

# Quantitative Examination of Vocal Folds, Perspectives for Deep Learning High-Speed Image Analysis and Optical Coherence Tomography with Ultra-High Resolution

Niels Israelsen, Danish Technical University Lyngby  
Christian Frederik Larsen, Copenhagen Business School Frederiksberg  
Mette Pedersen Medical research Center Copenhagen F

**30<sup>th</sup>** Congress of  
**Union of the European  
Phoniatricians**



27-30 April 2023  
Xanadu Convention Center, Antalya | Türkiye

# What we have

# High-speed video for imaging

Our commercial laryngoscopy setup with 256 x 256 pixels, recorded with 4.000 frames per second

From Richard Wolf GmbH in Knittlingen, Germany (Endocam 5562), which features a high-speed camera mounted on a rigid scope.

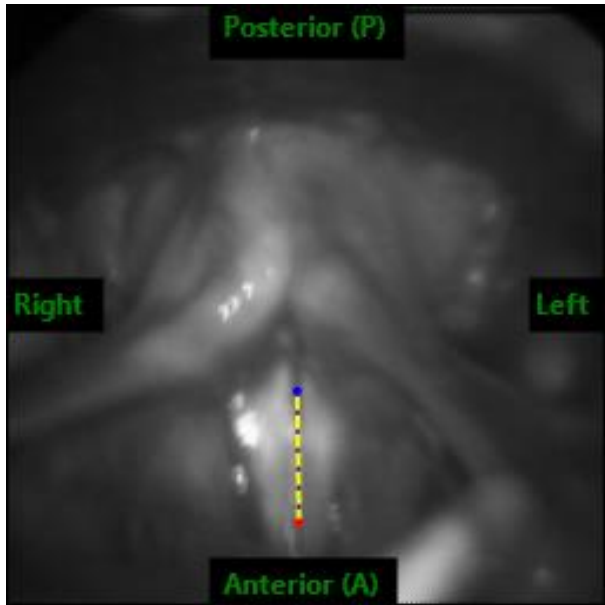
Full register



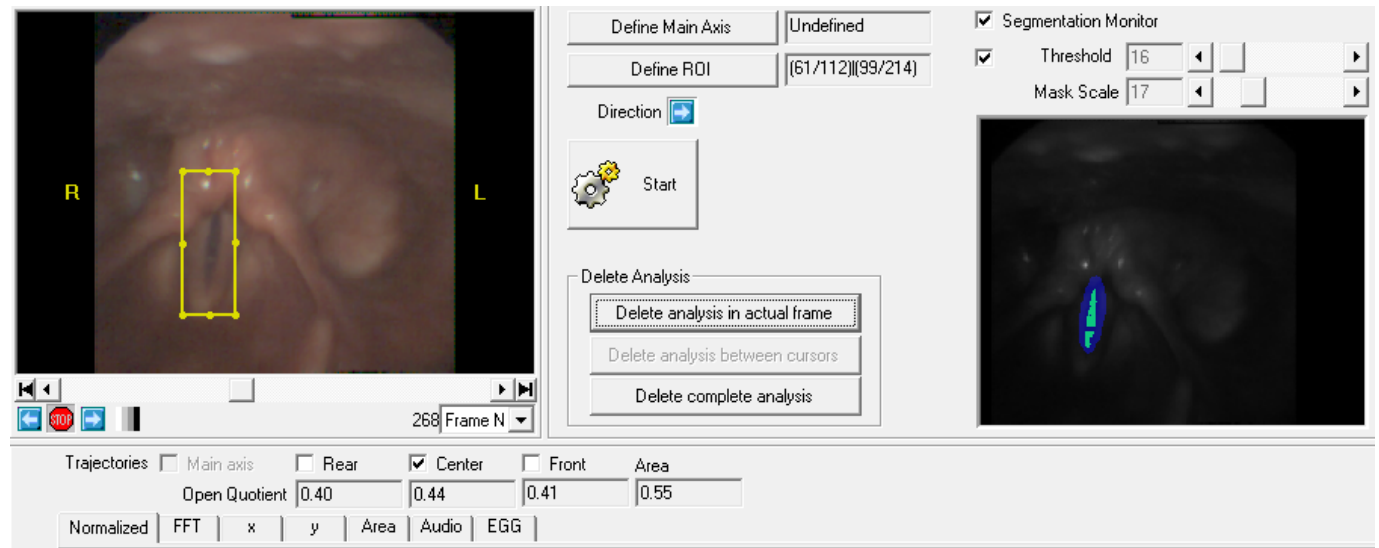
Upper register



# Analysis of the vocal folds



A normal larynx with markings of the center of glottis.



Region of interest and markings of the edges of the vocal folds with HSV.

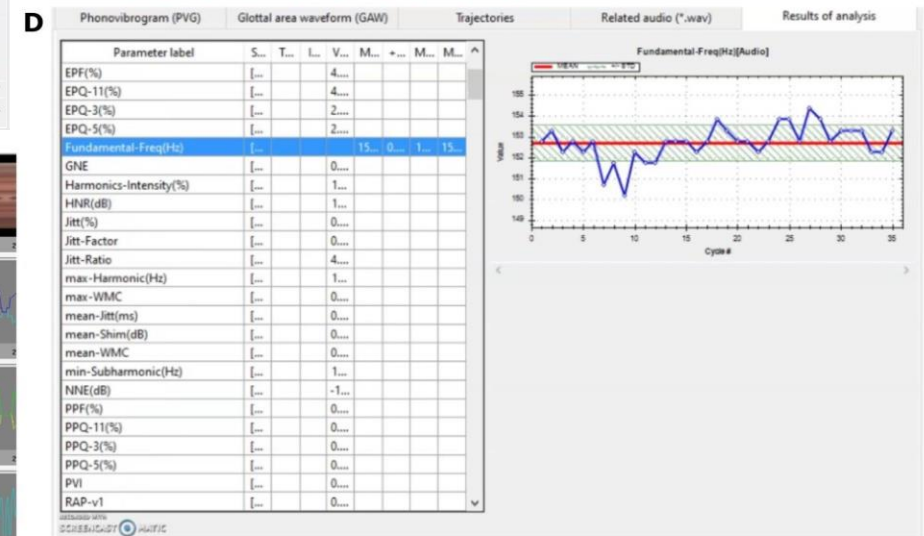
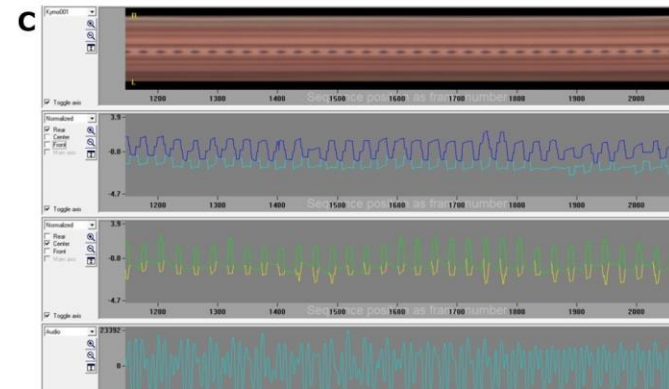
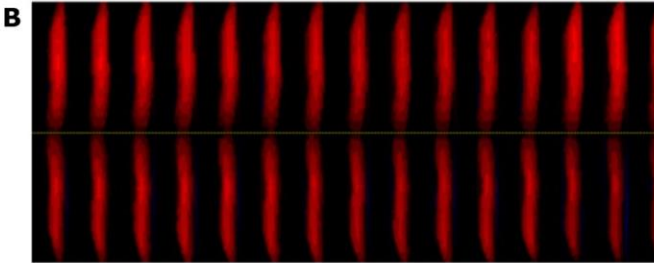
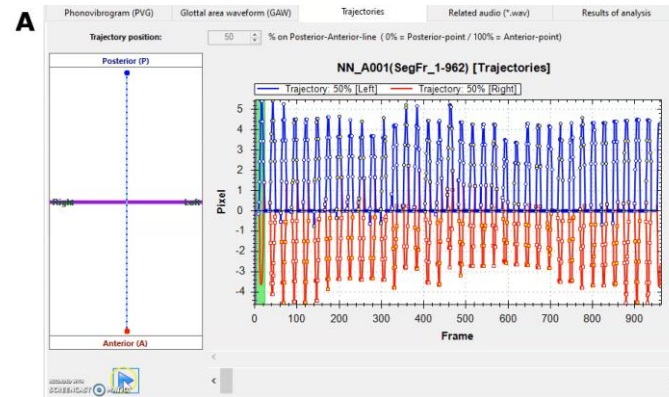
# Quantification of vocal fold analysis

(A) A high-speed video from a **software reproduction (GlottalAnalysisTools)**

(B) **Phonovibrogram**

(C) The top curve is a **high-speed video kymogram**, then the **closure of the rear of both right and left vocal folds** in a time period, the closure of the vocal folds in **the middle**, and at the bottom, the **corresponding acoustic image** is shown.

(D) **183** measurements.



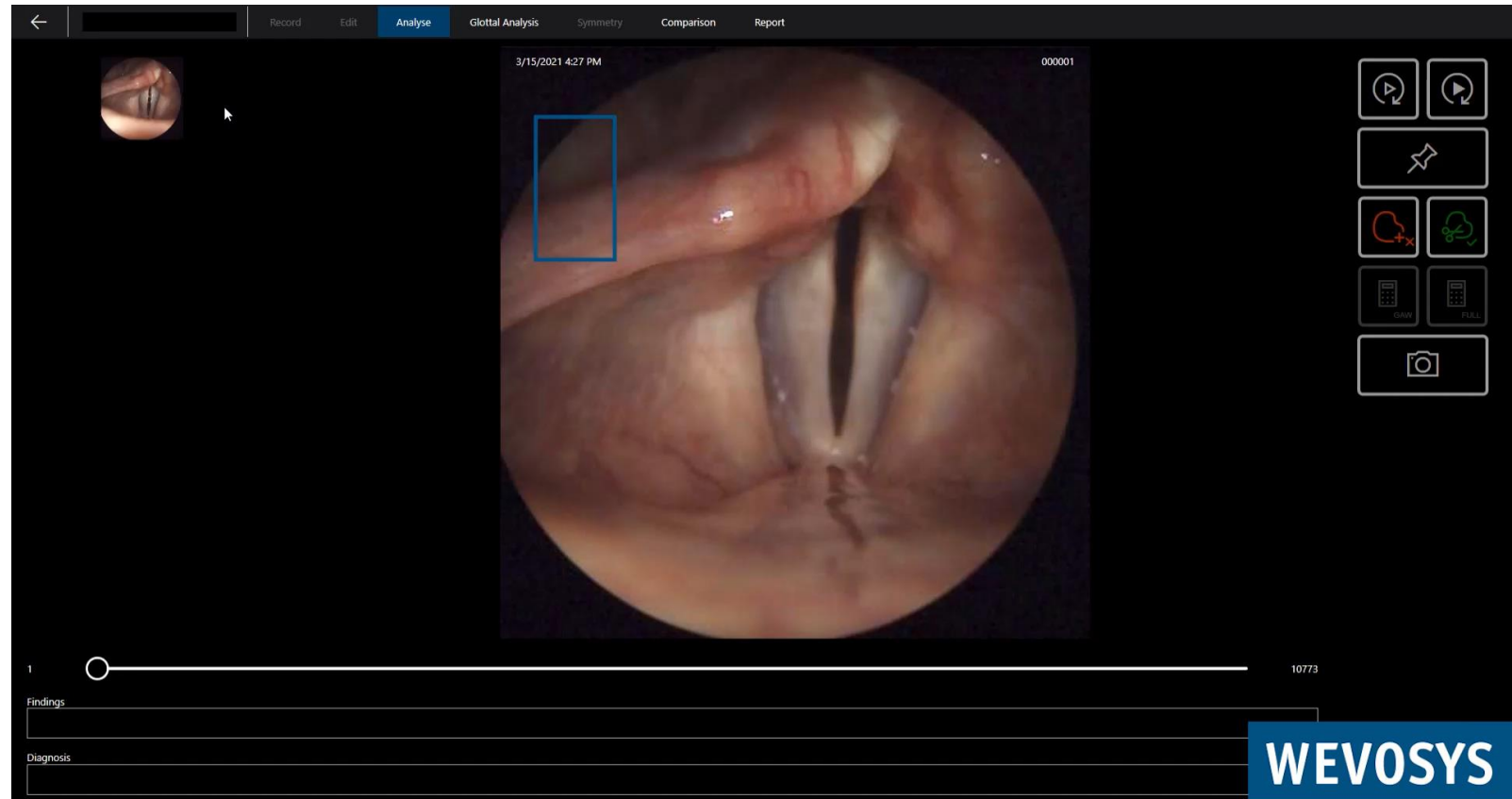
What we would like to have

# High-speed video for imaging

lingWAVES 4 HSV, a commercially available - Wevosys for high-speed video endoscopy

1.440 x 1.024 pixels in color.  
At 4.000 fps, up to 8.000 fps is possible.

Glottis is ROI.





# Deep learning software for analysis of the distance between vocal folds

(a) High-Speed Video

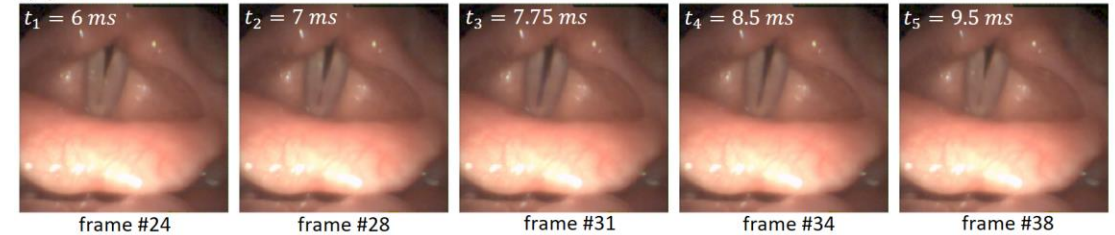
(b) neural network segmentation generated with U-LSTM\_5^CE, with the calculation of the glottal area

(c) overlay of segmentation results

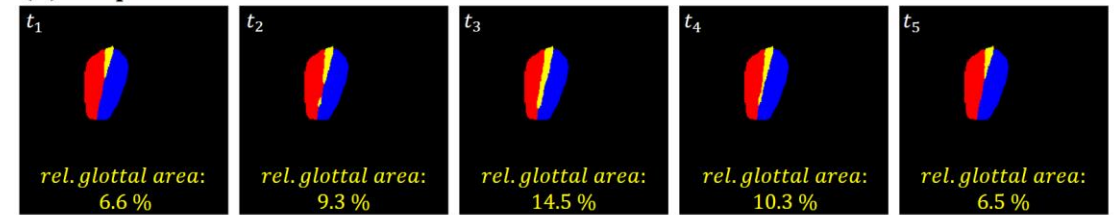
(d) mean and standard deviation for the normalized relative area between the vocal folds for the entire sequence, equal to the sum of the area between the vocal folds + right vocal fold + left vocal fold.

The yellow lines indicate where the pictures are taken in the oscillation.

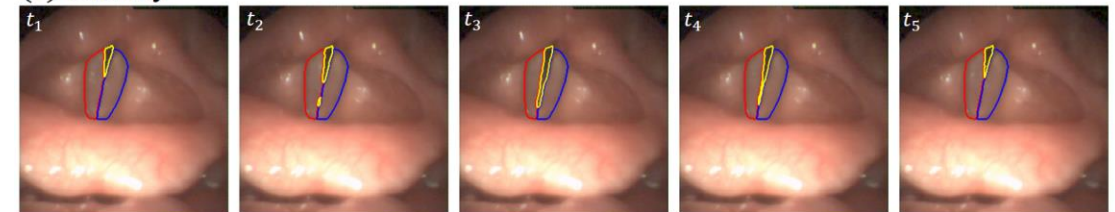
(a) HSV frames



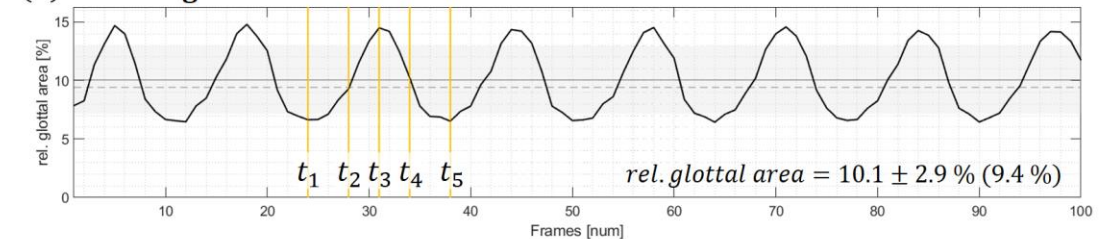
(b) NN predictions



(c) overlay

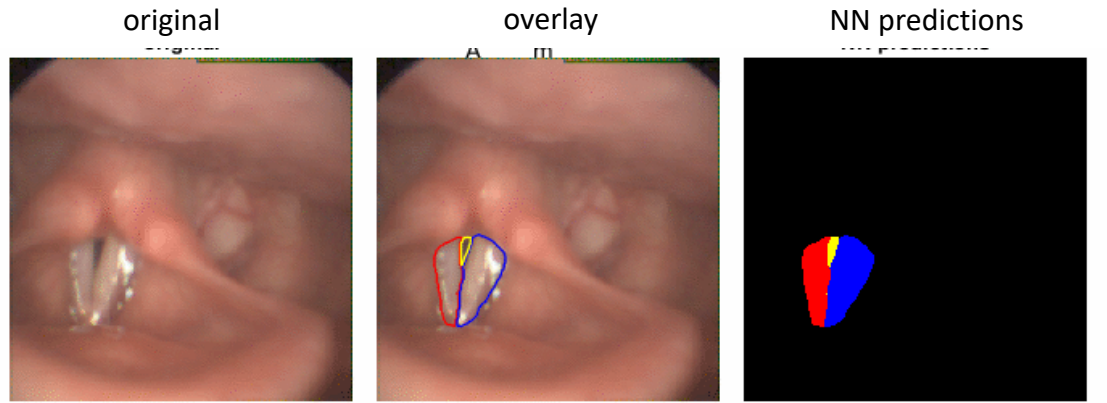


(d) relative glottal area

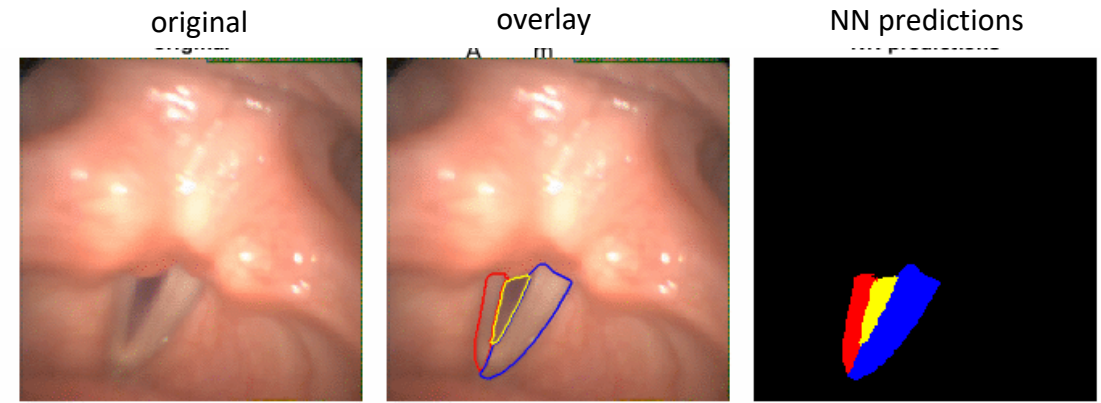
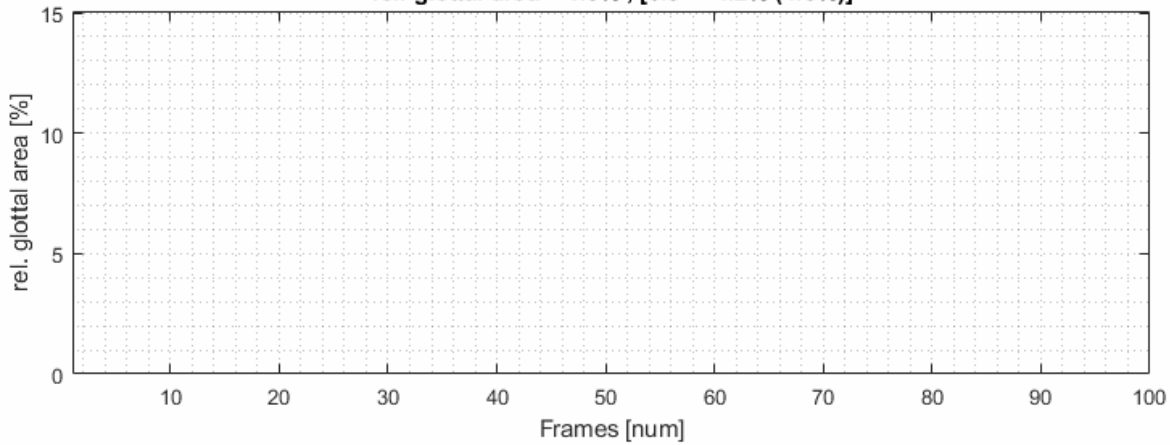




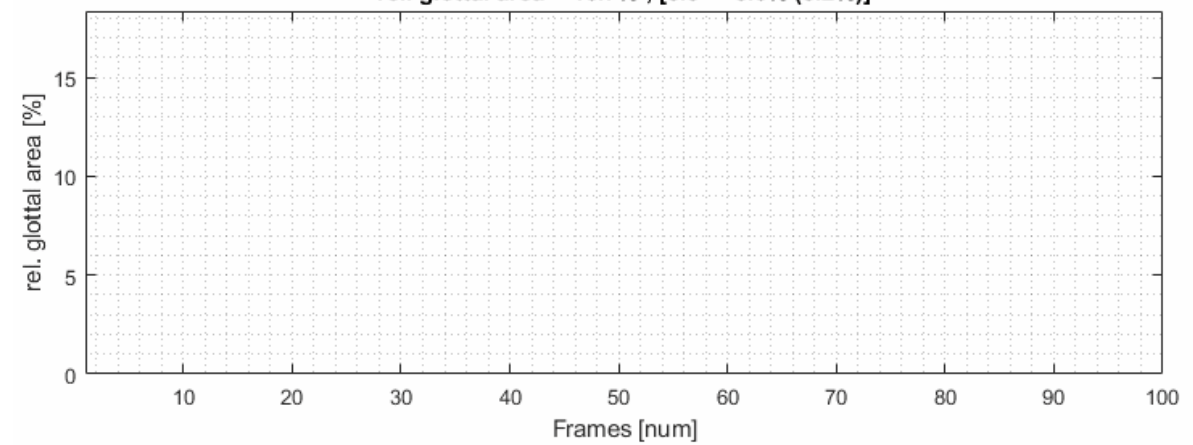
# Deep learning software for analysis of the distance between vocal folds



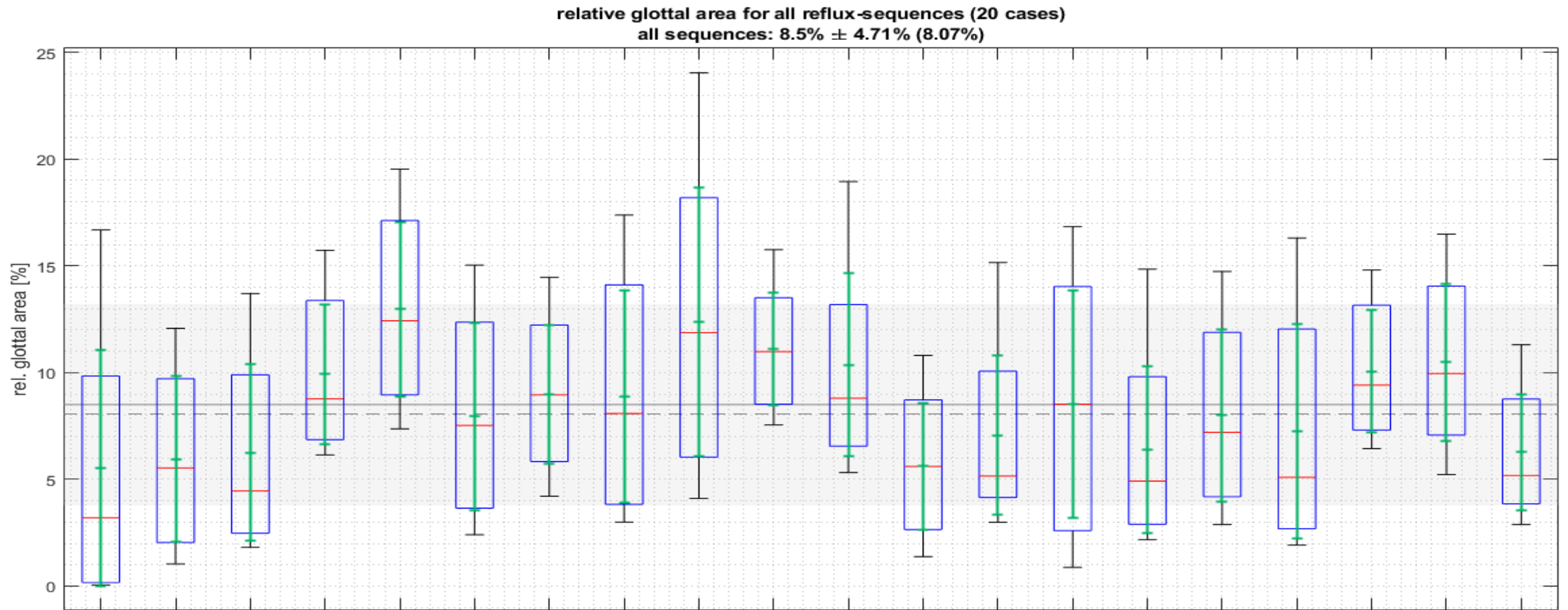
rel. glottal area = 4.3% , [6.3 +- 4.2% (4.5%)]



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



# Analysis of the distance between vocal folds



20 videos of reflux patients with the whole glottis movement of closure analyzed with Matlab

# Optical Coherence Tomography setup

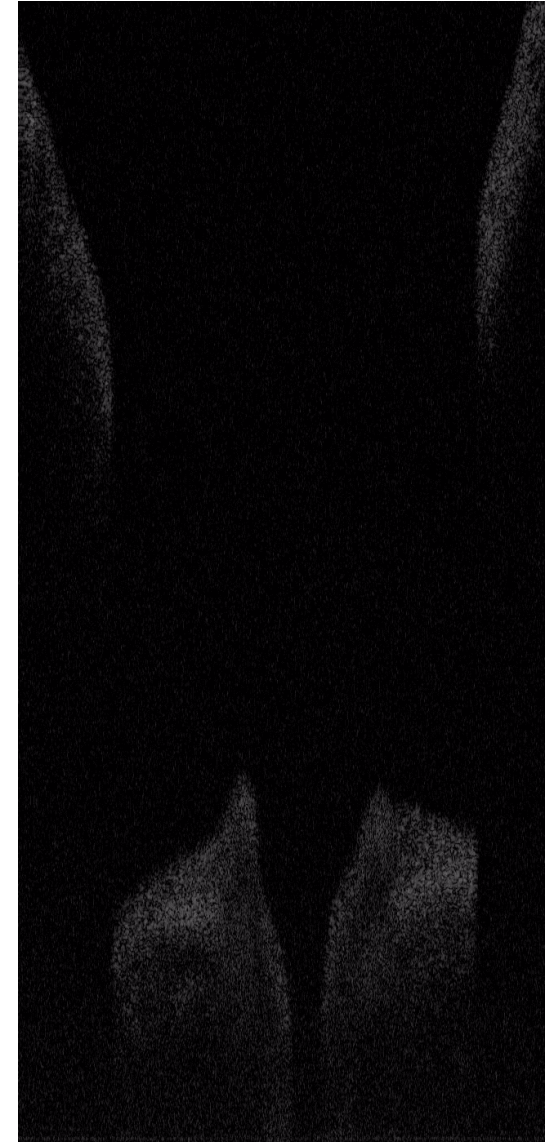


Optical coherence tomography with a resolution of 250 cross-section images per second -gives artefacts.

Single mucosal movement (e.g. smooth or irregular).

Vocal fold cells (normal or with swelling etc.)

Edges of the vocal folds



With permission from Brian Wong and Lily Chen

What we would like to use in the clinic

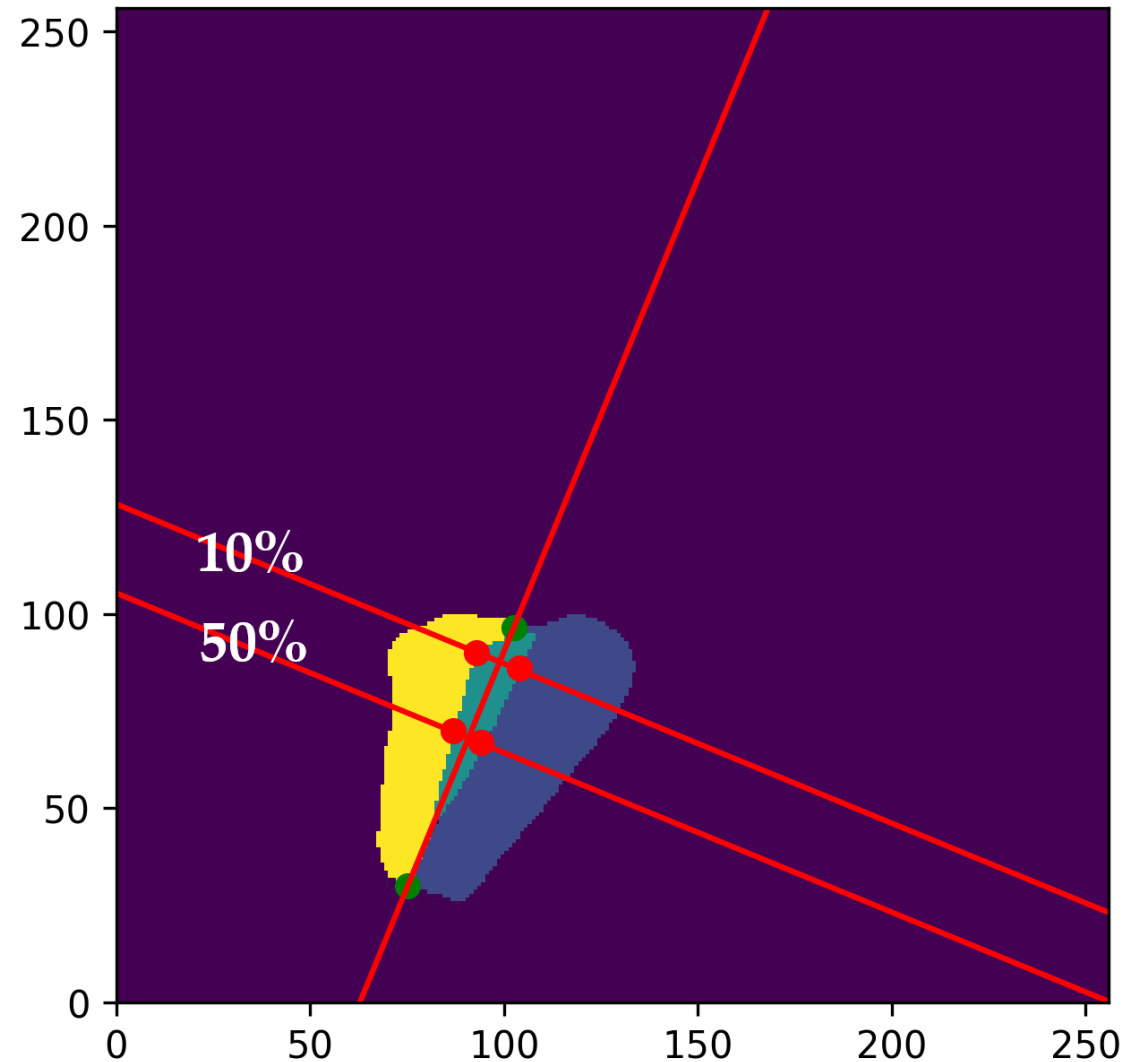
# Post network calculations

Post-network calculations.

U-LSTM\_5^CE, performs the segmentation.  
It does not provide **localized** information on specific distances between the vocal folds.

To extract this data, we propose an algorithm to serve as post-network calculations.

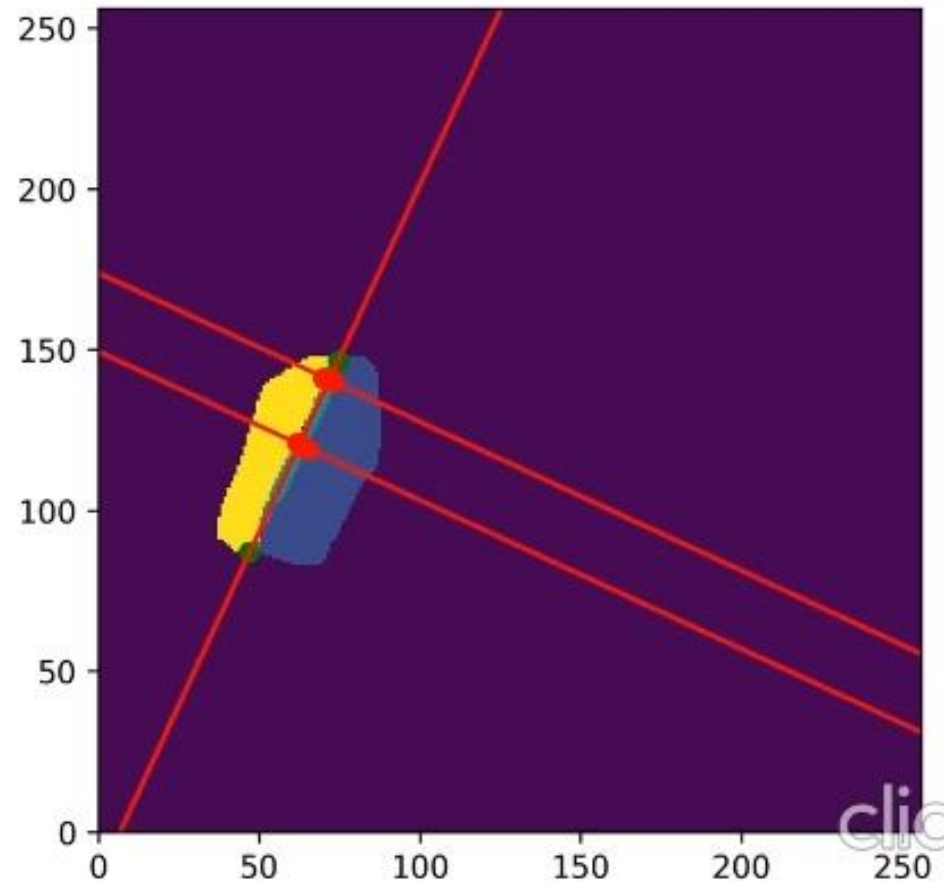
Our calculations are presented on the distances between the vocal folds as transverse lines on the glottis at 10% from the rear of the glottis with a comparison at 50% from the rear of the glottis, as examples.



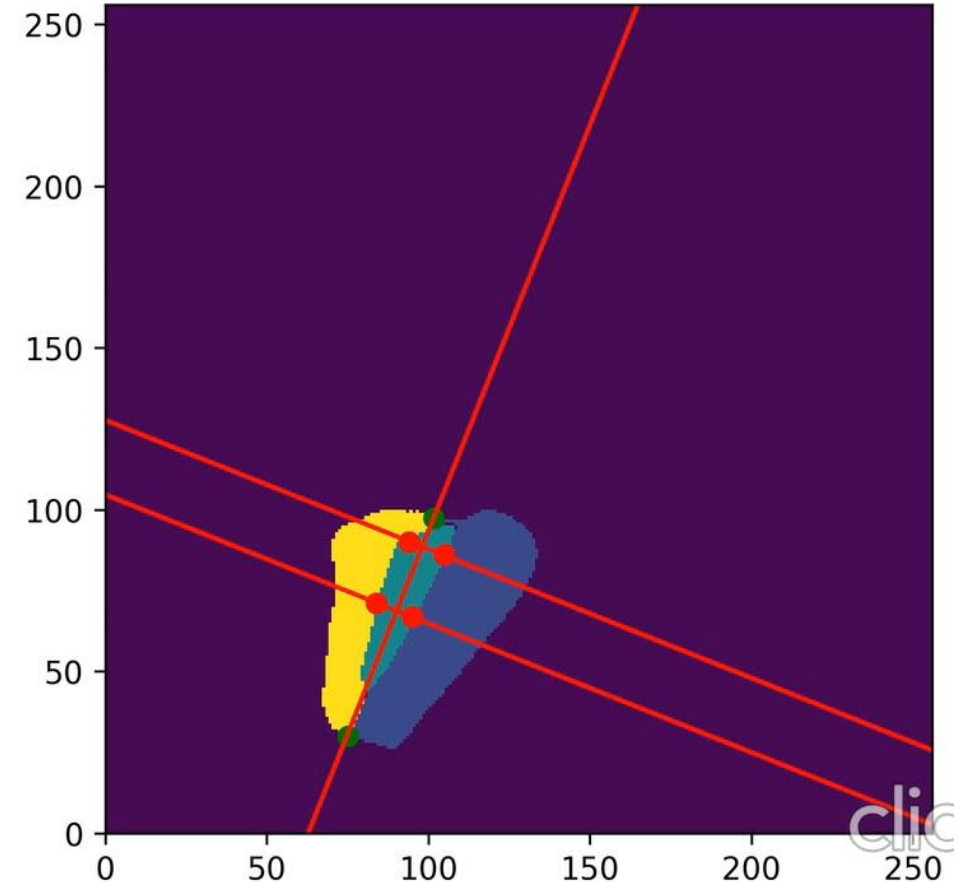
# Post network calculations - Video

Videos comprised of the resulting images with post-network calculations.

Insufficient closure of the rear



Full closure



# Result of deep learning analysis

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>

The distance between the vocal folds is measured in pixels for every frame, and the Least Square (LS) Mean is calculated for all subjects/videos.

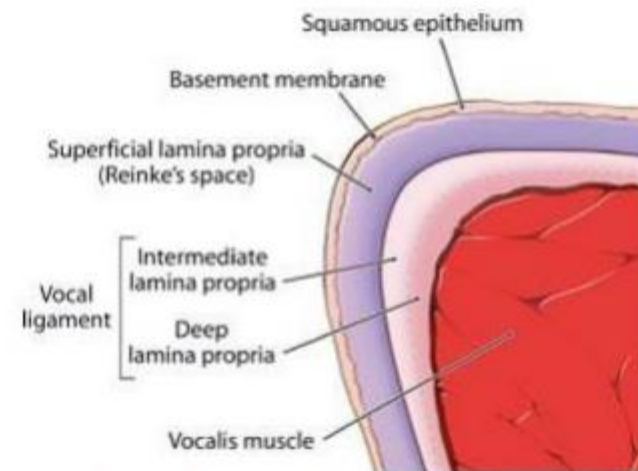


# Ultra High-Resolution Optical Coherence Tomography

UHR-OCT setup has been constructed that can combine HSV (4.000 frames per second)

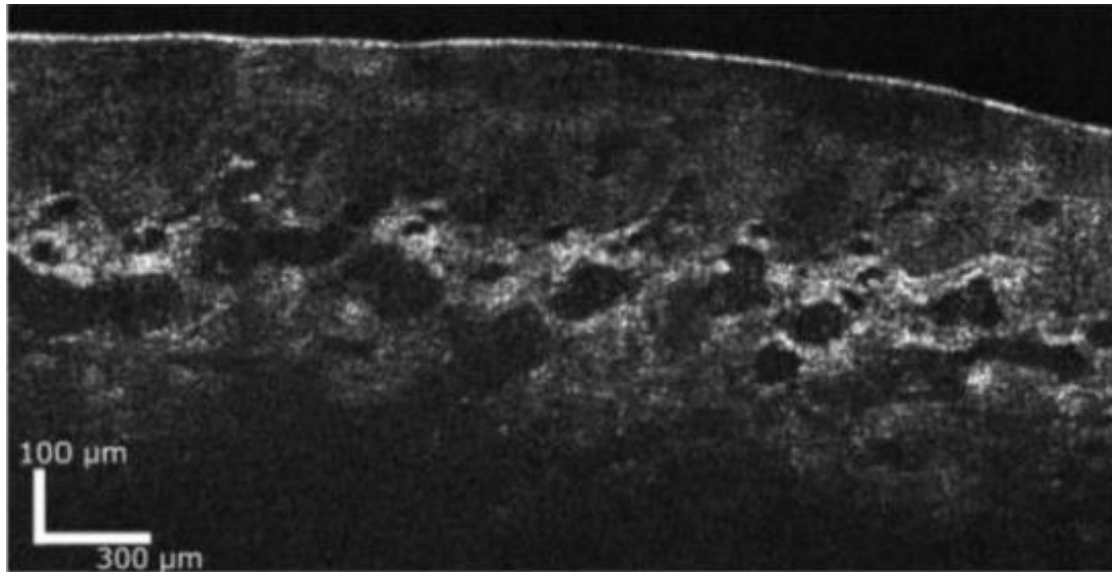
Oral mucosa (inside of lower lip) with epithelia, glands, and blood vessels has been presented.

A probe for the larynx is under construction.

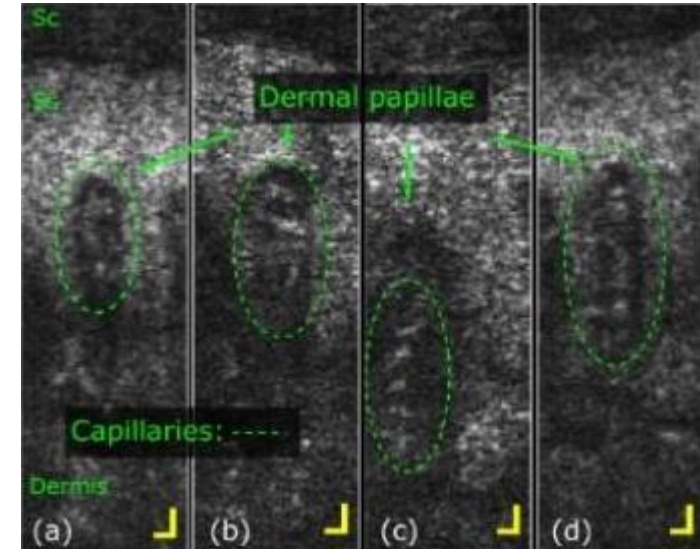


Source: <http://www.clevelandvoiceandsleep.com>

# Ultra High-Resolution Optical Coherence Tomography



In vivo UHR-OCT has a spatial resolution of less than 5 micron and can reach between 0,4 and 1 mm in the tissue.



Individual skin papillae and capillaries of the hand, the scale is 20  $\mu\text{m}$ -  
In cooperation with dermatologists

## 3 key take-aways

- 1. There is now constructed a UHR-OCT setup that can combine to HSV during phonation (4.000 frames per second).**
- 2. The high resolution of OCT provides precise information of the cellular levels, for a better understanding of dysfunction and mucosal changes also in the larynx, especially for the vocal folds.**
- 3. Deep learning is necessary to analyze larger amounts of HSV data. Calculations have been made for many years, manually.**
- 4. We can now measure distances between the vocal folds at a defined place, of large amounts of frames, usable for function analysis and UHR-OCT.**

Thank you

## References

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

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