

# Kvantitativ Undersøgelse af Stemmebånd

Høj-Hastigheds Video Film, AI og  
UHR-Optisk Kohærens-Tomografi

På basis af en status artikel i Ugeskrift for Læger



Niels Israelsen, Christian Frederik Larsen, Mette Pedersen

25 Marts 2023 Dansk skuespillerforbund, Tagensvej 85, Nørrebro, København

Hvad vi har

# Billeddannelse af stemmebånd

Fuldregister



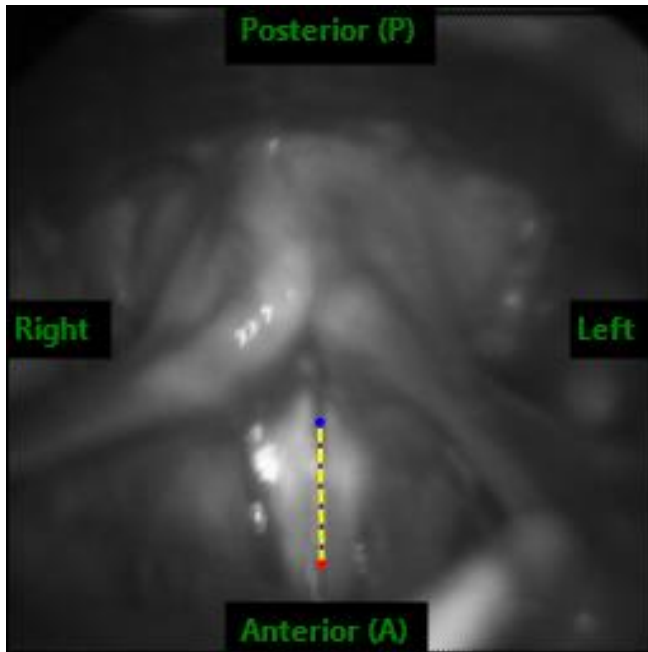
Vores kommercielle opsætning med 256 x 256 pixels, optaget med 4.000 billeder per sekund.

Fra Richard Wolf GmbH i Knittlingen, Tyskland (Endocam 5562), højhastigheds kameraet er monteret på et stift endoskop.

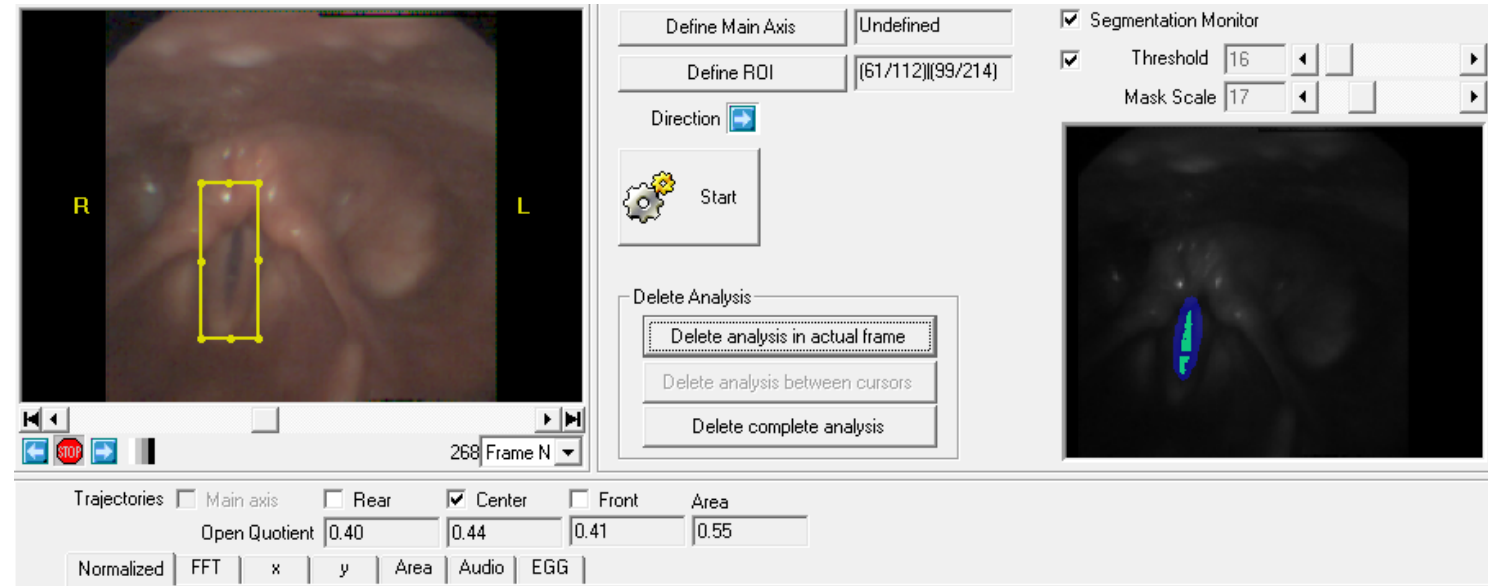
Randregister



# Analyse af stemmebåndene



Viser en normal strube med manuel markering af midten af stemmeridsen



Region af interesse og markeringer af kanterne på stemmebåndene med HSV.

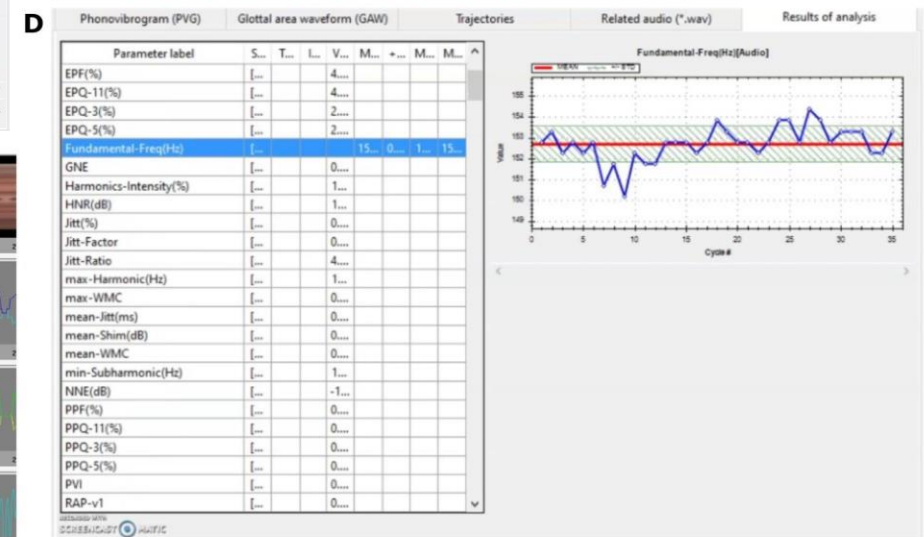
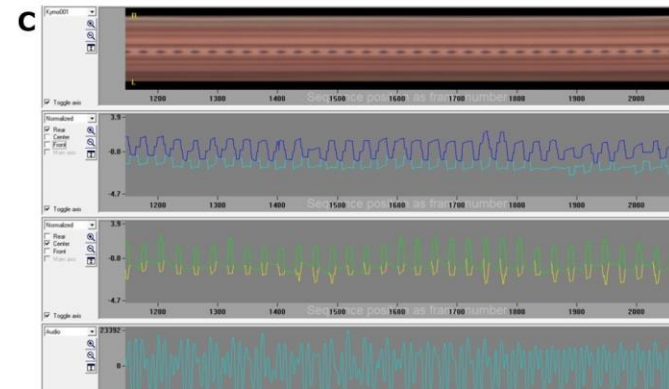
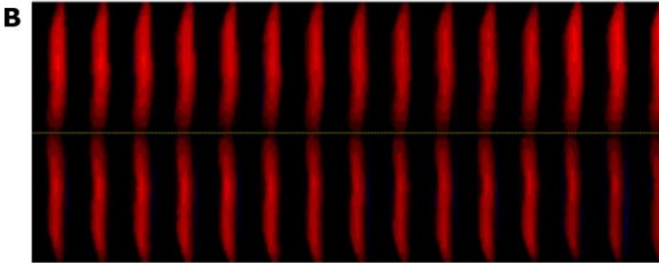
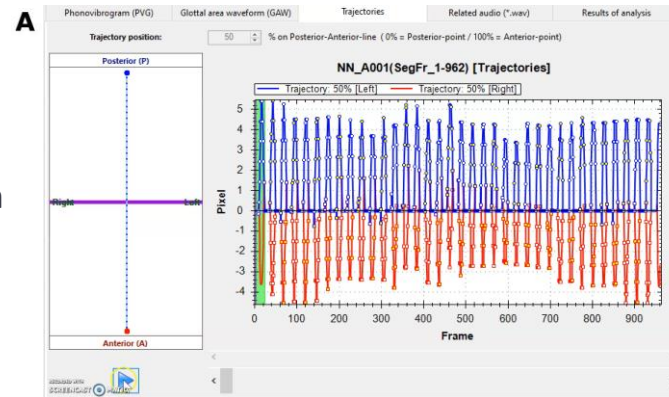
# Kvantitativ analyse af stemmebånd

(A) En High-Speed Video fra en software reproduktion (GlottalAnalysisTools)

(B) Phonovibrogram

(C) Den øverste kurve er et kymogram, dernæst ses lukning af højre og venstre stemmebånd bagtil i tid, herunder ses lukning af stemme båndene i midten, og nederst ses det korresponderende akustiske billede.

(D) 183 målingerne.



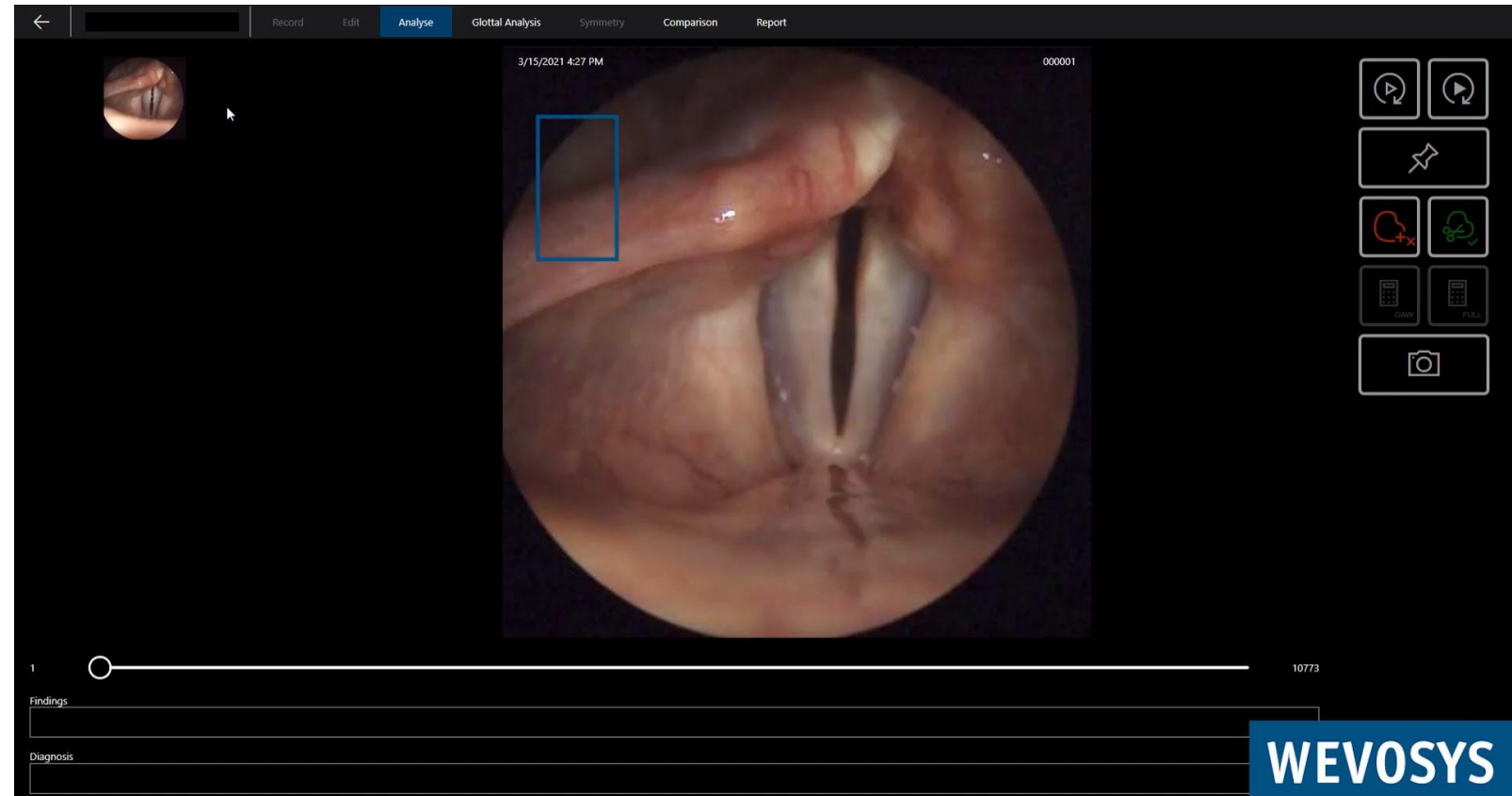
Hvad vi gerne vil have

# Billeddannelse af stemmebånd

LingWAVES 4 HSV er et kommercielt højhastigheds videoendoskopi fra Wevosys.

1.440 x 1.024 pixel – i farve.  
4.000 billeder i sekundet. Op til 8.000 billeder i sekundet er muligt.

Det interessante område er hele stemmeridsen.



# Optisk kohærens-tomografi udstyr med skoper til HSV og OCT

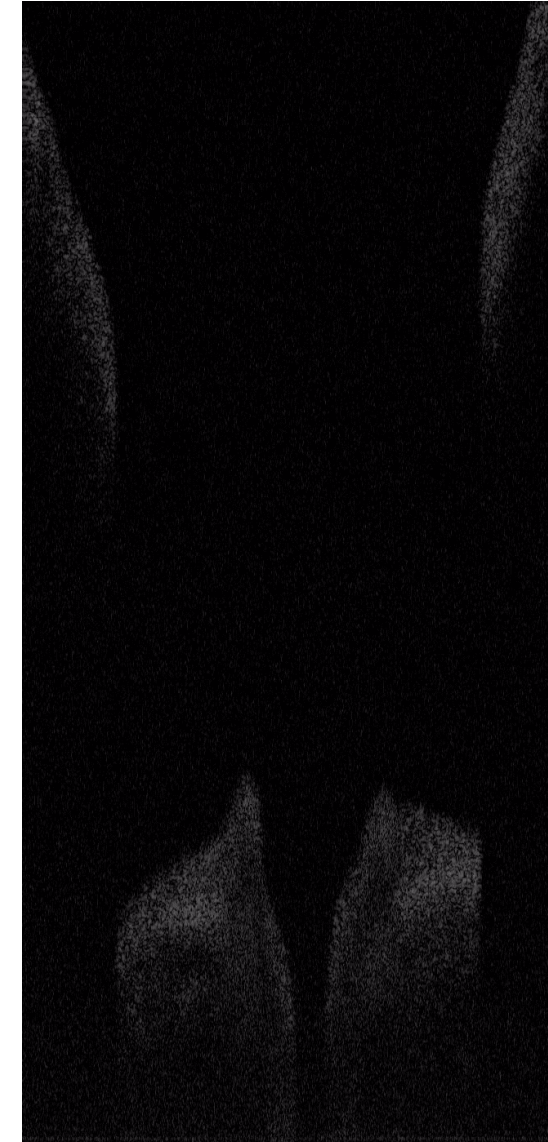


Optisk kohærens-tomografi med opløsning, 250 billeder pr sekund – forårsager artefakter (en slags forvrængning).

Slimhindebevægelse (jævn/ikke jævn)

Enkeltlag af stemmebånds celler (normal eller med hævelse, ændringer ect.,)

Kanten på stemmebåndene.



Med tilladelse fra Brian Wong og Lily Chen

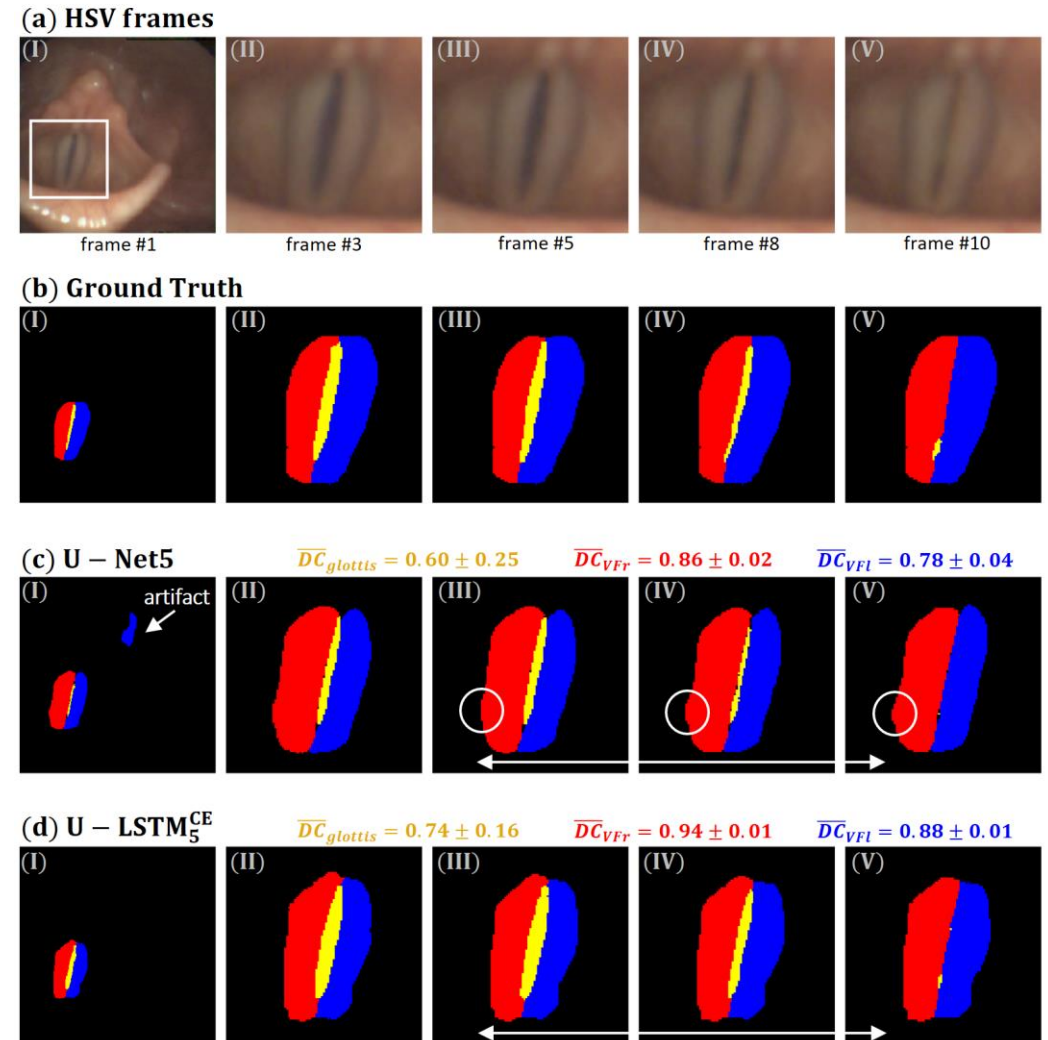


Hvad vi gerne vil bruge I klinikken fremover

# Deep learning software

Deep learning er nødvendig for at håndtere store mængder data.

- a) Højhastigheds video, 4 ud af 100
  - b) Ground Truth,
  - c) U-Net segmentering,
  - d) U-LSTM segmentering.
- 
- Dice koefficienten for hver klasse repræsenterer gennemsnittet og standardafvigelsen for hele sekvensen af 100 frames.



# Deep learning software til analysis af arealet mellem stemmebåndene

Billedet viser en højhastighedsvideo af en enkelt svingning fra en patient med lukkedeficit bagerst fra vores database.

(a) Højhastigheds video

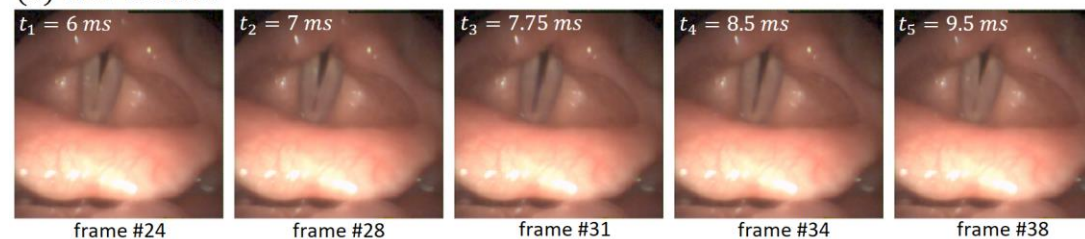
(b) U-LSTM\_5^ACE, med beregning af arealet for stemmeridsen.

(c) Overlay af segmenteringsresultaterne

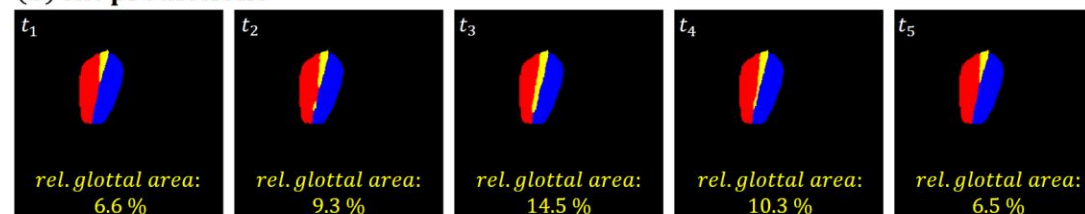
(d) Gennemsnittet og standardafvigelsen for det normaliserede relative område mellem stemmebåndene for hele sekvensen, svarende til summen af området mellem stemmebåndene + højre stemmebånd + venstre stemmebånd.

De gule linjer angiver, hvor billederne er taget i svingningen.

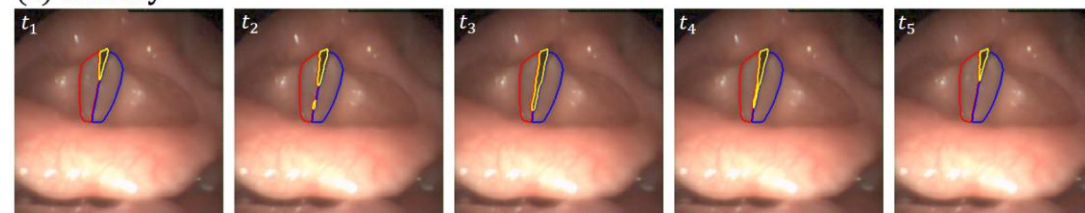
(a) HSV frames



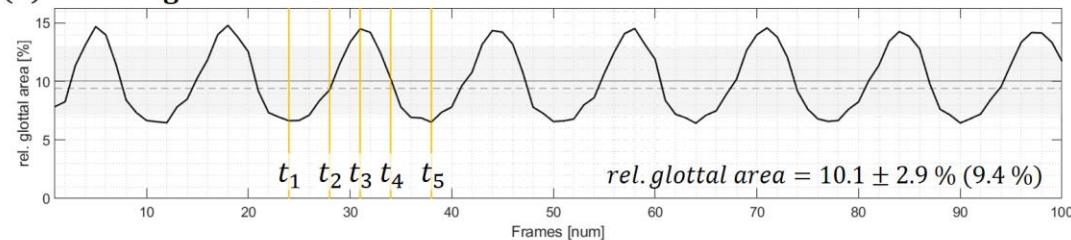
(b) NN predictions



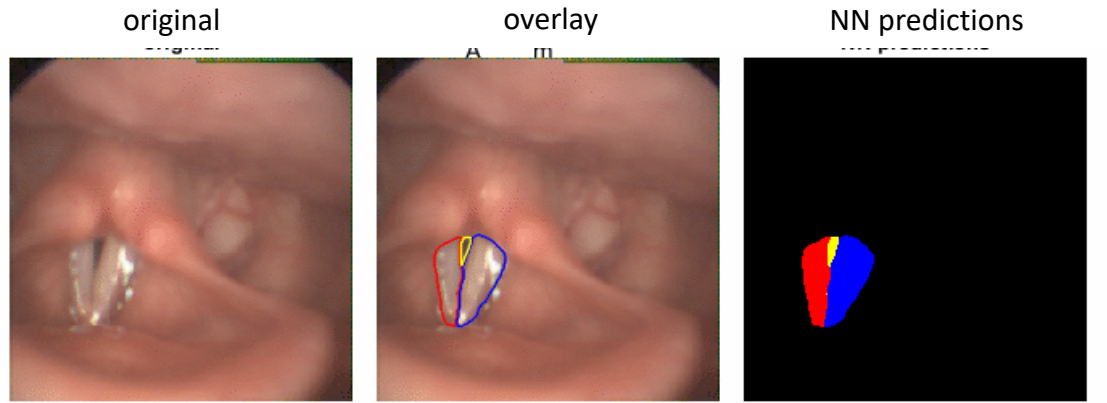
(c) overlay



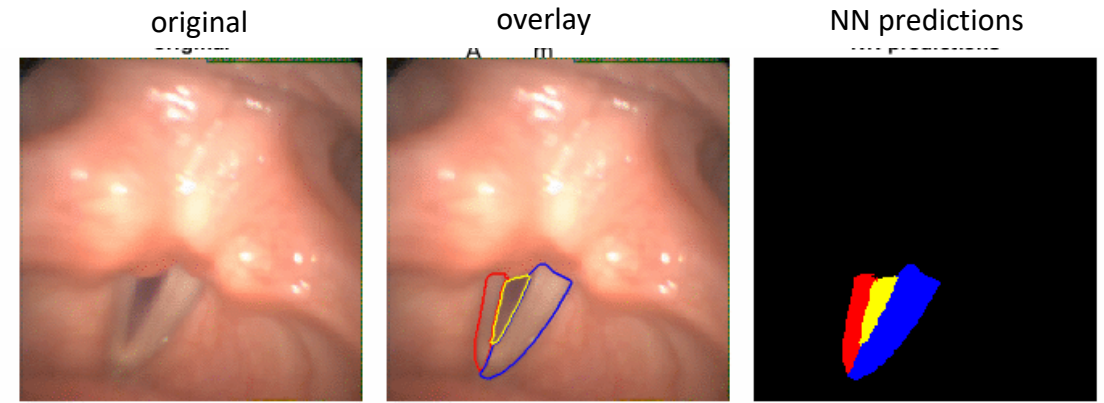
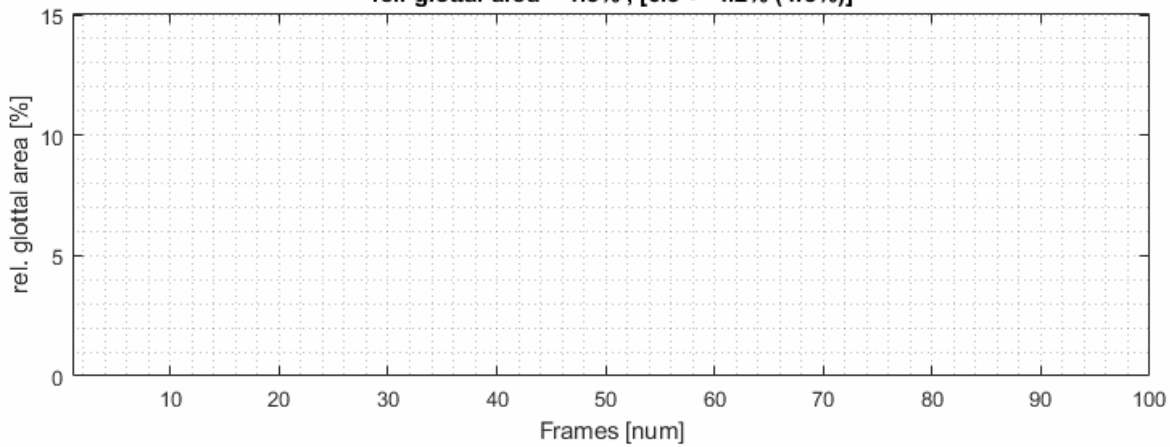
(d) relative glottal area



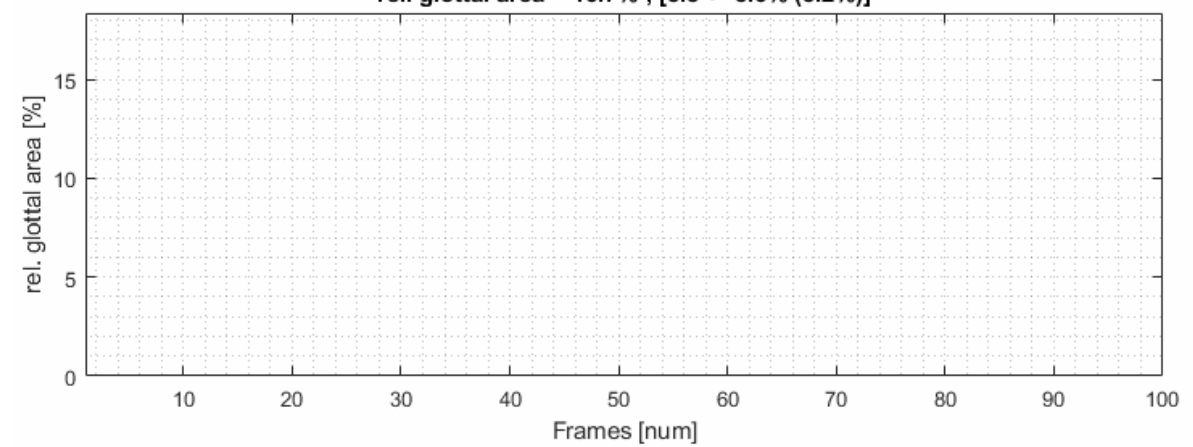
# Deep learning software til analysis af arealet mellem stemmebåndene



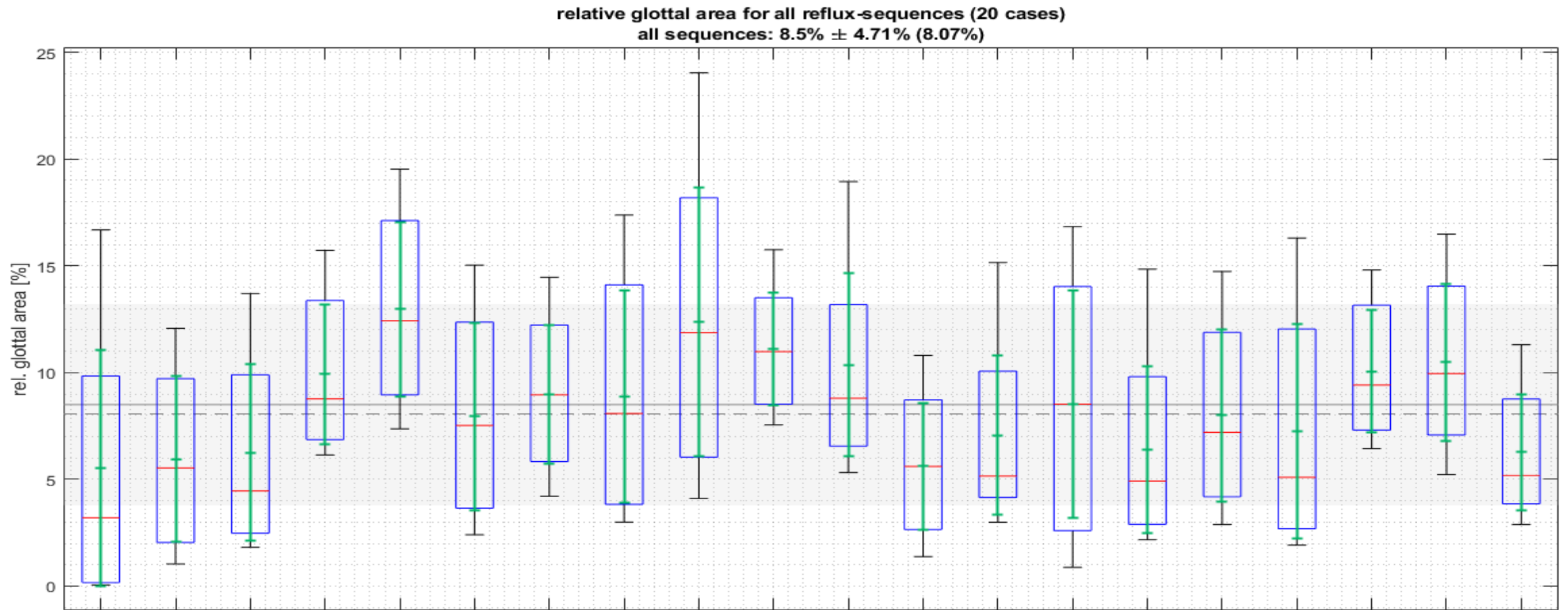
rel. glottal area = 4.3% , [6.3 +- 4.2% (4.5%)]



rel. glottal area = 16.7% , [5.5 +- 5.6% (3.2%)]



# Analyse af arealet mellem stemmebåndene



20 videoer af patienter med reflux, hvor hele stemmeridsens bevægelse med hensyn til lukke er blevet analyseret med Matlab.

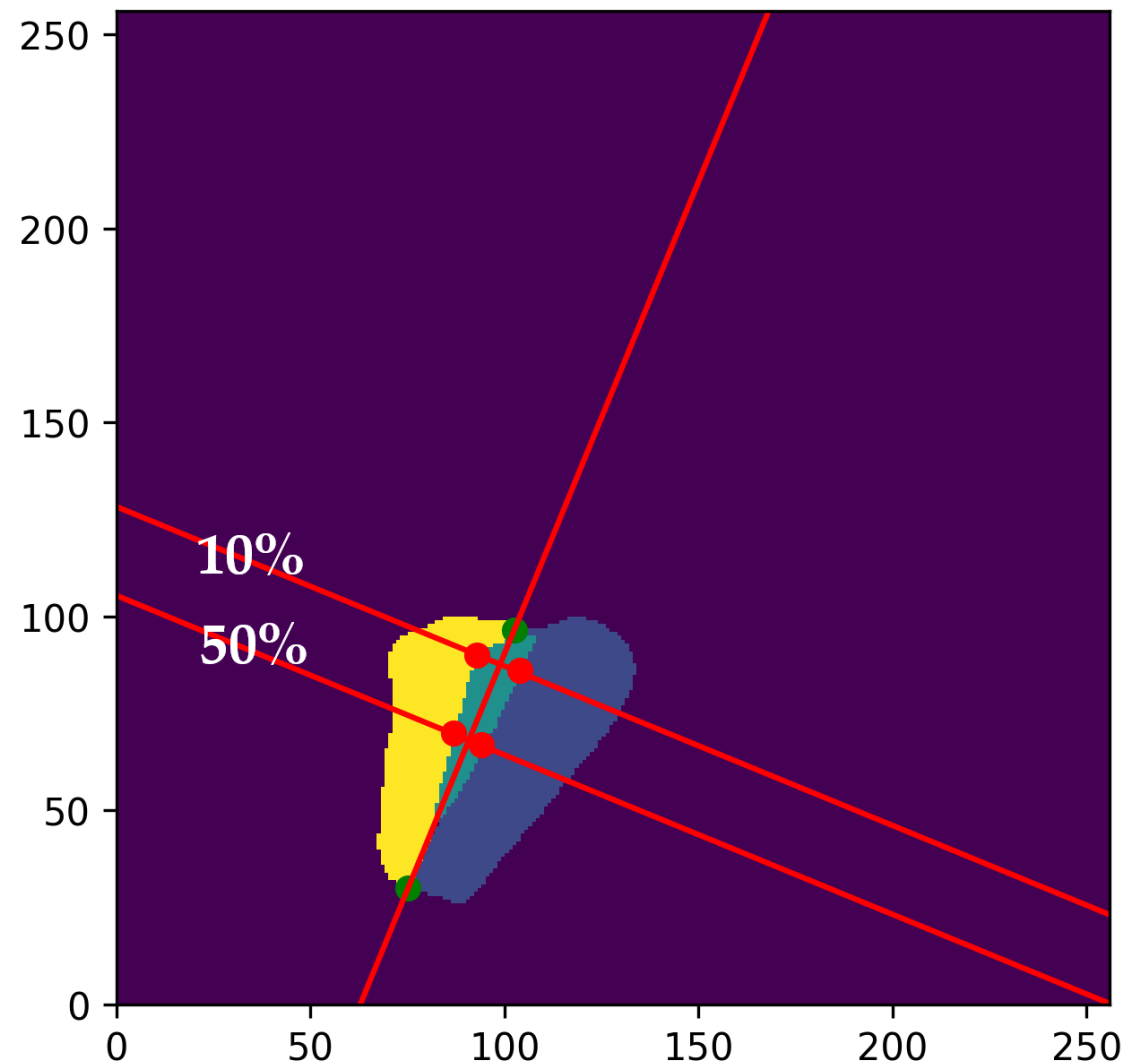
# Efter netværksberegninger

Efter netværksberegninger.

U-LSTM\_5^CE udfører segmenteringen.  
Den giver ikke **lokaliseret** information om specifikke afstande mellem stemmebåndene.

Vi foreslår en algoritme til at beregne dette.

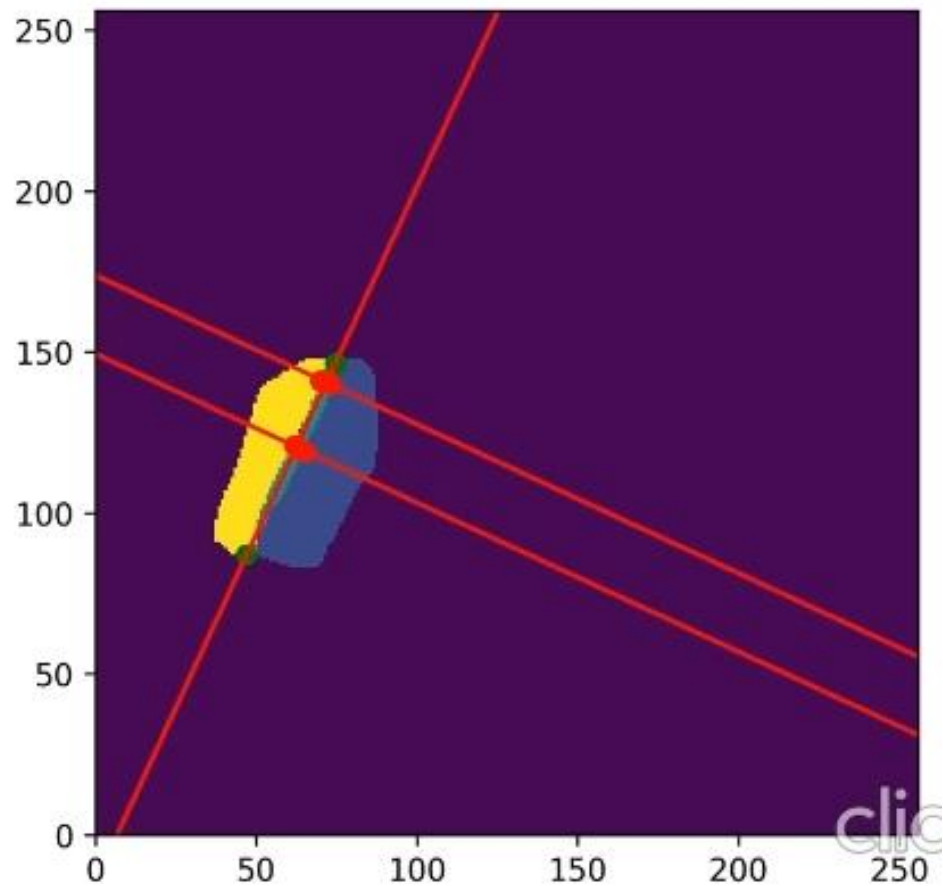
Vores beregninger vises som tværgående linjer mellem stemmebåndene, 10% fra det bagerste af stemmeridsen, og 50% fra det bagerst af stemmeridsen til sammenligning.



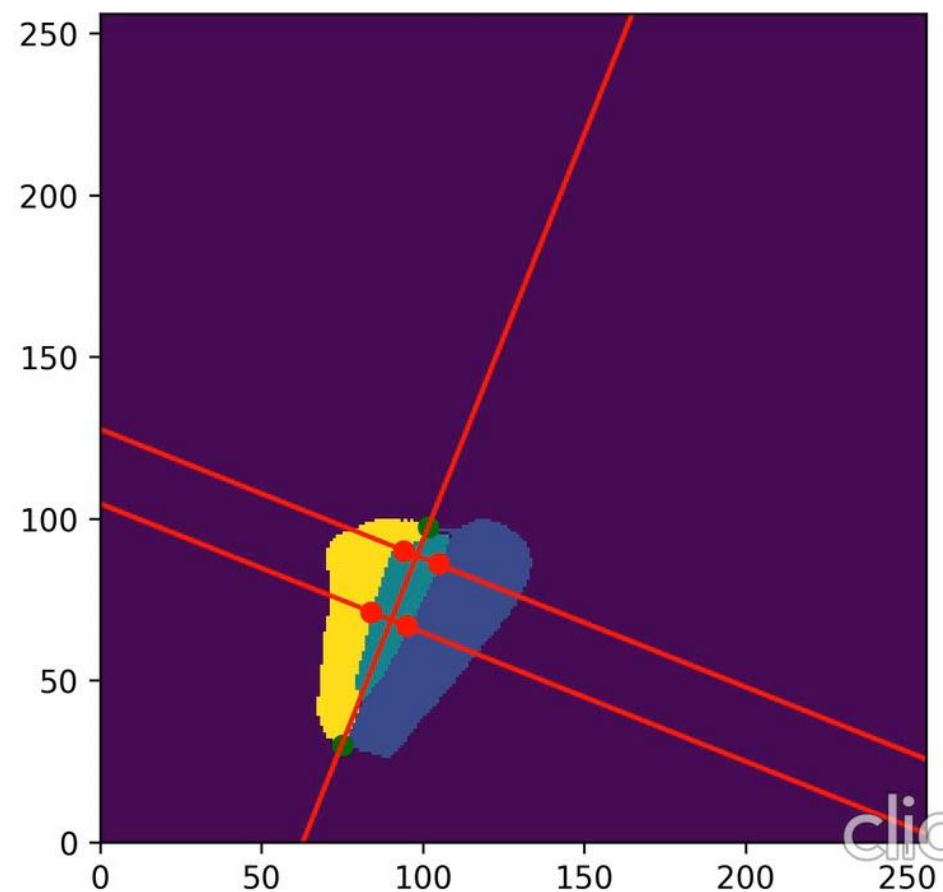
# Post network calculations - Video

Videoer af resultaterne fra efter netværksberegningerne.

Lukkedeficit bagtil



Fuld lukning mellem stemmebåndene



# Resultatet for deep learning målingerne 10% og 50% afstand bagfra mellem stemmebåndene

	Parameter	N	n	LS Mean	Standard Error	95% Confidence Interval	P-Value
<b>a</b>	10%						
	Normal subjects/videos.	20	2000	1.66	0.33		
	Subjects/videos with a rear glottal gap.	30	3000	4.44	0.27		
	Diff (subjects/videos with a rear glottal gap–normal subjects/videos).			<b>2.77</b>	0.42	(1.93 - 3.62)	<b>&lt; 0.0001</b>
<b>b</b>	50%						
	Normal subjects/videos.	20	2000	2.83	0.30		
	Subjects/videos with a rear glottal gap.	30	3000	2.83	0.24		
	Diff (subject/videos with a rear glottal gap–normal subjects/videos).			<b>0.26</b>	0.39	(-0.52 - 1.04)	<b>0.50</b>

Afstanden mellem stemmelæberne måles i pixels for hvert billede, og det Least Squares (LS) Mean beregnes for alle subjects/videoer.

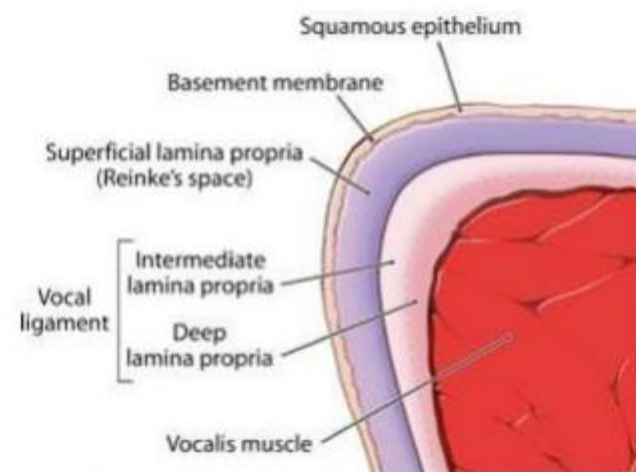


# Optisk kohærens-tomografi udstyr med skoper til HSV og OCT

En UHR-OCT opsætning er blevet konstrueret, der kan kombinere HSV (4.000 billeder i sekundet).

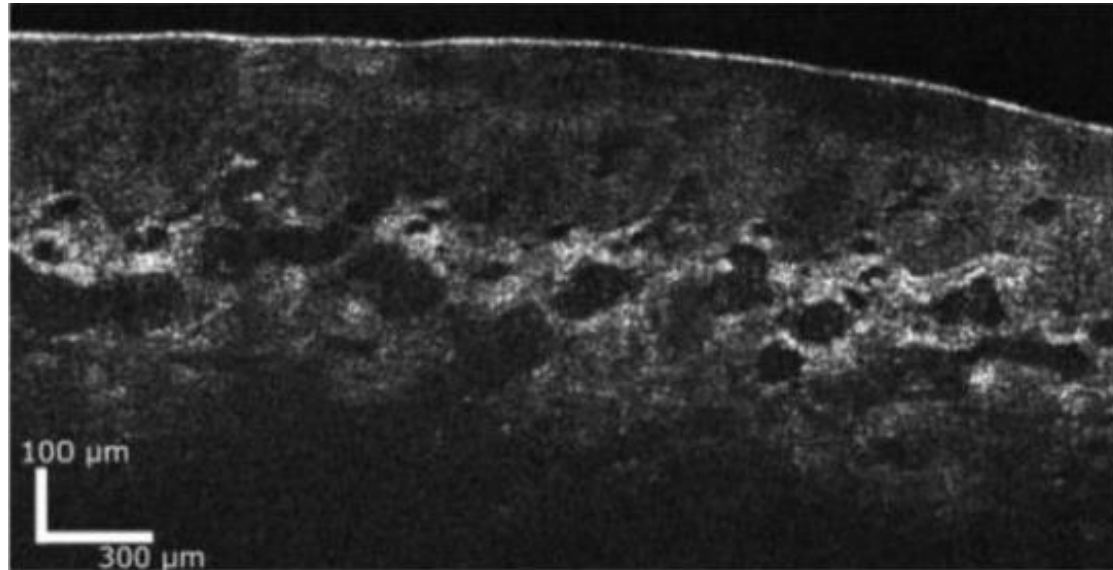
Mundslimhinde (inden i den nedre læbe) med epitel, kirtler og blodkar er blevet præsenteret.

En sonde til strubehovedet er under konstruktion.

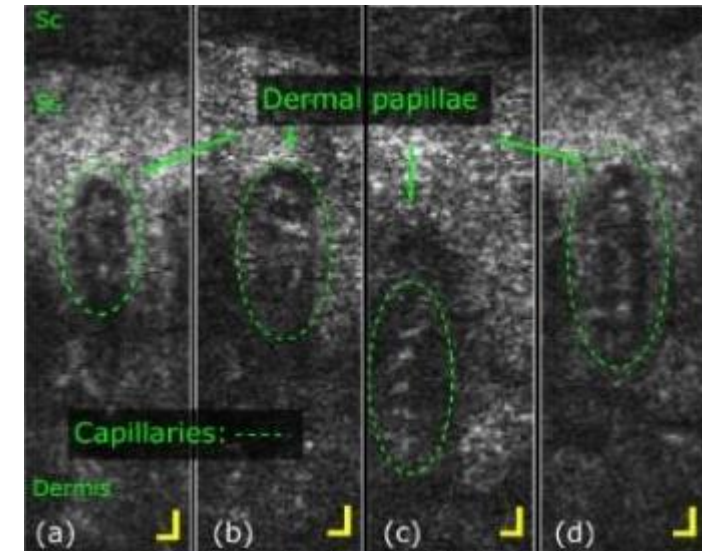


Source: <http://www.clevelandvoiceandsleep.com>

# Optisk kohærens-tomografi udstyr med skoper til HSV og OCT



In vivo optagelse UHR OCT fra mundslimhinde (inderside af underlæbe) med epitel, kirtler og blodkar, med ultrahøj opløsning sv. til en rumlig opløsning mindre end 5 my og en dybde i vævet mellem 0,4 og 1 mm.



Individuelle hudpapiller og kapillærer i hånden, målestokken svarer til 20 μm. I samarbejde med dermatologer

## Hovedbudskaber

- 1. Der er nu konstrueret et UHR-OCT-opstilling, der kan kombineres med HSV under fonation (4.000 billeder per sekund).**
- 2. Den høje opløsning af OCT giver præcis information om celleniveauerne for en bedre forståelse af dysfunktion og slimhindens ændringer for stemmelæberne.**
- 3. Dyb læring er nødvendig for at analysere større mængder HSV-data. Beregninger er blevet foretaget manuelt i mange år. Vi kan nu måle afstande mellem stemmelæberne på et defineret sted, med store mængder billeder, der kan bruges til funktionsanalyse og UHR-OCT.**

Tak

## References

Israelsen N, Larsen CF, Pedersen M. Kvantitativ undersøgelse af stemmebånd med højhastighedsvideo og optisk kohærens-tomografi *Ugeskr Læger status artikel* (Danish weekly medical journal) 2022;184:V02210146

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

Pedersen M, Larsen C, Eeg M (2022) Objective Quantitative Analysis of Laryngeal Glottal Gaps using High-Speed Video in Glottal Analysis Tools, a Case-Control Study. Research in Health Science. 7. p1. DOI: 10.22158/rhs.v7n4p1.

Larsen CF, Pedersen M (2022) Comparison of convolutional neural networks for classification of vocal fold nodules from high-speed video images. Eur Arch Otorhinolaryngol. 2022 Nov 11. doi: 10.1007/s00405-022-07736-6

Pedersen M, Larsen CF, Madsen B, Eeg M (2023) Localization and quantification of glottal gaps on deep learning segmentation of vocal folds. Sci Rep 13, 878. doi: 10.1038/s41598-023-27980-y

# 4 Neural Netværk

4 Neurale Netværk (CNN) blev sammenlignet for deres evne til at klassificere stemmebåndsknuder. For at teste om de kan hjælpe i rutinemæssig diagnostik.

Billedet viser nøjagtigheden under træningen af de 4 Neurale Netværk (CNN). Nøjagtigheden måles både på træningsdataene og på et separat datasæt til validering.

CNN5 er en brugerdefineret CNN med 5 lag, som viste sig at have den højeste nøjagtighed.

Nøjagtighed for træningsdata er visualiseret i røde linjer.

Nøjagtighed for valideringsdata er visualiseret i blå linjer.

