Quantitative Examination of Vocal Folds, 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

30th Congress of Union of the European Phoniatricians
27-30 April 2023
Xanadu Convention Center, Antalya I Türkiye

Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

High-speed video for imaging

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

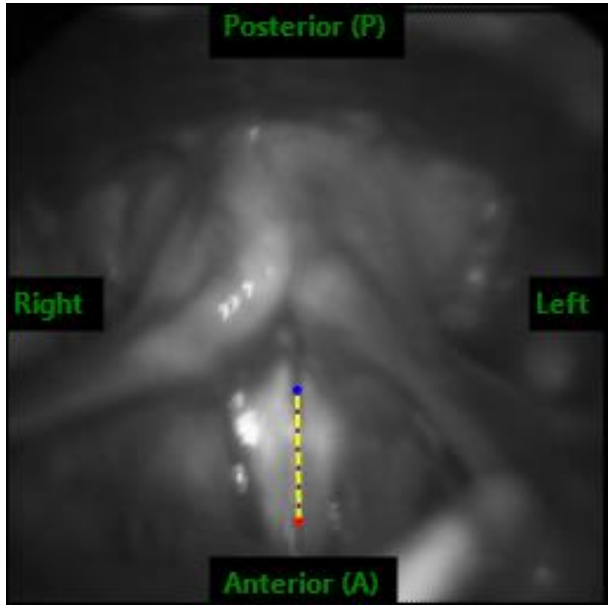
From Richard Wolf GmbH in Knittlingen, Germany (Endocam 5562), which features a high-speed camera mounted on a rigid scope.

Full register

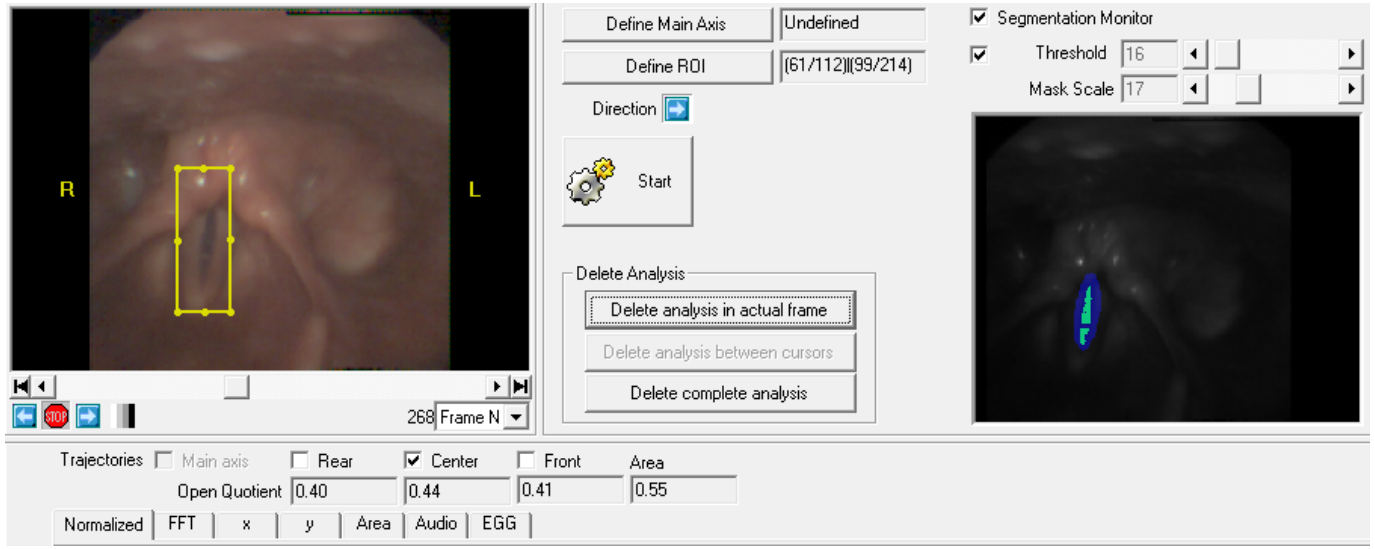
Upper register

Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

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.

Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

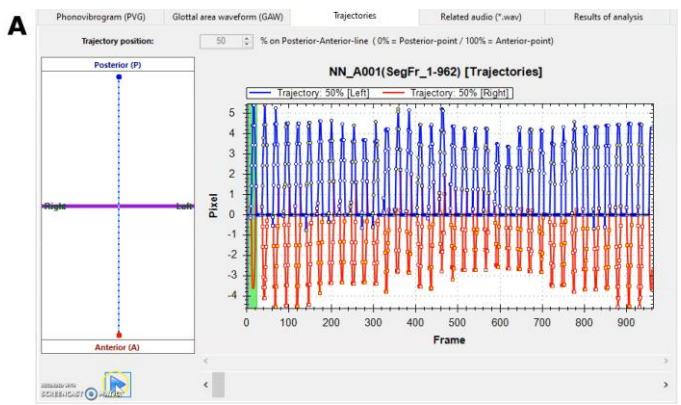
Quantification of vocal fold analysis

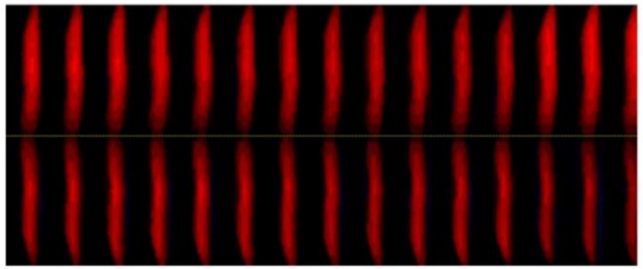
(A) A high-speed video from a **software reproduction (GlottalAnalysisTools)**

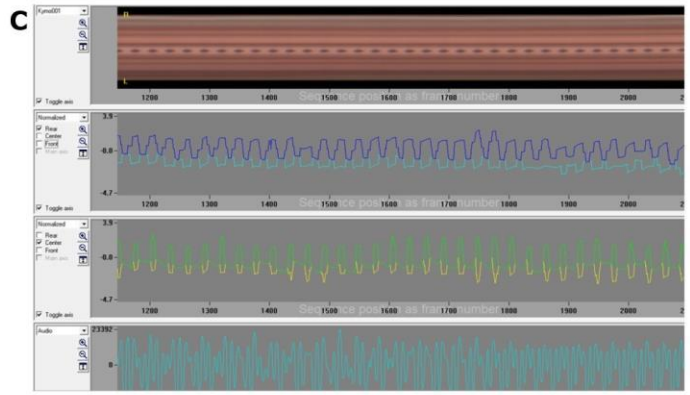
(B) **Phonovibrogram**

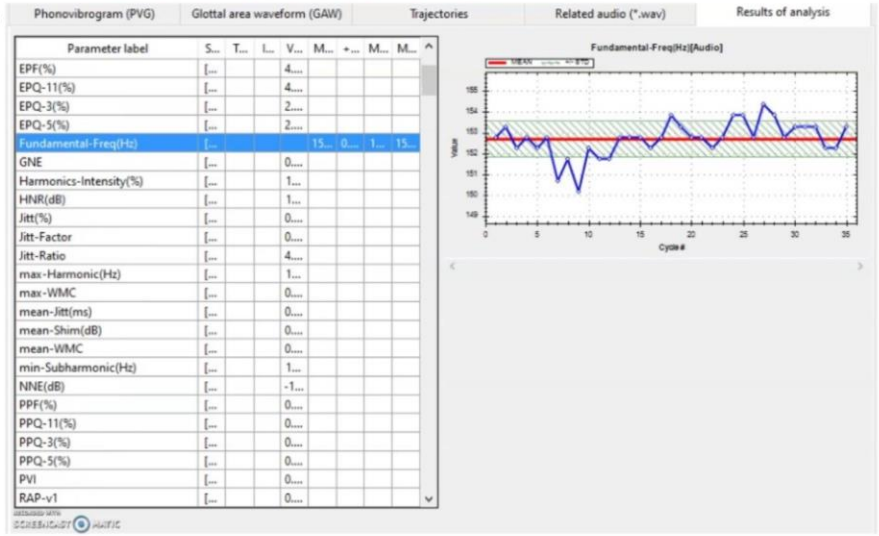
(C) The top curve is **a high-speed video kymogram**, then the **closure of the rear of both right and left vocal folds** in a time period, the closure of the vocal folds in **the middle**, and at the bottom, the **corresponding acoustic image** is shown.

(D) **183** measurements.









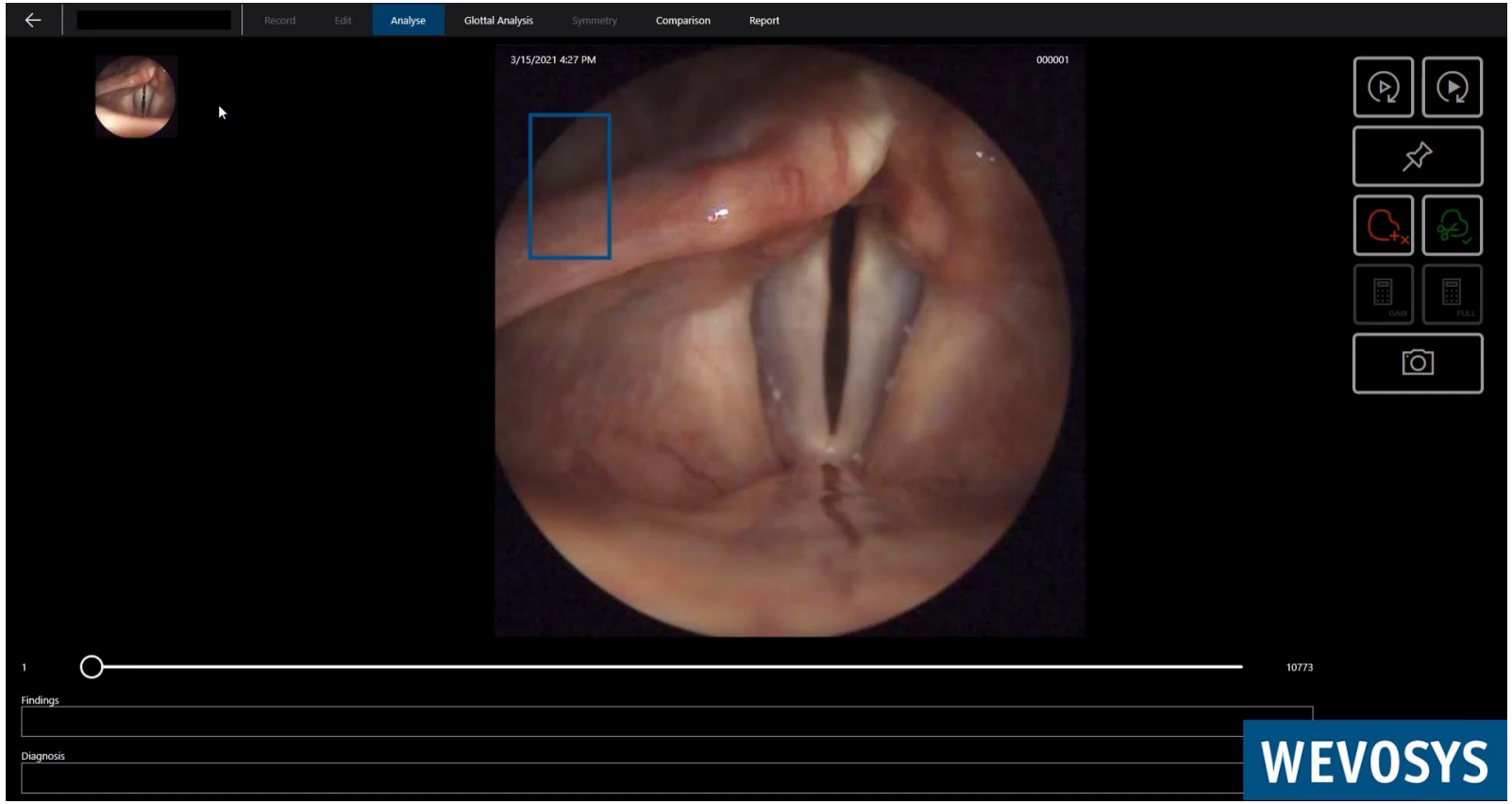
Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

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.



Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

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).

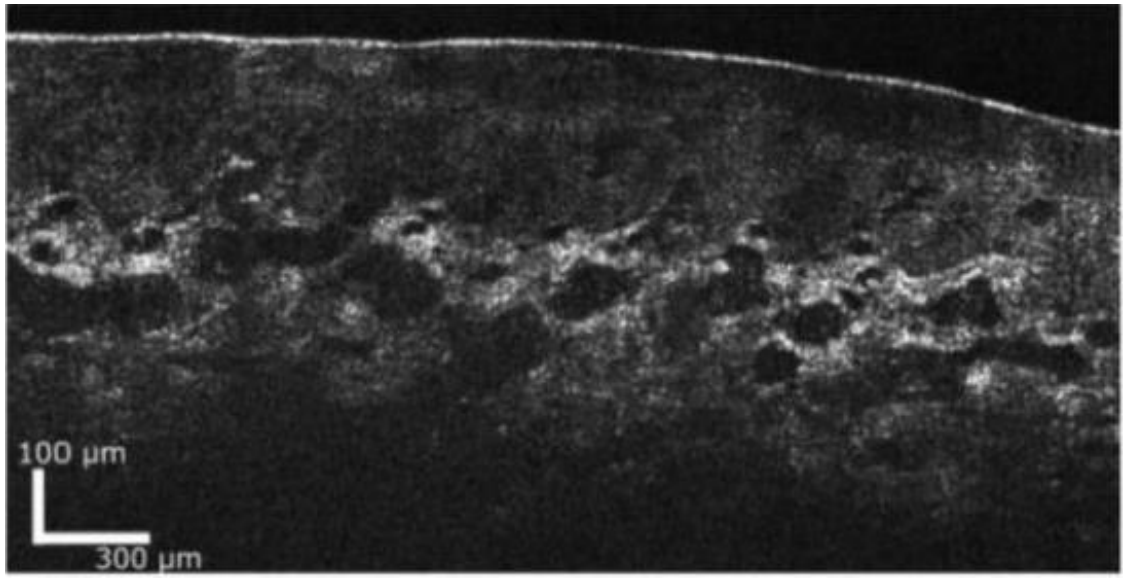
Vocal fold cells (normal or with swelling etc.)

Edges of the vocal folds

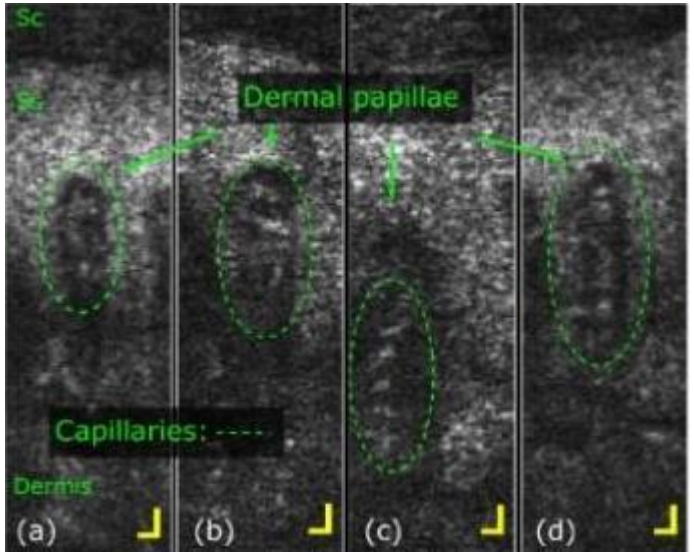
With permission from Brian Wong and Lily Chen

Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

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 µm-
In cooperation with dermatologists

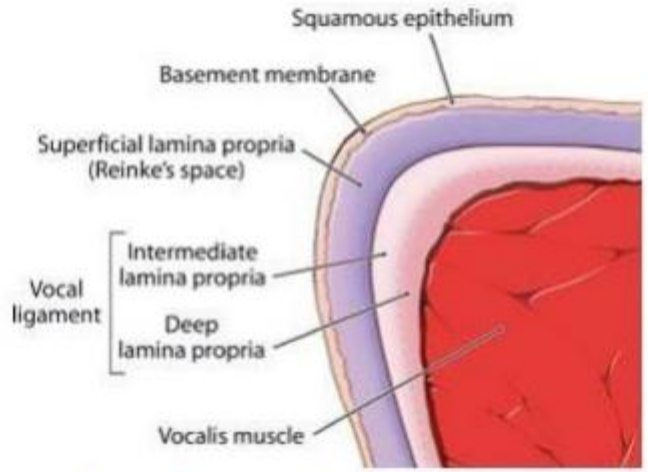
Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

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

Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

Deep learning software for analysis of the distance between vocal folds

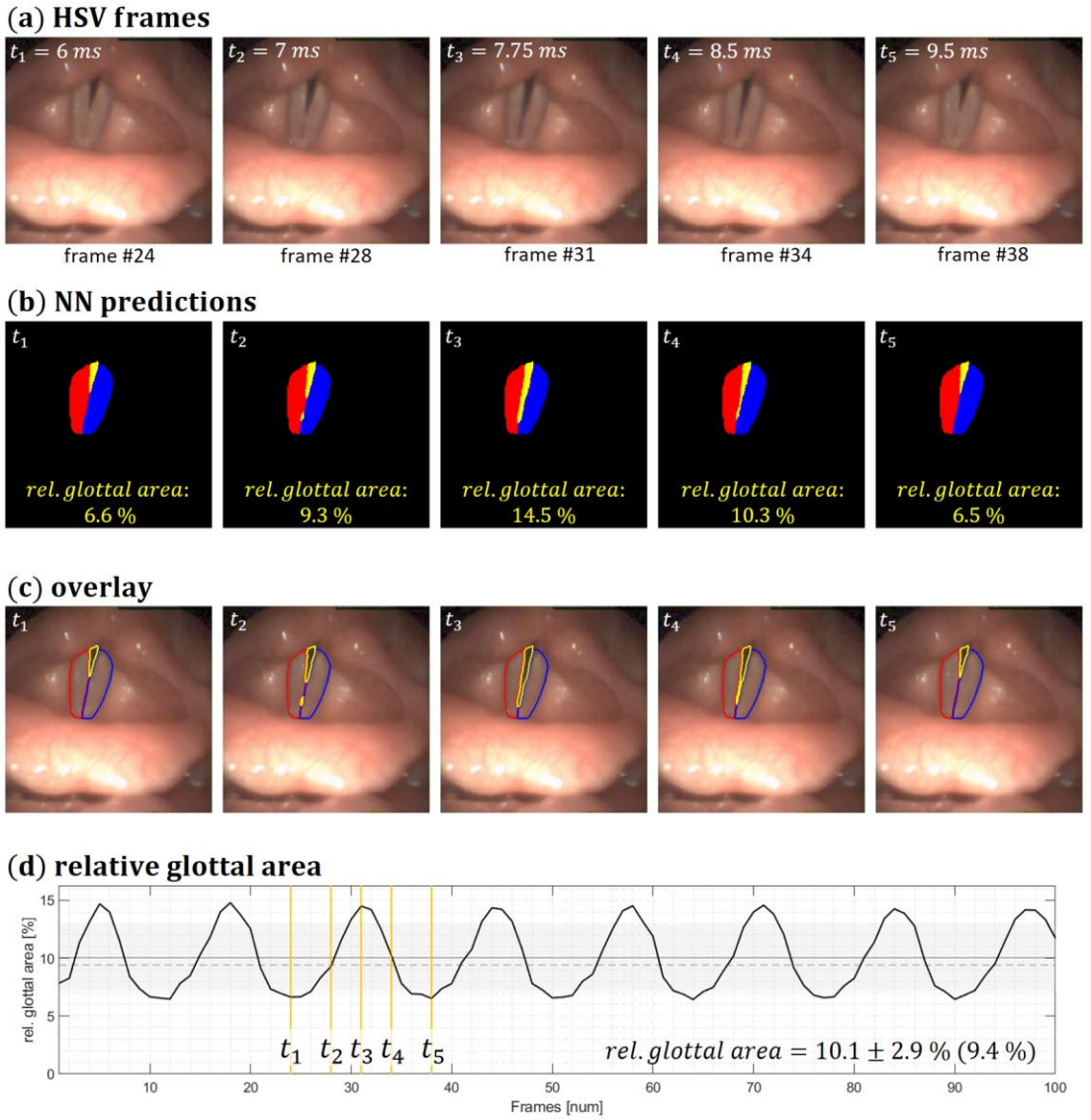
(a) High-Speed Video

(b) neural network segmentation generated with U-LSTM_5^CE, with the calculation of the glottal area

(c) overlay of segmentation results

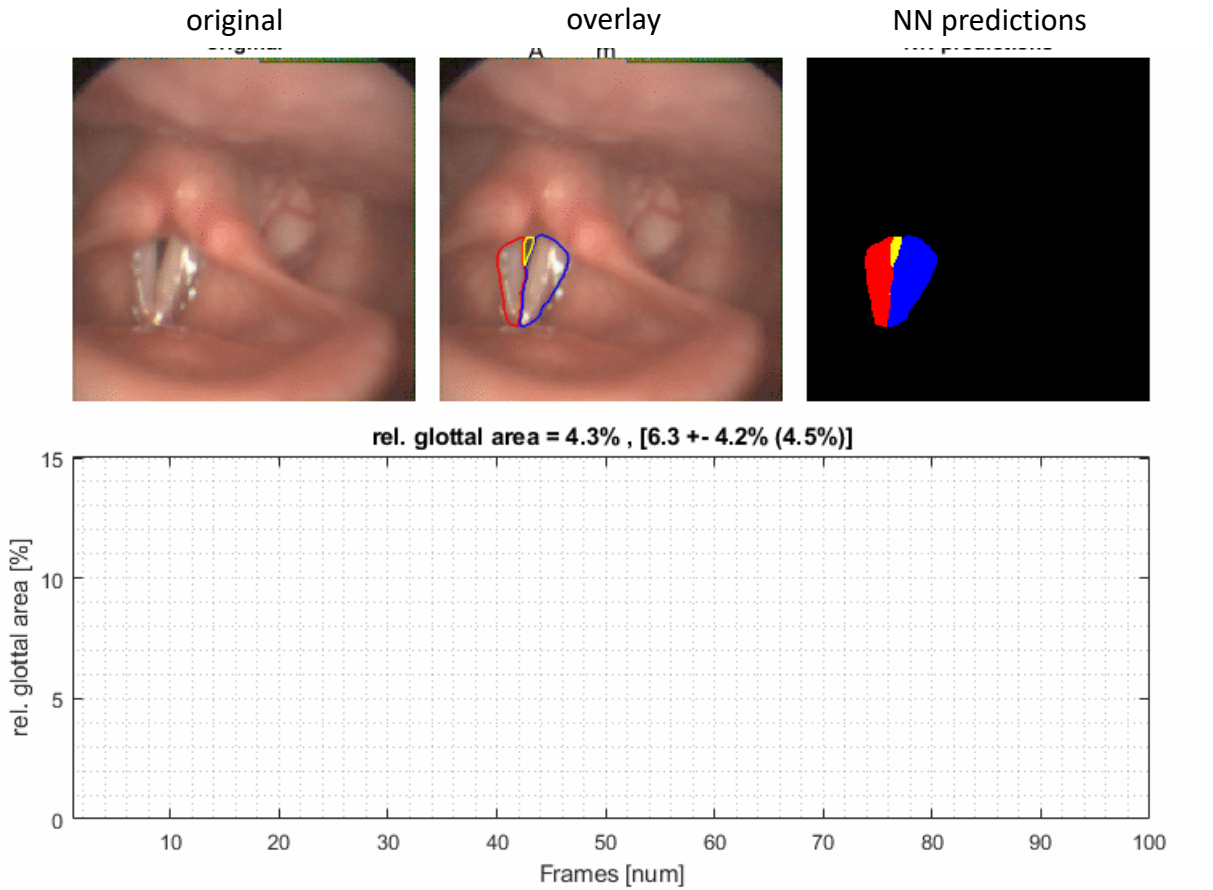
(d) mean and standard deviation for the normalized relative area between the vocal folds for the entire sequence, equal to the sum of the area between the vocal folds + right vocal fold + left vocal fold.

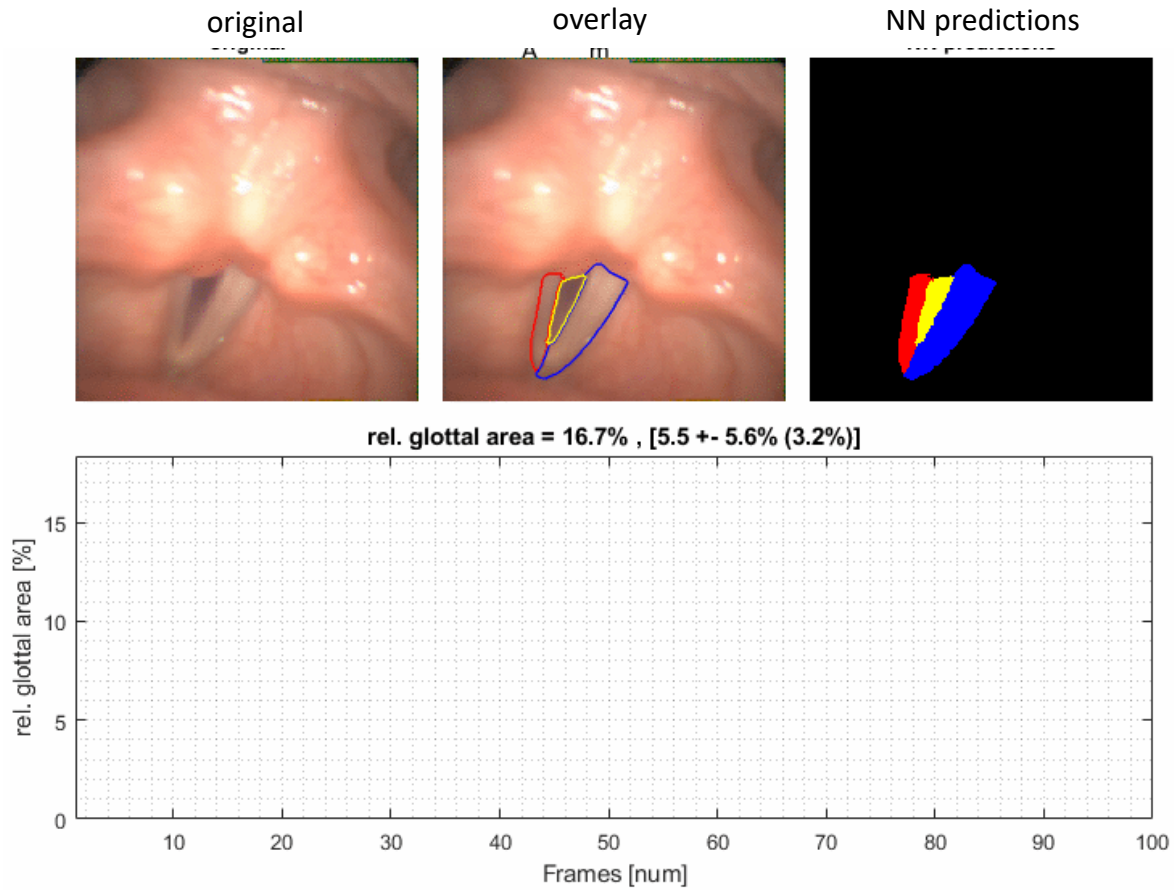
The yellow lines indicate where the pictures are taken in the oscillation.



Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

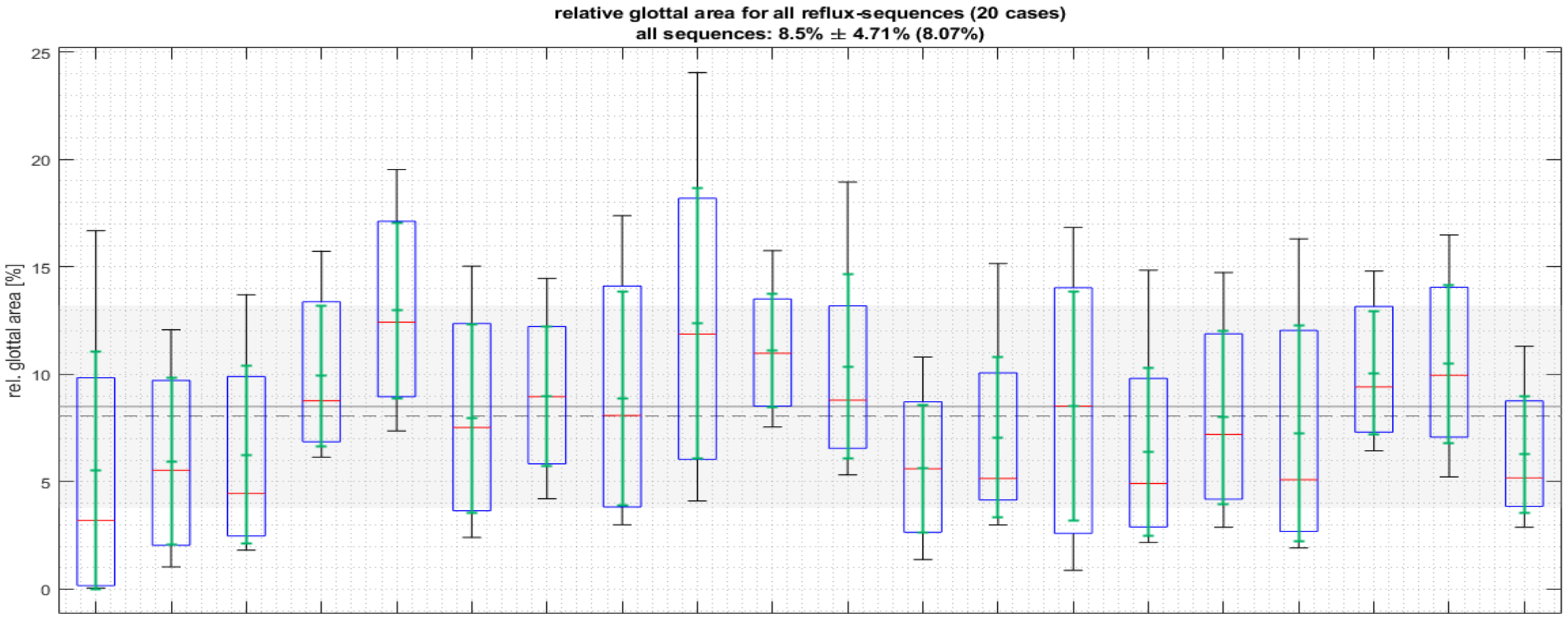
Deep learning software for analysis of the distance between vocal folds





Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

Analysis of the distance between vocal folds



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

Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

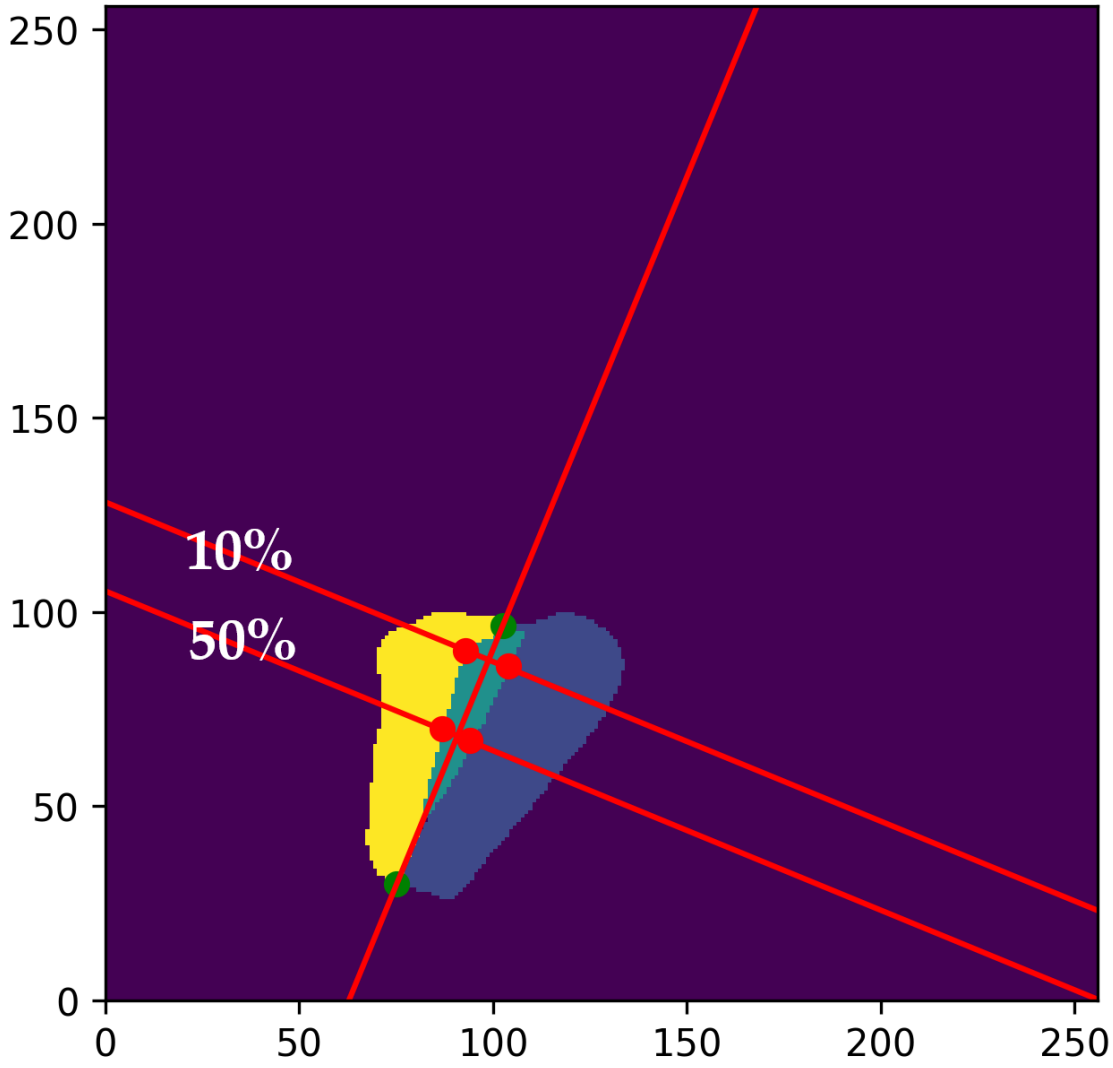
Post network calculations

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.

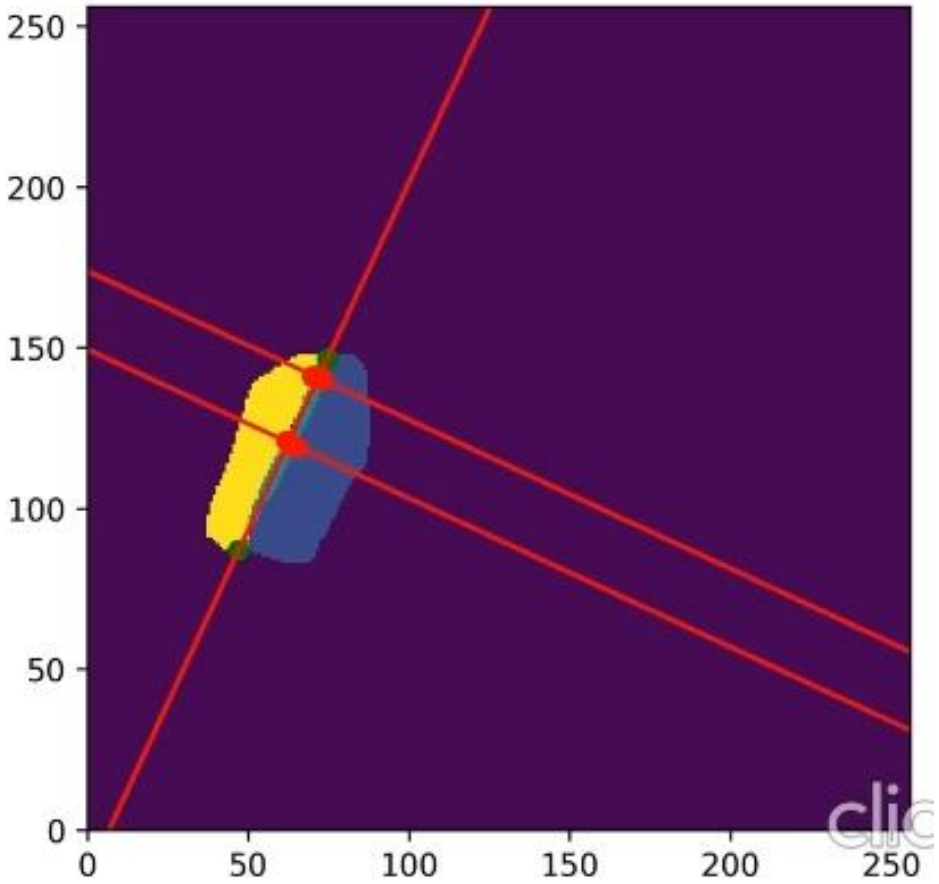


Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

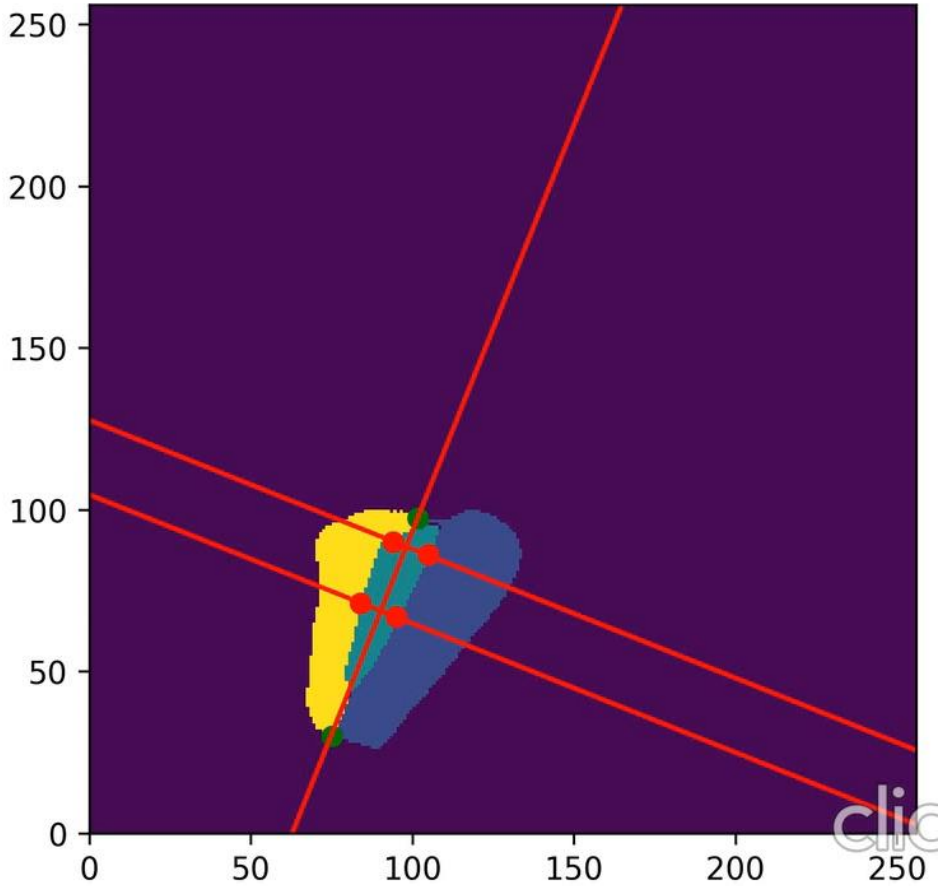
Post network calculations - Video

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

Insufficient closure of the rear



Full closure



Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

Result of deep learning analysis

	Parameter		N	n	LS Mean	Standard Error	95% Confidence Interval	P-Value
a	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).			**2.77**	0.42	(1.93 - 3.62)	< 0.0001
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).			**0.26**	0.39	(-0.52 - 1.04)	0.50

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.

Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

3 key take-aways

1. There is now constructed a UHR-OCT setup that can combine to HSV during phonation (4.000 phrames per second). 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.

2. 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. This can be supported with deep learning because of the big amount of data, and among others also for defining the region of interest.

3. Deep learning is necessary to analyze larger amounts of HSV data. Calculations have been made for many years, manually. 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.

Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

Thank you

Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

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

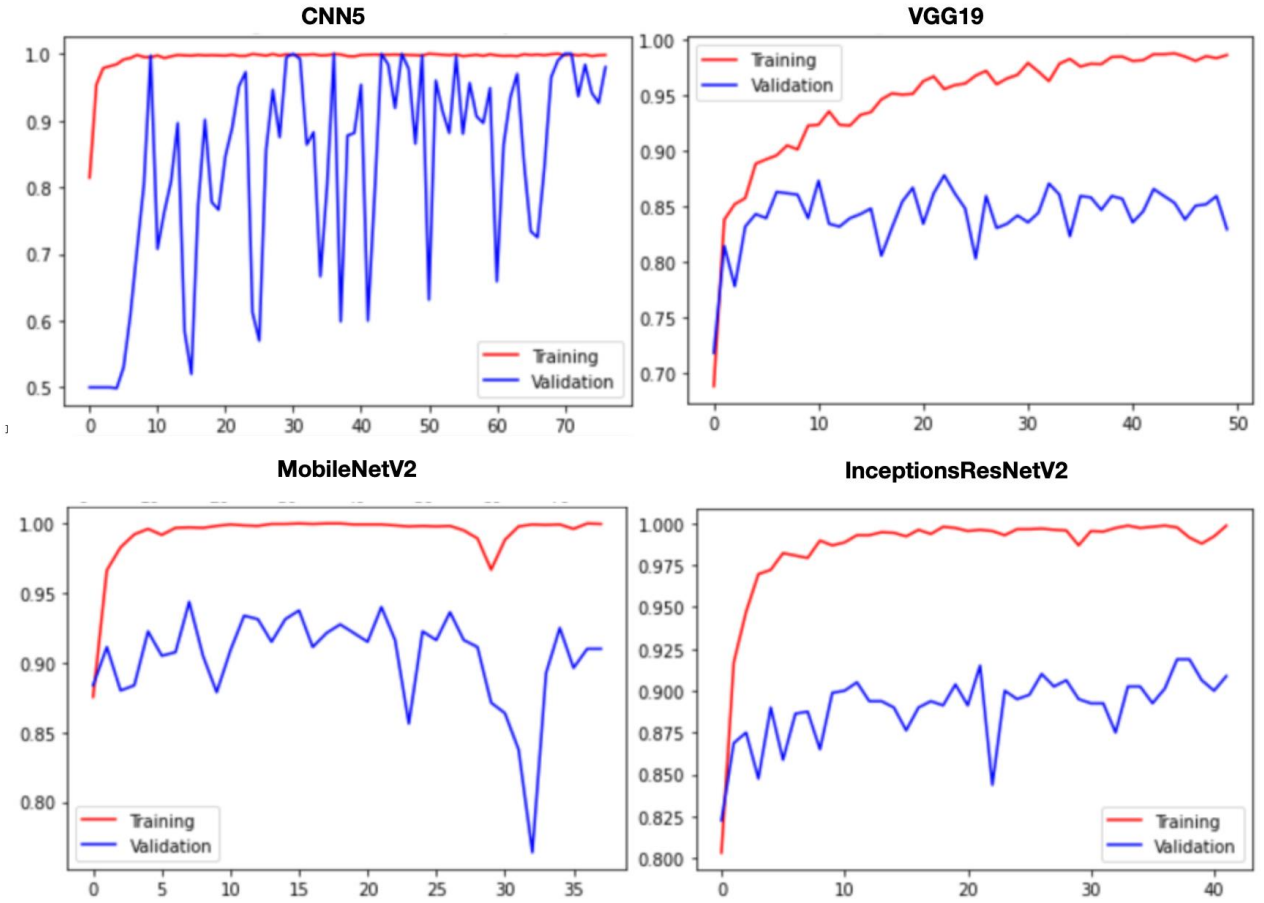
Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

Deep learning – 4 examined Convolutional Neural Networks (CNN)

4 Convolutional Neural Networks (CNN) were compared for their ability to classify vocal nodules. To test if CNN can aid in routine diagnostics.

The picture displays the accuracy during the training of 4 Convolutional Neural Networks (CNN). The accuracy is both measured on the training data and on a separate data set for validation.

CNN5 is a custom CNN with 5 layers, which proved to have the highest accuracy visualized in red and blue lines



Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

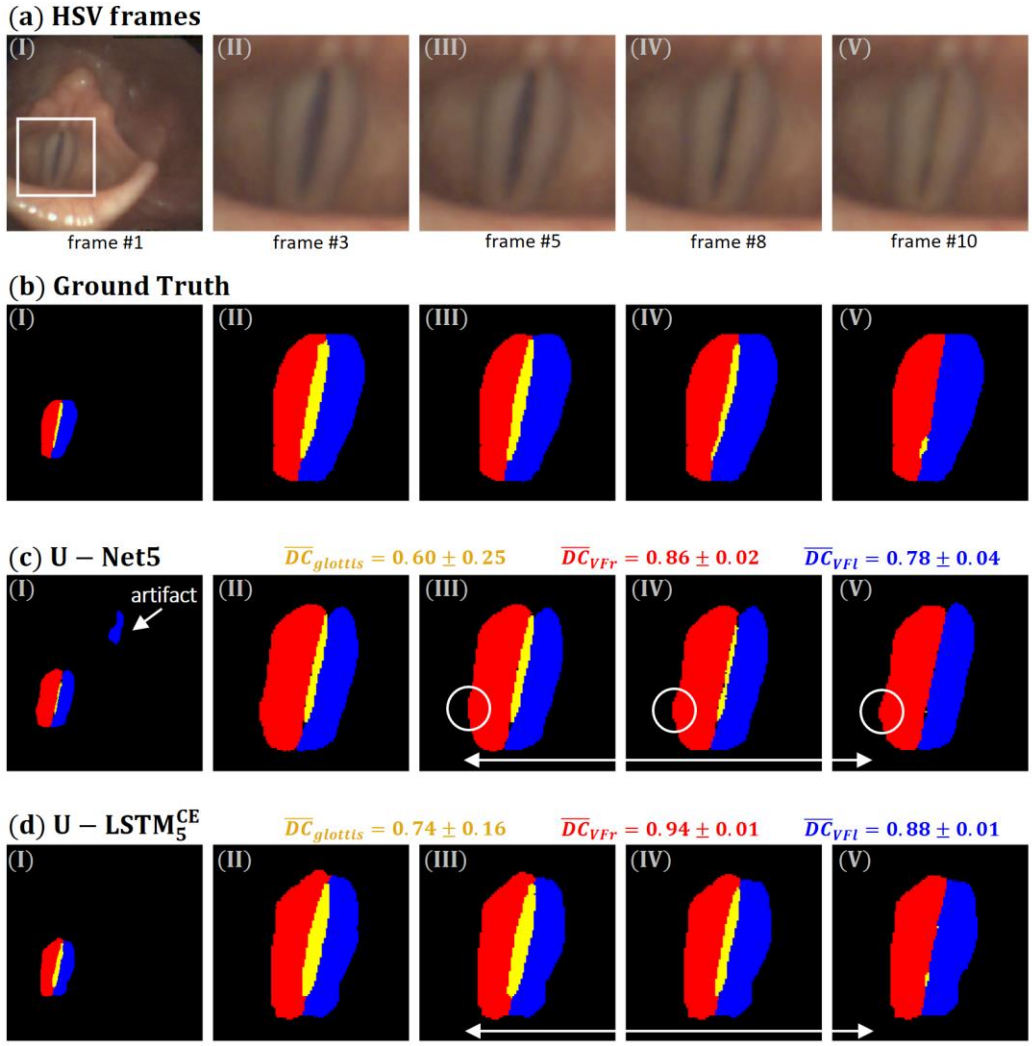
Deep learning software is needed for bigger amount of data

Deep learning is necessary for a large amount of data e.g., analysis of the distance between vocal folds

- Mona Fehling et al. made a comparison of deep learning network segmentation results of individual images of a normal subject.

- **a)** High-Speed Video, 4 out of 100
  b) Ground Truth,
- **c)** U-Net segmentation,
  d) U-LSTM segmentation.

- The dice coefficient for each class represents the mean and standard deviation for the whole sequence of 100 frames



Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org