

## Statusartikel

Ugeskr Læger 2022;184:V02210146

# Kvantitativ undersøgelse af stemmebånd med højhastighedsvideo og optisk kohærens-tomografi

Niels Israelsen<sup>1</sup>, Christian F. Larsen<sup>2</sup> & Mette Pedersen<sup>3</sup>

1) DTU Fotoniks, 2) Copenhagen Business School, 3) Medical Centre, Østergade 18, København

Ugeskr Læger 2022;184:V02210146

### HOVEDBUDSKABER

- Vi har værktøjer til kvantitativ analyse.
- Optisk kohærens-tomografi (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 OCT i ultrahøj opløsning in vivo, under fonation, til bedre diagnostik og behandlingsdokumentation.

Den menneskelige lyd dannelse 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]. Vi vil beskrive anvendelse af optisk kohærens-tomografi med ultrahøj opløsning (UHR-OCT) og deep learning (en gren inden for kunstig intelligens (AI)) til kvantitativ analyse af stemmebånd på basis af high-speed video (HSV).

Perspektivet er at koordinere metoderne, delvist i samarbejde med DTU Fotonik. 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 HSV af stemmebåndene og AI. Hermed kan der opnås store muligheder for diagnostik svarende til lignende forhold for øjenlidelser [4] og hudlidelser, hvor der er stor interesse for vævsforståelse [5-7].

De praktiserende læger har interesse i at vide, hvor vi står mht. diagnosemuligheder for stemmelidelser, nu hvor AI 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 differentialdiagnostik. 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 HSV og UHR-OCT både mht. stemmebåndsbevægelser og slimhindelidelser som en del af de 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

Muligheden for en eksakt diagnose for slimhindeforandringer i struben med ændret fonation til følge er forbedret væsentligt med HSV [1, 2]. Med dette apparatur kan der bl.a. analyseres en slowmotionvideo med 2.000-4.000 billeder af stemmebåndenes eksakte bevægelser pr. sekund. Stroboskopi har vundet udbredt anvendelse i laryngologien, men optager kun 25 billeder/s ( $\approx$  Hz), og derfor er HSV en stor forbedring, da gennemsnitstalefrekvenser for hhv. mænd og kvinder er på 110 Hz og 220 Hz. 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.

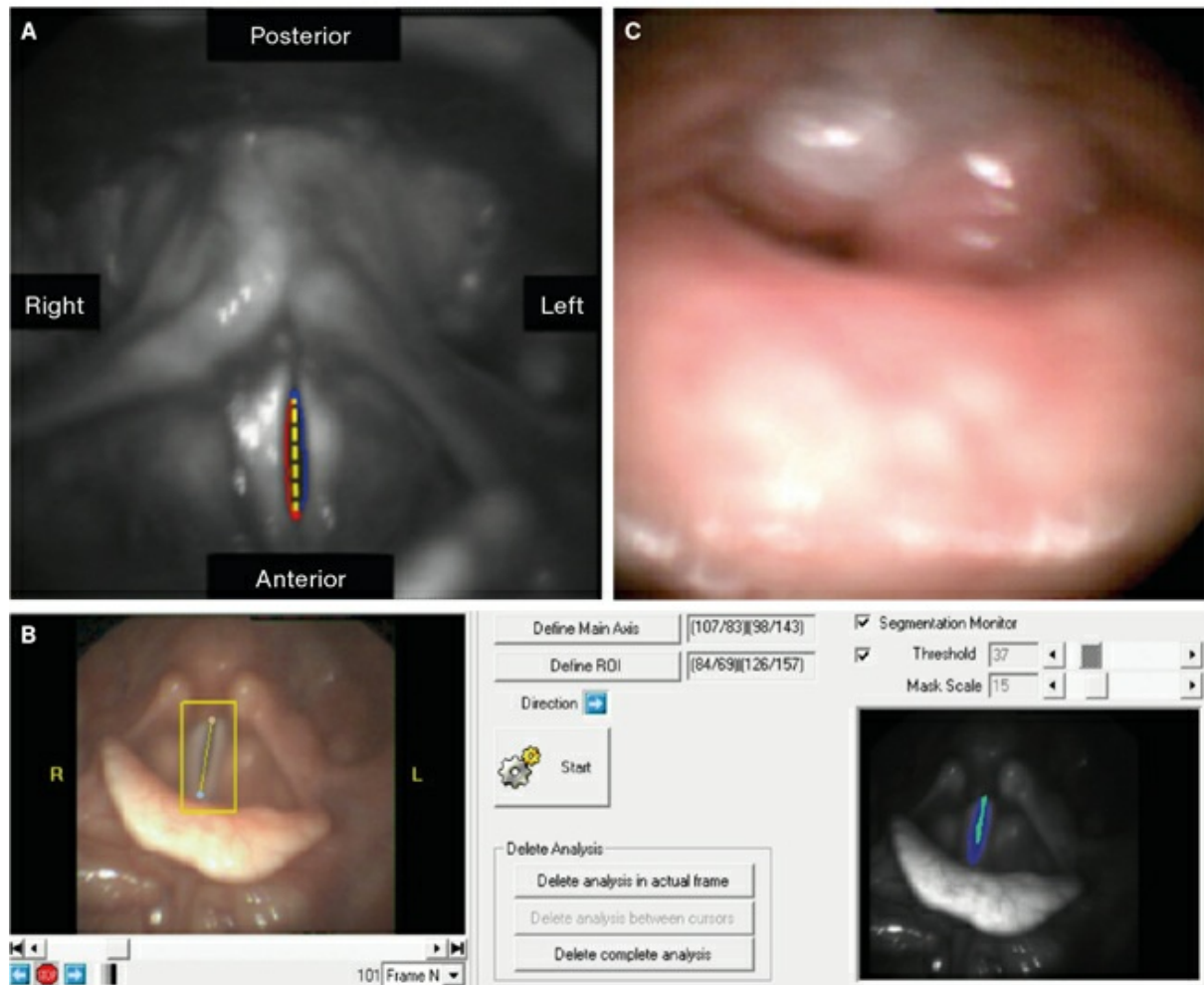
Vi ser med HSV ofte benigne lidelser med mere eller mindre tydeligt ødem af stemmebåndene [14]. Det kan være vanskeligt med HSV at give patienterne en sufficient diagnose, især til dem, der er afhængige af stemmebrug. Et andet interessant aspekt, der heller ikke kan forklares ved hjælp af HSV, er forståelse af hormoners indflydelse på vævet ved stemmens udvikling i puberteten [15].

HSV kan bruges til at optimere differentialdiagnostik af tumorer – benigne og maligne, arvæv, blødninger, traumer og lignende – fordi man har et retvisende billede af stemmebåndenes svingninger [16-18].

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 og 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, mhp. kvantitative beregninger. Disse visualiseringer har ikke været tilstrækkelige til f.eks. at forklare og opnå statistiske kendetegn og evidens [20]. Figur 1 C viser et billede fra en HSV (4.000 billeder/s) hos en patient, hvor slim kommer op fra mavesækken og irriterer struben. Dette forekommer på mindre end 0,2 s [21].

**FIGUR 1** **A.** En normal strube med markeringer af midten af stemmeridsen. **B.** Softwaremålinger, der markerer de frie kanter af stemmebåndene med high-speed video (HSV). **C.** Et billede fra en HSV, hvor slim kommer op i struben og forsvinder igen på 0,2 s.

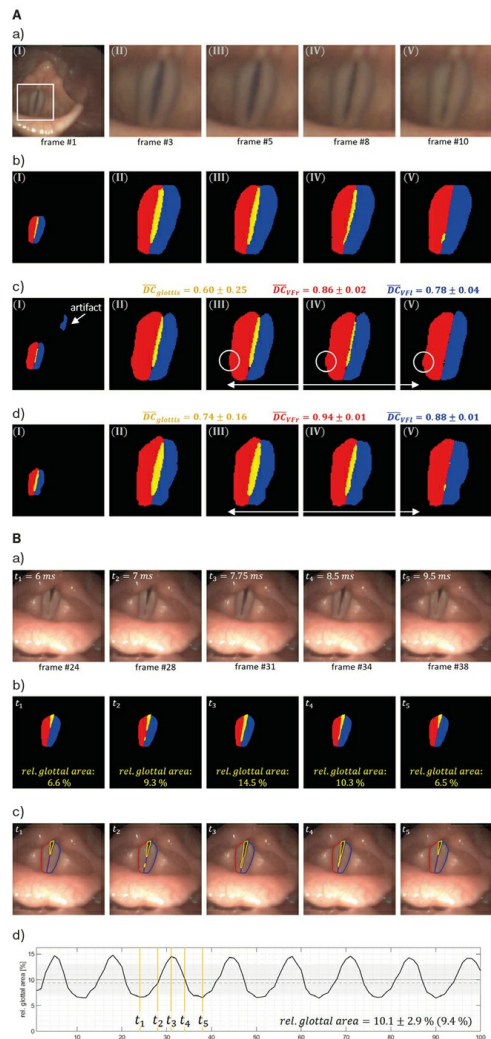


HSV = high-speed video

## DEEP LEARNING

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 den indeholder et feedbackloop, som gør den selvlærende, og dermed øges præcisionen [22]. *Fehling et al* [11] gennemgår de forskellige muligheder for deep learning og konkluderer, at U-long short-term memory (U-LSTM)-segmenteringstypen er bedst til analyse af stemmebånd (Figur 2).

**FIGUR 2 A.** En sammenligning af segmenteringsresultater af individuelle billeder fra en normal person fra vores database. a) HSV, b) ground truth, c) U-Net segmentering, og d) U-LSTM-segmentering. Dicekoefficienten for hver klasse repræsenterer gennemsnit og standarddeviationen for hele sekvensen (100 billeder). **B.** Segmenteringsresultatet for en HSV af en enkelt svingningscyklus fra en patient med insuffisient lukkefunktion bagtil fra vores database, a) HSV, b) neuralt netværks segmentering genereret med U-LSTM<sup>SE</sup>, c) overlægning af segmenteringsresultater, d) gennemsnit og standardafvigelse 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.



HVS = high-speed video; LSTM = long short-term memory  
 Analyse med deep learning givet med tilladelse fra Mona Fehling,  
 University of Applied Sciences, Trier

I Figur 2 vises, hvordan deep learning kan bruges til kvantitativt at differentialdiagnosticere 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. Det, 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 < 1 min for deres neurale netværk. Dette vil spare store mængder tid i den daglige behandling af patienter for øre-næse-

hals-læger og især øge præcisionen til diagnostik af benigne og maligne tumorer/leukoplakier, sulcus og mange andre stemmelidelser.

I Figur 2 vises eksempler på den kliniske betydning af deep learning, som er baseret på HSV og analyseret af *Mona Fehling*. Figur 2A: a) billeder fra en HSV fra en normal person fra vores egen database, b) ground truth er specialisters manuelle segmentering af stemmebåndene og bliver brugt til at måle, hvor præcist et givent neural netværk (deep learning) er til at segmentere, c og d) segmenteringsresultater for to neurale netværk, hvor U-LSTM er mere præcist [11]. Figur 2B: a) HSV og dele af en enkeltcyklus hos en patient med insufficient lukkefunktion bagtil, b) det valgte neurale netværk (U-LSTM) bruges til at estimere venstre og højre stemmebånd, og arealet imellem dem, c) estimeringen af det neurale netværk lægges oven på det originale billede til sammenligning og d) det relative areal mellem stemmebåndene på en kurve givet i tid (ud af 100 billeder pr. 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].

## OPTISK KOHÆRENS-TOMOGRAFI MED ULTRAHØJ OPLØSNING

OCT er en nyere skanningsmetode end UL-skanning. Den bruger lys i stedet for lyd, og der kan derfor opnås en langt bedre rumlig opløsning [23-25]. UHR-OCT har en rumlig opløsning på  $< 5 \mu\text{m}$  og en dybderækkevidde ned i vævet på 0,4-1 mm. OCT har været anvendt til patienter i narkose, altså uden fonation [8, 9]. Fordelen ved OCT er, at man bedre kan differentialdiagnosticere, 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 man undgår alle risici, der er forbundet ved et invasivt indgreb.

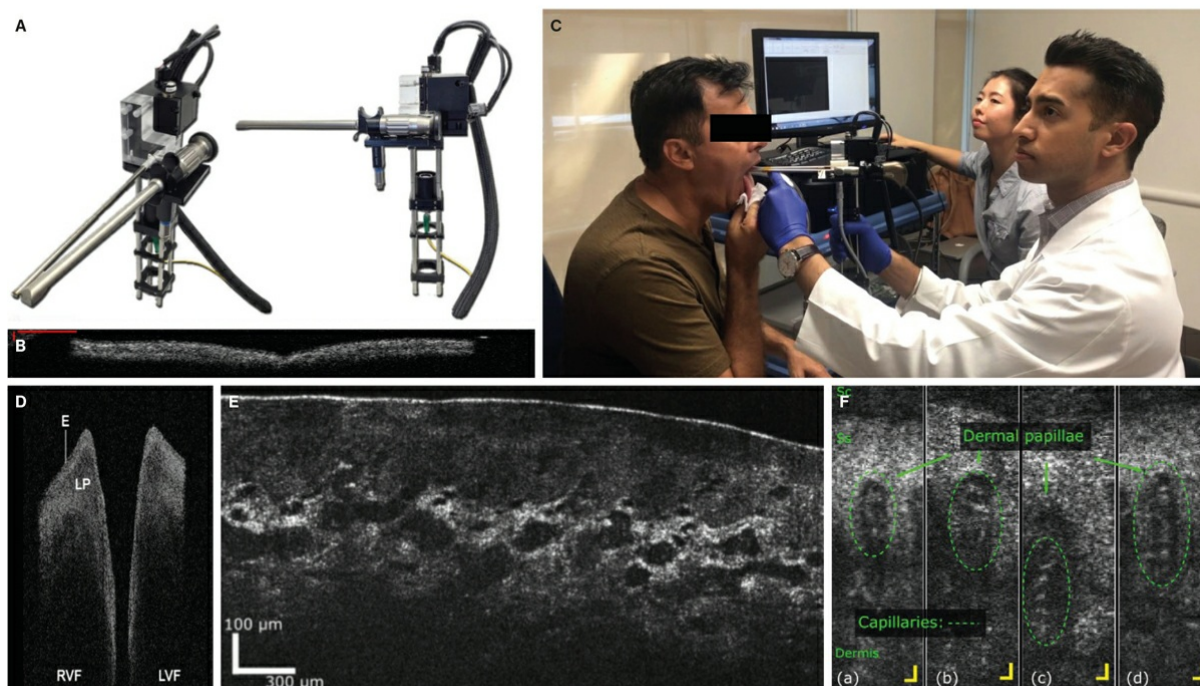
Indtil nu har der ikke kunnet dokumenteres evidensbaseret behandlingseffekt for den subjektive gene: hæshed. På OCT er det muligt at se de cellulære lag i stemmebåndene under fonation inklusive kanternes regelmæssighed, og dette er et meget stort fremskridt. Desværre er hastigheden af OCT typisk kun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

Der er nu konstrueret en UHR-OCT-opsætning, som kan kombineres med HSV under fonation (4.000 billeder/s) [10]. Den høje OCT-opløsning giver nøjagtig information om celledag, hvilket medfører 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 strubslimhinden under fonation kan have en form og længde som et laryngoskop. Det er derfor muligt at koble det sammen med laryngoskopet til HSV. Det indeholder en laserpointer til at sigte mod det område, som skal afbildes [25, 27]. En probe til afbildning af stemmebåndene under fonation indeholder en linjeskanningsprocedure (i modsætning til traditionel laserpletskanning) kombineret med et todimensionalt kameraspektrometer for at opnå tilstrækkelig hastighed og er forsynet med en superkontinuumkilde 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.

I **Figur 3** ses en OCT-opstilling samt billeder fra en film af OCT med 200 billeder/s. Disse optagelser illustrerer fordelene ved ultrahøj opløsning [16].

**FIGUR 3** Optisk kohærens-tomografi (OCT)-opstilling **A.** Udstyret til OCT på en probe fra Beckman Institutet i Californien. **B.** Et tværsnit af lukkede stemmebånd med OCT. **C.** Opstillingen til OCT i Beckman Institute. **D.** Et billede fra en film af OCT med 200 billeder/s. **E.** En in vivo-optagelse af UHR-OCT fra en mundslimhinde (underside af underlæbe) med epitel, kirtler og blodkar [10, 28]. **F.** Individuelle hudpapiller og kapillærer i hånden, målestokken svarer til 20  $\mu\text{m}$  [6]. Disse optagelser illustrerer fordelene ved ultrahøj opløsning [16].



A-D bringes med tilladelse fra Brian Wong, Beckman Institute, Californien.

## 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 [17-20].

Pga. de store mængder data, som HSV indeholder, har vi præsenteret nogle nye forskningsresultater, hvor de forskellige teknologier og deres muligheder inden for 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. stemmebåndsbevægelser og slimhinfunktion som en del af de 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 afbildes kun hver fjerde stemmebåndsbevægelse med et enkelt billede. Stroboskopi var et skridt fremad, og der forsøges 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 mhp. bedre evidensbaseret kvantitativ analyse.

**Korrespondance** *Mette Pedersen*. E-mail: m.f.pedersen@dadlnet.dk

**Antaget** 5. november 2021

**Publiceret på ugeskriftet.dk** 7. februar 2022

**Interessekonflikter** ingen. Forfatterens ICMJE-formularer er tilgængelige sammen med artiklen på ugeskriftet.dk

**Referencer** findes i artiklen publiceret på ugeskriftet.dk

**Artikelreference** Ugeskr Læger 2022;184:V02210146

## SUMMARY

### Quantitative examination of vocal cords with high-speed video and optical coherence tomography

Niels Israelsen, Christian F. Larsen & Mette Pedersen

Ugeskr Læger 2022;184:V02210146

To directly relate tissue abnormalities to dysfunctional voicing, it is decisive to temporally resolve the vocal fold movement during phonation on the microscopic level. High-speed video (HSV) can record the vocal folds with 2,000-4,000 fps. Ultra-high resolution optical coherence tomography can distinguish cellular layers with a resolution better than 5  $\mu\text{m}$  within a tissue depth of 1 mm. In this review, we propose combining the two technologies and apply deep learning-based image segmentation to establish statistical evident and reproducible documentation for voice-related diseases.

## REFERENCER

1. Woo P. Stroboscopy and high-speed imaging of the vocal function 2. udg. Plural Publishing, 2021.
2. Pedersen M, Eeg M, Jønsson A, Mamood S. Chapter 8, Working with Wolf Ltd. HRES 5562 analytic system for high-speed recordings. I: Izdebski K, Yan Y, Ward RR et al. Normal and abnormal vocal folds kinematics: HSDP, OCT & NBI. CreateSpace Independent Publishing Platform, 2015:57-65.
3. Projektbeskrivelse af ShapeOCT (2015-2019), med 25,7 mio. fra Innovationsfonden. <https://www.dtu.dk/om-dtu/nyheder-og-presse/dynamo1/2018/06/tema-5-laserteknologi-giver-detajlerede-billeder-af-hudkraeft?id=6aaa51f4-bffa-4f3b-a847-9f4387c65eea> (12. jan 2022).
4. Ran AR, Tham CC, Chan PP et al. Deep learning in glaucoma with optical coherence tomography: a review. *Eye*. 2021;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. *Front Med (Lausanne)* 2020;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. *Biomed Opt Express*. 2018;9(5):2240-65.
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;14(11):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. *Ann Otol Rhinol Laryngol*. 2006;115(4):277-84.
10. Israelsen NM, Jensen M, Jønsson AO, Pedersen M. Ultrahigh resolution optical coherence tomography for detecting tissue abnormalities of the oral and laryngeal mucosa: a preliminary study. I: Proceedings of 11th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications, 2016:195-7.
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):e0227791.
12. Kist AM, Gómez P, Dubrovskiy D et al. A deep learning enhanced novel software tool for laryngeal dynamics analysis. *J*.

- Speech Lang Hear Res. 2021;64(6):1889-903.
13. Yousef AM, Deliyiski DD, Zacharias SRC et al. A hybrid machine-learning-based method for analytic representation of the vocal fold edges during connected speech. *AppSci (Basel)* 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):654-7.
  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. *Int Forum Allergy Rhinol*. 2013;3(1):10-8.
  18. Eysholdt U. Laryngoscopy, stroboscopy, high-speed video and phonovibrogram., page 364-376. I: am Zehnhoff-Dinnesen A, Wiskirska-Woznica B, Neumann K, Nawka T, red. *Phoniatics I*. Springer, 2020:364-76.
  19. Pedersen M, McGlashan J. Surgical versus non-surgical interventions for vocal cord nodules. *Cochrane Database Syst Rev*. 2012;(6)CD001934.
  20. Pedersen M. Which mathematical and physiological formulas are describing voice pathology: an overview. *J Gen Pract*. 2016;4:3.
  21. Woisard V. Gastro-esopharyngeal reflux influences on larynx and voice., page 263-271. I: am Zehnhoff-Dinnesen A, Wiskirska-Woznica B, Neumann K, Nawka T, red. *Phoniatics I*. Springer, 2020:263-71.
  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 Surg Med*. 2019;51:412-22.
  23. Sergeev AM, Gelikonov GV, Gelikonov FI et al. In vivo endoscopic OCT imaging of precancer and cancer states of human mucosa. *Opt Express*. 1997;1(13):434-40.
  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. I: Wong BJP, Ilgner J, red. *Biomedical optics in otorhinolaryngology: head and neck surgery*. Springer, 2016:479-90.
  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. *Sci Rep*. 2016;6:22792.
  26. Pedersen MF. Electroglottography compared with synchronized stroboscopy in normal persons. *Folia Phoniatr*. 1977;29(3):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)* 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). *Lasers in Dentistry XXIV, 2018, Proceedings SPIE 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, red. *Optical Biopsy XVII: toward real-time spectroscopic imaging and diagnosis*. SPIE 2019;10873.