Technical risk and bias of high-speed video and stroboscopy

Mette Pedersen MD PhD

Round Table – Laryngology

October 30, 09:00 – 10:30

Deep learning applied to laryngeal evaluation

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

High-Speed Videos

Pedersen M, Larsen CF. (2021) *Accuracy of Laryngoscopy for Quantitative Vocal Fold Analysis in Combination with AI, A Cohort Study of Manual Artefacts*, Scholarly Journal of Otolaryngology 6(3): DOI: 10.32474/SJO.2021.06.000237

Videostroboscopy (our own experience and references)

Phoniatrics 1. (2020) *Fundamentals of Voice Disorders, Disorders of Language and Hearing Development*. Springer.

Woo, P. (2021) *Stroboscopy and High-Speed Imaging of Vocal Function* (2nd edition). Plural publishing.

Future considerations

Paderno A, Gennarini F, Sordi A, Montenegro C, Lancini D, Villani FP, Moccia S, Piazza C. (2022) *Artificial intelligence in clinical endoscopy: Insights in the field of videomics*. Front Surg. Sep 12;9:933297. DOI: 10.3389/fsurg.2022.933297.

Israelsen N, Larsen CF, Pedersen M. (2022) [*Quantitative examination of vocal cords with high-speed video and optical coherence tomography*]. Ugeskr Laeger. Feb 7;184(6):V02210146. Danish. PMID: 35179120.

Yousef AM, Deliyski DD, Zacharias SRC, de Alarcon A, Orlikoff RF, Naghibolhosseini M. (2021) *A Hybrid Machine-Learning-Based Method for Analytic Representation of the Vocal Fold Edges during Connected Speech*. Appl Sci (Basel). Feb;11(3):1179. DOI: 10.3390/app11031179.

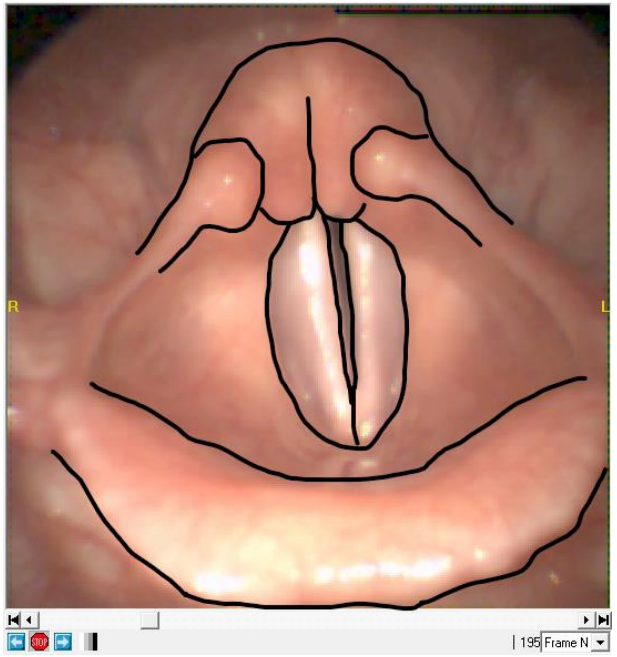
Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

A cohort of high-speed video endoscopies was evaluated to elucidate - how many of our videos could be used for collecting quantitative data, in our case for deep learning.

The important areas of the larynx for evaluation of vocal fold function are drawn in the picture below

Full view of the:

- Vocal folds
- Arytenoids
- Epiglottis
- False vocal folds
- …

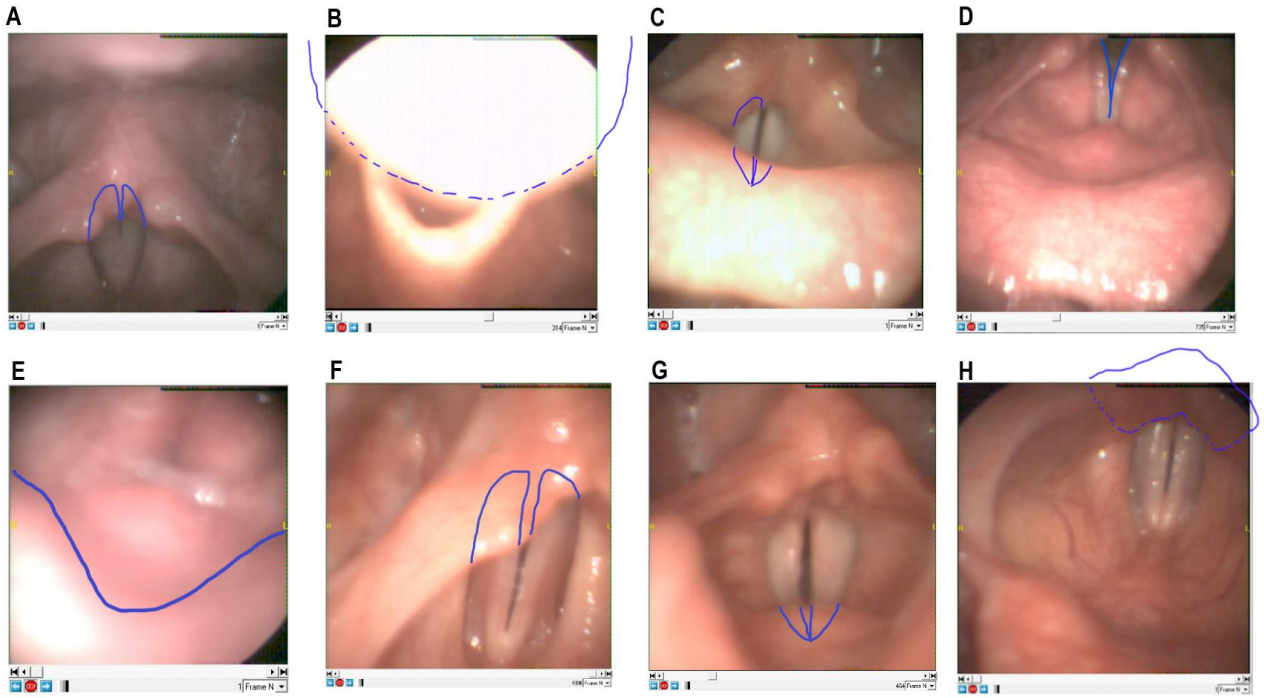


Medical Research Center, Falkoner Allé 80, Copenhagen F, Denmark
mettepedersen.org

Various reasons shown for the videos to be unusable for deep learning

The most common reasons were these 8:

A. Rear part of the vocal folds is not seen
B. Epiglottis and/or uvula blocking vision
C. Various parts of the vocal folds are not seen
D. No vibration of the vocal folds
E. Persistent constricted larynx
F. Pictures are taken from an oblique angle
G. Front part of the vocal folds is not seen
H. Part of the arytenoid region is not seen

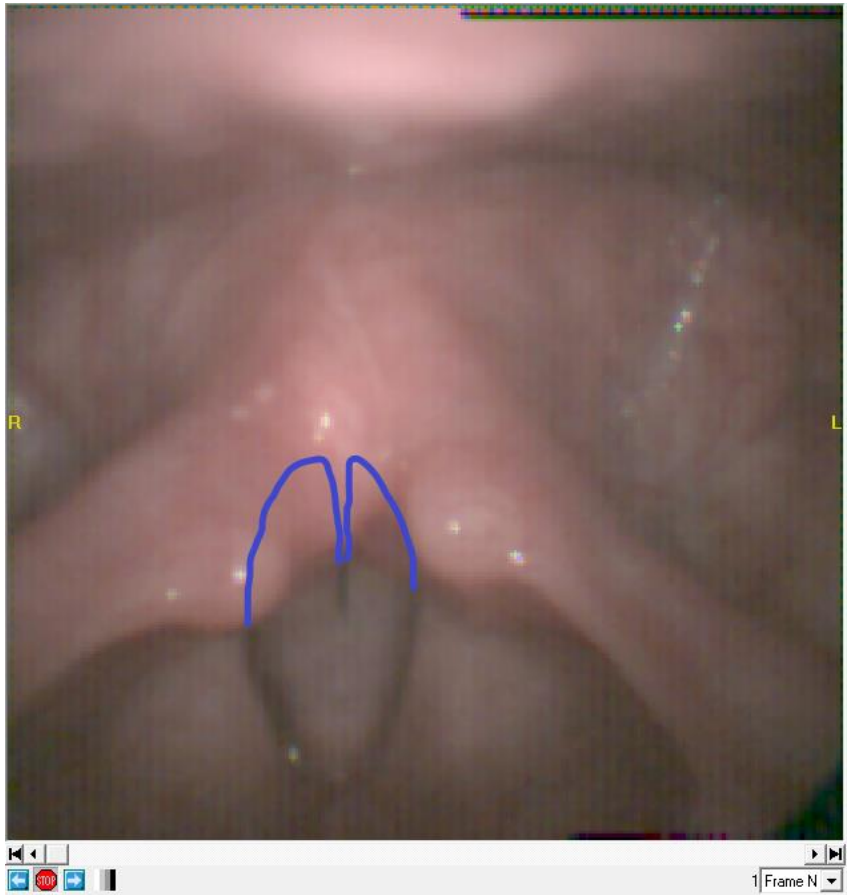


Each reason will be elaborated

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

Rear part of the vocal folds is not seen

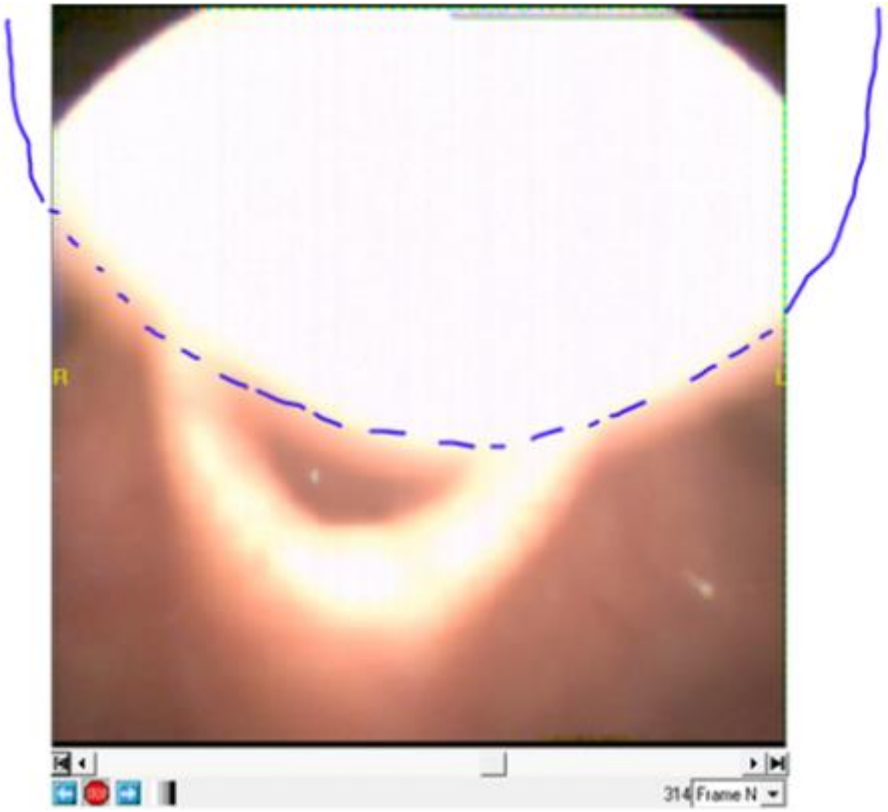
(We attempted to draw the missing rear part of the vocal folds for visualization purposes)



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

Epiglottis and/or Uvula blocking vision

(We attempted to draw uvula for visualization purposes)



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

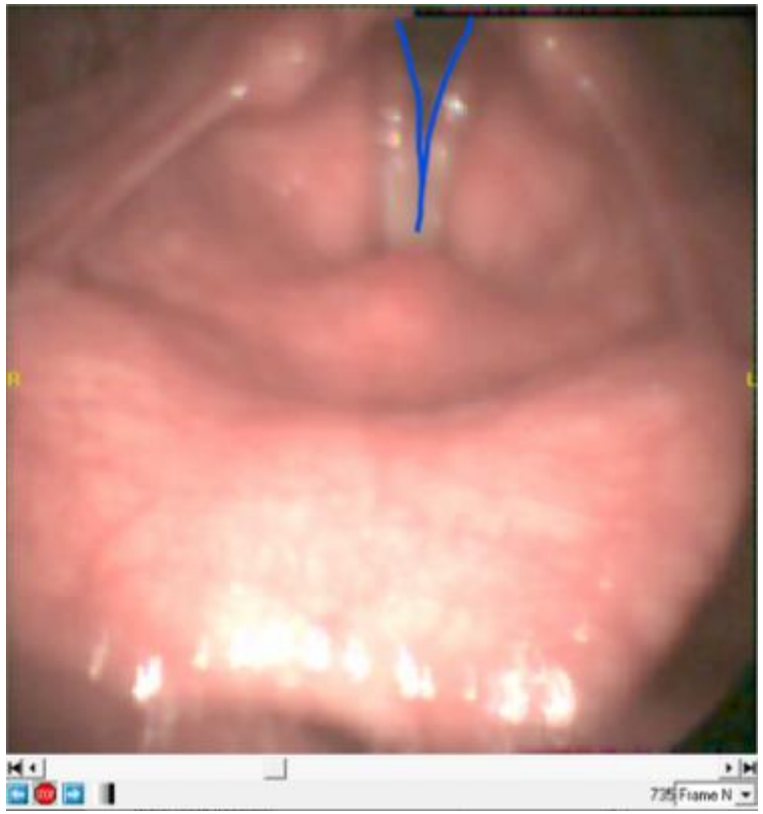
Various parts of the vocal folds are not seen (front and rear)

(We attempted to draw the missing front part of the vocal folds for visualization purposes)

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

No vibration of the vocal folds

(We attempted to draw the edges of the vocal folds for visualization purposes)



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

Persistent constricted larynx, vocal folds not seen

(We attempted to draw the top of epiglottis for visualization purposes)

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

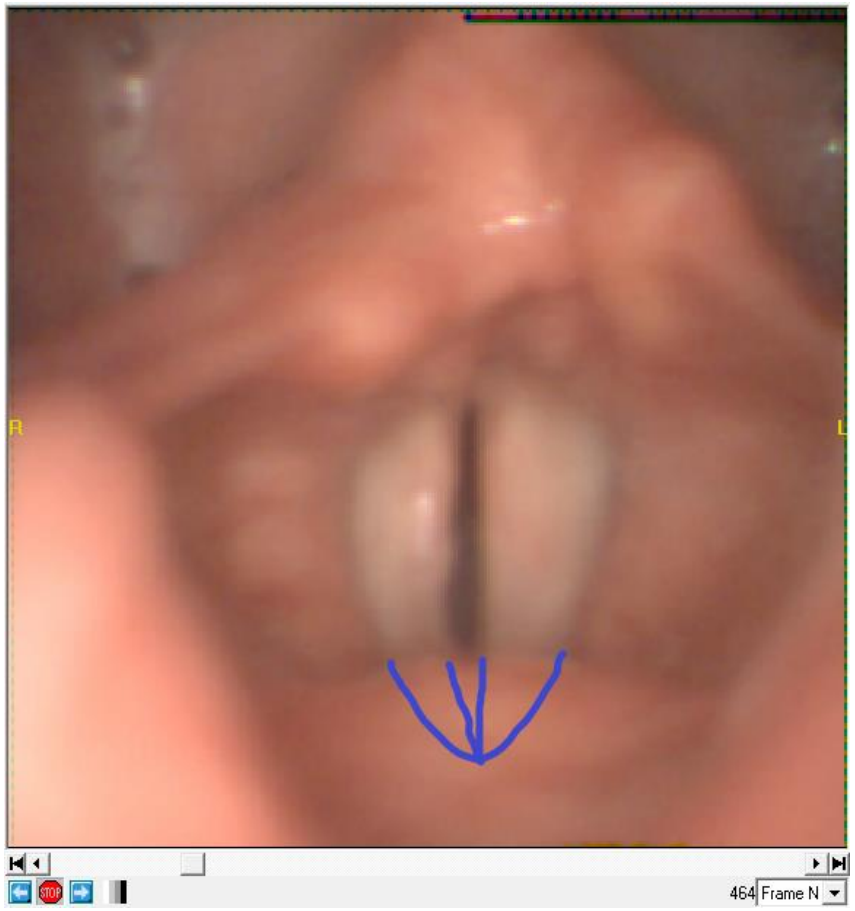
Pictures are taken from an oblique angle

(We attempted to draw the missing part of the vocal folds for visualization purposes)

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

The front part of the vocal folds is not seen

(We attempted to draw the missing part of the vocal folds for visualization purposes)



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

Part of the arytenoid region is not seen

(We attempted to draw the missing area of the arytenoid region for visualization purposes)

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

Statistics

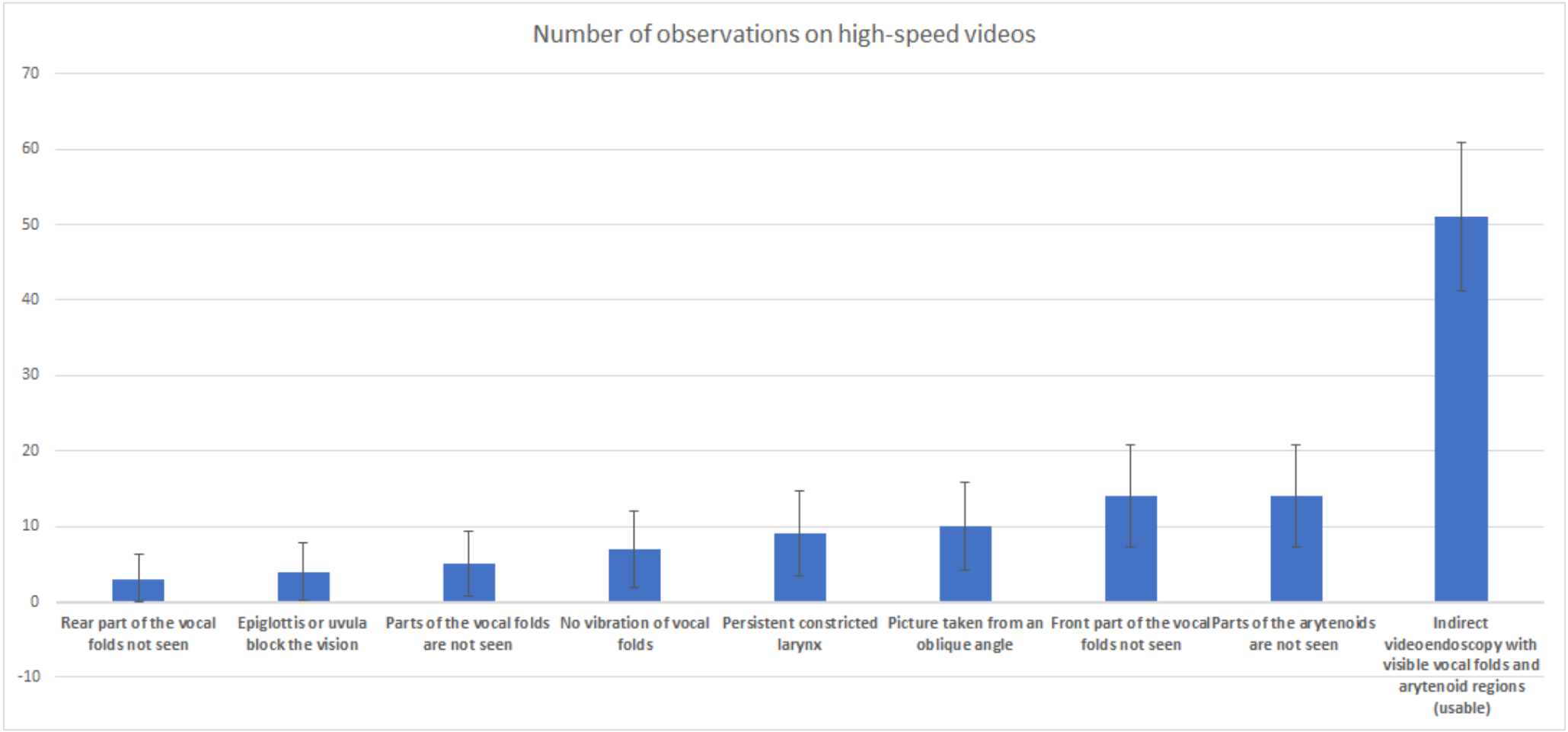
Assuming the assessments are independent with regard to whether there is a finding,
the total number of assessments with a given finding is binomially (usable / non-usable) distributed.

With 100 assessments, an observed incidence of 1, 10, and 25 findings
will result in estimated 95% confidence intervals of [0%-3%], [4%-16%] and [17%-33%] ect, respectively.
--
95% confidence intervals were calculated with the **Wald test** using the asymptotic Normal distribution assumption of the estimated proportion in the binomial distribution.

Based on these calculations 100 randomized videos were sufficient to be used for calculations.

A randomization was made from our database of **15.732 videos** out of **7.909 patients**,
using a randomization algorithm from a computational platform (**Wolfram Alpha**)
to select 100 videos for the study.

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



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

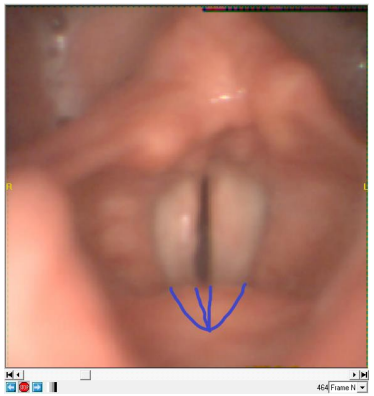
Observations on high-speed videos for the usable versus non-usable videos with 95% confidence intervals, showed that only 51% were usable (Wald 95% ci: 41,2% – 60,8%).

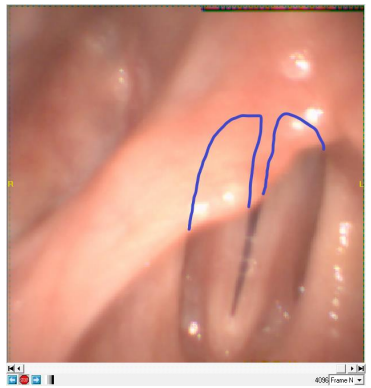
The largest groups of the non-usables are:

1. **Oblique angle pictures (10%)**

2. **Insufficient pictures of the front of the vocal folds (14%)**

3. **Insufficient pictures of the arytenoids (14%)**





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

The cohort study utilized high-speed videos with an average of several hundred pictures - recorded over a period of 12 years, in an otorhinolaryngology medical center.

The HSV was recorded on a sustained vowel /ah/, with the Endocam 5562 equipment from WOLF at a rate of 4.000 frames per second in 256 x 256 pixels (equipment portrayed at the top right corner).

Newer equipment with more pixels and higher frame rates is available (WEVOSYS, portrayed at the bottom right corner).

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

HSV of female patient, aged 38

Full register, F0 = 168Hz, dB = 78

Upper register, F0 = 320Hz, dB = 74

Other function related mistakes are:

1. The use of the upper register, even if the lower was intended
   (the visualization is different in the 2 registers)
2. A sustained phonation is necessary
   Abnormalities must be observed consistently in stroboscopy
   Transient vibratory abnormalities need HSV (or videokymography)
3. Frequency and decibels are not given (often automatic!)

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

The causes for the videos not being usable in our study were mostly not related to the equipment including resolution and framerate - or the patient

but related to the examiner.

Therefore most of the results could be extrapolated to e.g. stroboscopy, NBI, white light laryngoscopy, and video endoscopy**.**

(Personally, if I had known the importance back when the recordings were being made, I could in many cases have made an extra effort and positioned the scope differently to obtain an optimal recording.)

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

The reason for our study to be conducted on high-speed video:

For **functional** evaluation, real-time HSV is preferable
compared with the virtual slow motion as seen in stroboscopy.

video endoscopy and stroboscopy
are mostly sufficient for the **classification** of disorders.

The handling of the picture is the same in all cases.

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

Why are good videos necessary?

Insufficient video films are bias in research

Classification

1. It is underlined that machine learning or deep learning is useful for:
Large cohort studies
Organizing unstructured data.

Function analysis

2. We have suggested combining HSV with ultrahigh-resolution optical coherence tomography.
3. Connected speech assessment based on vocal fold movements has been suggested.

Good diagnostics and research areas require proper data material for evidence-based clinical studies.

1. Paderno A, Gennarini F, Sordi A, Montenegro C, Lancini D, Villani FP, Moccia S, Piazza C. (2022) Artificial intelligence in clinical endoscopy: Insights in the field of videomics. Front Surg. Sep 12;9:933297. DOI: 10.3389/fsurg.2022.933297
2. Israelsen N, Larsen CF, Pedersen M. (2022) [Quantitative examination of vocal cords with high-speed video and optical coherence tomography]. Ugeskr Laeger. Feb 7;184(6):V02210146. Danish. PMID: 35179120
3. Yousef AM, Deliyski DD, Zacharias SRC, de Alarcon A, Orlikoff RF, Naghibolhosseini M. (2021) A Hybrid Machine-Learning-Based Method for Analytic Representation of the Vocal Fold Edges during Connected Speech. Appl Sci (Basel). Feb;11(3):1179. doi: 10.3390/app11031179

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

Thank you for your attention

LINKS

High-Speed Videos (our own study)

Pedersen M, Larsen CF. (2021) *Accuracy of Laryngoscopy for Quantitative Vocal Fold Analysis in Combination with AI, A Cohort Study of Manual Artefacts*, Scholarly Journal of Otolaryngology 6(3): DOI: 10.32474/SJO.2021.06.000237

Videostroboscopy (our own experience and references)

Phoniatrics 1. (2020) *Fundamentals of Voice Disorders, Disorders of Language and Hearing Development*. Springer.

Woo, P. (2021) *Stroboscopy and High-Speed Imaging of Vocal Function* (2nd edition). Plural publishing.

Future aspects

Israelsen N, Larsen CF, Pedersen M. (2022) [*Quantitative examination of vocal cords with high-speed video and optical coherence tomography*]. Ugeskr Laeger. Feb 7;184(6):V02210146. Danish. PMID: 35179120.

Paderno A, Gennarini F, Sordi A, Montenegro C, Lancini D, Villani FP, Moccia S, Piazza C. (2022) *Artificial intelligence in clinical endoscopy: Insights in the field of videomics*. Front Surg. Sep 12;9:933297. DOI: 10.3389/fsurg.2022.933297.

Yousef AM, Deliyski DD, Zacharias SRC, de Alarcon A, Orlikoff RF, Naghibolhosseini M. (2021) *A Hybrid Machine-Learning-Based Method for Analytic Representation of the Vocal Fold Edges during Connected Speech*. Appl Sci (Basel). Feb;11(3):1179. doi: 10.3390/app11031179.