimhinden [10]. En probe til undersøgelse af strubeslimhinden under fonation kan have en form og længde som et laryngoskop. Det er derfor muligt at koble det sammen med laryngoskobet til HSV. Det indeholder en laserpointer til at sigte mod det område som skal afbilledes [25, 27]. En probe til afbildning af stemmebåndene under fonation indeholder en linjescan-procedure (i modsætning til traditionel laser plet scan) kombineret med et 2D kamera spektrometer for at opnå tilstrækkelig hastighed, og er forsynet med en superkontinuum kilde for at opnå en tilstrækkelig dybdeopløsning. Med dette gives således mulighed for direkte diagnostik med UHR-OCT af struben og stemmebåndene under fonation. Biopsier er en invasiv procedure og UHR-OCT ville derfor være at foretrække ved godartede lidelser, da undersøgelsen kan foretages uden anæstesi og med mindst muligt ubehag for patienten. Der kan dog være tolkningsproblemer indtil en standardisering foreligger.
Figur 3 A. Udstyret til OCT på en probe vises med tilladelse fra Beckman Instituttet i Californien. **B**. Et tværsnit af lukkede stemmebånd med OCT. **C.** Opstillingen til OCT i Beckman Instituttet. **D**. Et billede fra en video af OCT med 200 billeder i sekundet.
Nederst ses en *in vivo* optagelse lavet med UHR-OCT systemet fra DTU Fotonik [10]. **E.** Fra mundslimhinde (inderside af underlæbe) med epitel, kirtler og blodkar, **F**. Individuelle hudpapiller og kapillærer i hånden. Disse optagelser illustrerer fordelen ved ultrahøj opløsning [16].

Figur 3 - OCT-opstilling



FIGUR 3 OCT-opstilling **A**. Udstyret til OCT på en probe fra Beckman Instituttet i Californien. **B**. Et tværsnit af lukkede stemmebånd med OCT. **C**. Opstillingen til OCT i Beckman Instituttet. **D**. Et billede fra en film af OCT med 200 billeder i sekundet. (A-D Med tilladelse fra Brian Wong, Beckman Instituttet) **E**. In vivo optagelse UHR OCT fra mundslimhinde (inderside af underlæbe) med epitel, kirtler og blodkar [10, 28]. **F**. Indviduelle hudpapiller og kapillærer i hånden, målestokken svarer til 20 µm [6].

Diskussion og konklusion

Hensigten med denne artikel er at give en introduktion til nogle muligheder for at kombinere HSV med andre metoder til diagnostik af de øvre luftveje, specielt når stemmelidelser har nedsat funktion af stemmebåndene til følge. Det er vedvarende svært at differentiere årsager til patientklager over hæshed tilstrækkeligt [17, 18, 19, 20].

På grund af de store mængder data som HSV indeholder har vi præsenteret nogle nye forskningsresultater hvor de forskellige teknologier og deres muligheder indenfor rammen laryngologi er beskrevet. Vi har diskuteret hvorledes stemmebåndenes afgrænsning og arealet imellem dem ved segmentering kan gøres automatisk, så man i laryngologien kan opnå en højere diagnostisk hastighed, som er mere præcis til differentiering af mange diagnoser.

Vi har præsenteret muligheder for bedre vævsanalyse, der kan give hjælp til løsninger for en eksakt diagnose. Man kan kombinere HSV og UHR-OCT, både mht. stemmebånds bevægelser og slimhindefunktion som en del af øvre luftveje, fordi UHR-OCT har høje frekvenser som svarer til HSV [10]. Der kan suppleres med deep learning. Stroboskopi forårsager artefakter pga. for lav billedfrekvens i forhold til stemmebåndenes bevægelser og er vanskelig at kvantificere [1]. Oftest afbilledes kun hver 4. stemmebåndsbevægelse med et enkelt billede.

Stroboskopi var et skridt fremad, og der forskes i at kombinere stroboskopi og OCT [29]. Næste skridt for forståelsen af strubens funktionen under fonation er at bruge UHR-OCT og deep learning sammen med HSV, med henblik på bedre evidensbaseret kvantitativ analyse.

Litteratur

30. Woo P. Stroboscopy and High-Speed Imaging of the Vocal Function. Plural publishing 2nd ed. 2021.
31. Pedersen M, Eeg M, Jønsson A et al. Chapter 8, Working with Wolf Ltd. HRES 5562 Analytic system for high-speed recordings, Normal & Abnormal Vocal folds Kinematics: HSDP, OCT & NBI, 2015;1:57-65.
32. DTU fotonik. ShapeOCT (2015-2019), med 25,7 mio. fra Innovationsfonden, Grant/Award Number: 4107-00011A.
33. Ran AR, Tham CC, Chan PP et al. Deep learning in glaucoma with optical coherence tomography: A review. Eye. 2021 Jan;35(1):188-201.
34. Del Amor R, Morales S, Colomer A et al. Automatic segmentation of epidermis and hair follicles in optical coherence tomography images of normal skin by convolutional neural networks. Frontiers in Medicine. 2020 Jun 4;7:220.
35. Israelsen NM, Maria M, Mogensen M et al. The value of ultrahigh resolution OCT in dermatology - delineating the dermo-epidermal junction, capillaries in the dermal papillae and vellus hairs, Biomedical Optics Express, 2018; 9(5). 2240-2265.
36. Israelsen NM, Mogensen M, Jensen M et al. Delineating papillary dermis around basal cell carcinomas by high and ultrahigh resolution optical coherence tomography-A pilot study. J Biophotonics. 2021 Jul 10:e202100083.
37. Brian J Wong. In Vivo Optical Coherence Tomography of the Human Larynx: Normative and Benign Pathology in 82 Patients. Laryngoscope. 2005;115(11):1904-11.
38. Klein AM, Pierce MC, Zeitels SM et al. Imaging the Human Vocal Folds in Vivo with Optical Coherence Tomography: A Preliminary Experience. Annals of Otology, Rhinology & Laryngology. 2006;115(4):277-284.
39. Israelsen NM, Jensen M, Jønsson AO et al. Ultrahigh Resolution Optical Coherence Tomography for Detecting Tissue Abnormalities of the Oral and Laryngeal Mucosa: A Preliminary Study, MAVEBA Proceedings, 2016; 195-197.
40. Fehling MK, Grosch F, Schuster ME et al. Fully automatic segmentation of glottis and vocal folds in endoscopic laryngeal high-speed videos using a deep Convolutional LSTM Network. PLoS One. 2020;15(2):1–29.

41. Kist AM, Gómez P, Dubrovskiy D et al. A Deep Learning Enhanced Novel Software Tool for Laryngeal Dynamics Analysis. J. Speech, Language and Hearing Research. 2021;64(6):1889-1903.
42. Yousef AM, Deliyski DD, Zacharias SRC et al. A Hybrid Machine-Learning-Based Method for Analytic Representation of the Vocal Fold Edges during Connected Speech. Applied Sciences. 2021; 11(3):1179.
43. Watanabe T, Kaneko K, Sakaguchi K, Takahashi H. Vocal-fold vibration of patients with Reinke's edema observed using high-speed digital imaging. Auris Nasus Larynx. 2016;43(6):654-657
44. Garcia JA, Benboujja F, Beaudette K et al. Using attenuation coefficients from optical coherence tomography as markers of vocal fold maturation. Laryngoscope. 2016; 126(6): E218–23.
45. Pedersen M, Agersted A, Akram B et al. Optical coherence tomography in the laryngeal arytenoid mucosa for documentation of pharmacological treatments and genetic aspects: a protocol, Advances in Cellular and Molecular Otolaryngology, 2016; 4:1.
46. Roth DF, Abbott KV, Carroll TL et al. Evidence for primary laryngeal inhalant allergy: a randomized, double-blinded crossover study. International forum of allergy & rhinology. 2013;(1):10-8.
47. am Zehnhoff-Dinnesen A, Wiskirska-Woznica B, Neumann K et al. Phoniatrics I. Springer Berlin Heidelberg 2020.
48. Pedersen M, McGlashan J. Surgical versus non-surgical interventions for vocal cord nodules (Review), The Cochrane Library. 2012; 1-13.
49. Pedersen M. Which Mathematical and Physiological Formulas are Describing Voice Pathology: An Overview, Journal of General Practice, 2016; 4:3.
50. Woisard V: Gastro-esopharyngeal Reflux Influences on Larynx and Voice, page 263-271. In Phoniatrics I. am Zehnhoff-Dinnesen A, Wiskirska-Woznica B, Neumann K, Nawka T, editors Springer Berlin Heidelberg 2020.
51. Pham TT, Chen L, Heidari AE et al. Computational analysis of six optical coherence tomography systems for vocal fold imaging: A comparison study. Lasers in surgery and medicine 2019; 51:412-422.
52. Sergeev AM, Gelikonov GV, Gelikonov FI et al. In vivo endoscopic OCT imaging of precancer and cancer states of human mucosa. Optics express, 1997; 1,13: 434-440.
53. Just T, Guder E, Witt G et al. Confocal Endomicroscopy and Optical Coherence Tomography for Differentiation Between Low-Grade and High-Grade Lesions of the

Larynx in Biomedical Optics in Otorhinolaryngology: Head and Neck surgery. Eds. Springer New York, 2016; 479-490.
54. Coughlan CA, Chou L, Jing JC et al. In vivo cross-sectional imaging of the phonating larynx using long-range Doppler optical coherence tomography. Nature Science Reports. 2016; 6: 22792.
55. Pedersen MF. Electroglottography compared with synchronized stroboscopy in normal persons. Folia phoniatr; 1977; 29:191-200.
56. Donner S, Bleeker S, Ripken T et al. Automated working distance adjustment enables optical coherence tomography of the human larynx in awake patients, J Med imaging (Bellingham, Wash.) 2015; 2.2, 026003.
57. Wei W, Choi WJ, Men S et al. Wide-field and long-ranging-depth optical coherence tomography microangiography of human oral mucosa (Conference Presentation), Proceedings SPIE 10473, Lasers in Dentistry XXIV, 2018, 104730H
58. Maguluri GN, Mehta DD, Kobler JB et al. Optical biopsy of vocal folds during phonation using parallel OCT (Conference Presentation). In: Alfano RR, Demos SG, Seddon AB, editors. Optical Biopsy XVII: Toward Real-Time Spectroscopic Imaging and Diagnosis. SPIE; 2019; 13.

Summary (engelsk)

Title: **Quantitative examination of vocal folds, aspects of image analysis and OCT with Ultra-high resolution**

To directly relate tissue abnormalities to dysfunctional voicing, it is decisive to temporally resolve the vocal fold movement during phonation; and that on the microscopic level. High-speed video (HSV) can record the vocal folds with 2-4.000 fps. Ultrahigh resolution optical coherence tomography (UHR-OCT) can distinguish cellular layers with a resolution better than 5 µm within a tissue depth of 1 mm. We propose combining the two technologies and apply deep learning-based image segmentation to establish statistical evident and reproducible documentation for voice related diseases.

Hovedbudskaber – tre korte sætninger:

- Vi har værktøjer til kvantitativ analyse.
- OCT bliver brugt i laryngologien under narkose, men der er mulighed for at udføre OCT under fonation.

- Vi har mulighed for at kombinere high-speed video med UHR-OCT in vivo, under fonation, til bedre diagnostik og behandlings dokumentation.