LUPINE PUBLISHERS

Scholarly Journal of Otolaryngology

DOI: 10.32474/SJO.2021.06.000237

ISSN: 2641-1709

Research Article

Accuracy of Laryngoscopy for Quantitative Vocal Fold Analysis in Combination with AI, A Cohort Study of Manual Artefacts

Mette Pedersen[1]*, Christian F Larsen[2]

[1] *Medical Centre, Østergade, Copenhagen, Denmark*

[2] *Copenhagen Business School, Solbjerg Plads, Denmark*

***Corresponding Author:** Mette Pedersen, Medical Centre, Østergade 18, Copenhagen, Denmark

Received: April 01, 2021 **Published:** April 15, 2021

Abstract

Introduction: A cohort of high-speed videoendoscopies was evaluated for usability for deep learning. The aim of our study was to find the percentage of our high-speed videos (15.732) that could be used for deep learning (AI). A screening of the material showed that some videos had artefacts, making them non usable for deep learning.

Material: A randomization was made with Wolfram Alpha random number generator selecting between 15.732 videos from 7.909 patients. The various non usable videos are described including the rear parts of the vocal folds not seen, the epiglottis or uvula blocking vision, parts of the vocal folds not seen, no vibration of the vocal folds, persistent constricted larynx, picture taken from an oblique angle, the front part of the vocal folds not seen, and parts of the arytenoid region not seen.

Method: Assuming the assessments are independent with regards to whether there is a finding, the total number of assessments with a given finding is binomial 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%], respectively. 95% confidence intervals are calculated as Wald test using the asymptotic Normal distribution assumption of the estimated proportion in the binomial distribution. Assuming the incidence of findings for each of the different findings was below 25%, the expected length of the 95% confidence interval is 16%-point (33-17), with 200 and 500 assessments, the corresponding length is 14%-point and 8%-point, respectively. Based on these calculations 100 randomised films were sufficient to be used for calculations.

Results and Conclusion: The prospective cohort study of high-speed videos covered 12 years from the February 2007 to January 2019 in an otorhinolaryngology medical centre. 7.909 patients with a total of 15.732 high-speed video films of the larynx including the vocal folds had been consecutively sampled (4.000 frames per second, Richard Wolf Ltd. endocam 5562). Observations on high-speed video for the usable versus non usable videos with 95% confidence intervals, showed that only 51% were usable. The interesting result is that oblique angle pictures (10%) as well as insufficient pictures of the front of the vocal folds and arytenoids (14%) were the largest groups of the non-usable. They can be augmented by the examiner in the future. Various video and deep learning programs are discussed.

Keywords: Manual artifacts; deep learning; vocal fold analysis; quantitative measures

Abbreviations: AI: Artificial Intelligence; OCT: Optical Coherence Tomography

Introduction

Deep learning, a branch of Artificial Intelligence (AI), is a future possibility for quantitative measurements of the vocal folds, also for documenting treatment effects. We were interested in a co-operation for deep learning since we wanted our prospective cohort of high-speed video endoscopy results to be analysed [1]. The required videos had artefacts in many cases making them non usable for deep learning. Analysing our prospective cohort material of 15.732 videos seemed to be of interest, because several categories of non-usable videos were made based on inspection of the videos. Vocal folds and the arytenoids must be fully visible for optimal evaluation. We have made high-speed videos with stiff scopes for indirect laryngoscopy. But the results are usable for nasal endoscopies of the larynx as well as stroboscopy in combination with AI. A discussion of statistics ended up with a plan for randomisation for calculations. High-speed video is usable to quantify vocal fold measurements [2-9]. Difficulties with light and spatial cameras have been discussed [10-13]. Videos can be affected by patient movement, as well as with nonlinear distortions and phase asymmetry [14,15]. Artifacts, such as parts of the glottis concealed, and parts of the arytenoid cartilage can cover other parts. In the future deep learning might be trained to some extend to take abnormal films into account and correct them, but that will take some time, and there is a risk for bias [16,17]. The aim of our study was to find the percentage of our clinical material of 15.732 high-speed videos that could be used for deep learning (AI) and later Optical Coherence Tomography (OCT). Clinicians should be advised to take the necessary precautions to ensure that laryngoscopies are usable with AI.

Material

A cohort of high-speed videos was evaluated for usability for deep learning. The regularity of the high-speed videos was necessary for the vocal folds. The ideal video included vibrating vocal folds from front to rear without any hidden parts, including well defined arytenoids, relevant for pathology. The distortion could be the following: The vocal folds could be hidden in front or in the rear part. The film could be recorded from an oblique angle. Epiglottis or uvula could block the vision. There could be a persistent constricted larynx. The arytenoid region could be insufficiently presented. Vibration of the vocal folds could be lacking. Recordings of high-speed videos (Figures 1 & 2) were made with HRES Endocam 5562 from Richard Wolf Ltd., 4.000 frames per second and 256 x 256 pixels. 90-degree angle, with stiff scopes used. For randomisation Wolfram Alphas number generator was used. Patients´ high-speed videos were stored in 4 folders and the number generator was used to choose between the folders. For each folder the random number generator was used to generate a random between the total amount of patients in that folder; First folder 1.470 patients, 2.638 videos. Second folder 2.350 patients, 4.790 videos. Third folder 2.232 patients, 4.605 videos. 4th folder 1.857 patients, 3.699 videos. For some patients there were multiple recordings, and the random number generator was used to generate a random number between the multiple recordings. To correlate our study with the literature, we made a search to find comparable studies. The Search strategy in databases for papers on inspection results of larynx videos was made with the library of the Royal Society of Medicine, UK. The planned search words were -- Vocal folds – Larynx – Stroboscopy -- High-speed digital imaging -- Voice assessment – Recording – Examiner—Error-- Video endoscopy -- Observational error -- Measurement error-- Distortion. The final 9th (in italics) search included 13 publications with three from the larynx [14,17,18].

Method

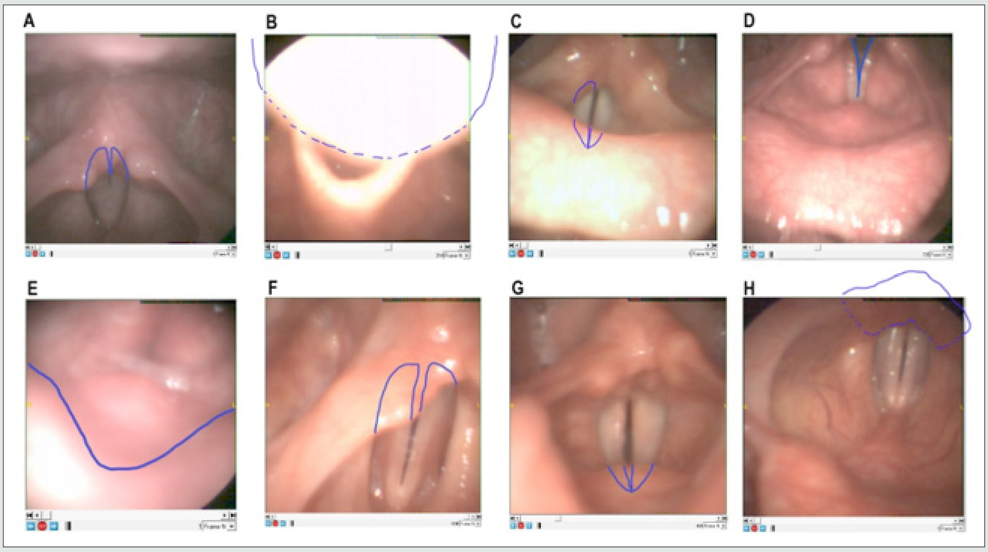


Figure 1: Presentation of distorted videos. A: Rear part of the vocal folds not seen. B: Epiglottis or uvula block the vision. C: Parts of the vocal folds are not seen. D: No vibration of vocal folds. E: Persistent constricted larynx. F: Picture taken from an oblique angle. G: Front part of the vocal folds not seen. H: Parts of the arytenoids are not seen.

Assuming the assessments are independent with regards to whether there is a finding, the total number of assessments with a given finding is binomial 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%], respectively. 95% confidence intervals are calculated as Wald

Citation: Mette Pedersen*, Christian F Larsen. Accuracy of Laryngoscopy for Quantitative Vocal Fold Analysis in Combination with AI, A Cohort Study of Manual Artefacts. Sch J Oto 6(3)-2021. SJO. MS.ID.000237. DOI: 10.32474/SJO.2021.06.000237

test using the asymptotic Normal distribution assumption of the estimated proportion in the binomial distribution. Assuming the incidence of findings for each of the different findings was below 25%, the expected length of the 95% confidence interval is 16%-point (33-17), with 200 and 500 assessments, the corresponding length is 14%-point and 8%-point, respectively. Based on these calculations 100 randomised films were sufficient to be used for calculations. The randomisation was made as mentioned, with Wolfram Alpha random number generator selecting between 15.732 videos from 7.909 patients. The total size of all recordings was 515 GB. To evaluate whether a video was usable, two experienced observers went through the randomised recordings and carefully categorized each finding. The normal findings included clear presentations of the different parts of the vocal folds and arytenoid region, including well defined vibrations. To be used in deep learning it is not possible to extrapolate from the findings what the full picture would have shown. To understand the various findings, we have drawn examples of the groups of non-usable videos as presented in Figure 1. Figure 2 shows an example of a usable video, with and without drawing of the relevant areas. Two experienced examiners evaluated each film, in some cases, more than one non usable finding was seen. A comparison of the different non usable groups for age and gender was made, and the number of frames for influence on the results.

Results

The prospective cohort study of high-speed videos covered 12 years from the February 2007 to January 2019 in an otorhinolaryngology medical centre. 7.909 patients with a total of 15.732 high-speed videos of the larynx including the vocal folds had been consecutively sampled (4.000 frames per second, Richard Wolf Ltd. endocam 5562). The statistical method of Wald test 95% confidence intervals allowed using 100 randomised videos to extrapolate from the total material. Using Wolfram Alphas random number generator for randomisation and JMP 16, 2021 (SAS institute) for statistical calculations, the following results from age, gender, number of frames were found. The mean age was 44 years for groups in total, range 9 - 82y, (CI 95% 40,6 – 47,5, Std 17,2) Figure 3. The mean age for men was 46 years, range 9 - 74y, (CI 95% 40,5 – 51,6, Std 15,5), the mean age for women was 43 years, range 13 - 82y, (CI 95% 38,6 – 47,5, Std 18,1) Figure 4. With ANOVA not statistically difference for age groups and number of frames was found, Figure 5. One way analysis of age groups by usable versus non usable videos are presented and shows no statistical difference between groups, Figure 6. Age distribution for usable versus non usable videos is shown in Figure 7. In an analysis of usability versus gender, for the total amount of female videos and the total amount of male videos. 56,72% of female and 39,39% of male were usable. Fisher's exact 2-sided test (0,1369) showed no significant difference between genders (Table 1). To check if videos improved with more than one examination, a hypothesis was that patients became more familiar with the equipment and relaxed more, allowing for a better recording to be made. Another hypothesis was that as patients' symptoms improved, it became easier to get a good recording. An analysis of usable by number of videos in one patient (1st, 2nd, 3rd etc examination), Pearsons test (0,8185) showed no statistical difference between groups (Table 2). The usable videos include regular movements of the vocal folds, clear presence of the arytenoids and the false vocal folds (Figure 2).

Table 1: Contingency analysis of usability versus gender. For the total amount of female videos (F) and the total amount of male videos (M) is shown how many that were usable (Y) in percentage and how many that were non usable (N). Fisher's exact 2-sided test (0,1369) showed no statistical difference between groups (JMP 16, 2021 SAS institute).

Contingency Analysis of Usable by Examination number			
Contingency Table			
Count Row %	N	Y	Total
F	29 43,28	38 56,72	67
M	20 60,61	13 39,39	33
Total	49	51	100
Tests			
N	DF	-Log Like	R Square (U)
100	1	1,3344020	0.0193
Test		**Chi Square**	**Prob>ChiSq**
Likelihood Ratio		2,669	0,1023
Pearson		2,665	0,1032
Fisher's Exact Test		**Prob**	**Alternative Hypothesis**
Left		0,0781	Prob (Usable=Y) is greater for gender = F than M
Right		0,9676	Prob (Usable=Y) is greater for gender = M than F
2-Tail		0,1369	Prob (Usable=Y) is different across gender

Table 2: Contingency analysis of usable (Y) and non-usable (N) by video examination number of one patient (1st, 2nd, 3rd etc examination), Pearsons test (0,8185) showed no statistical difference between groups (JMP 16, 2021 SAS institute).

Contingency Analysis of Usable by Examination Number			
Contingency Table			
Count Row %	N	Y	Total
1	37 51,39	35 48,61	72
2	7 41,18	10 58,82	17
3	4 50,00	4 50,00	8
4	1 50,00	1 50,00	2
6	0 0,00	1 100,00	1
Total	49	51	100
Tests			
N	DF	-Log Like	R Square (U)
100	4	0.96702439	0.0140
Test		**Chi Square**	**Prob>ChiSq**
Likelihood Ratio		1,934	0,7479
Pearson		1,546	0,8185

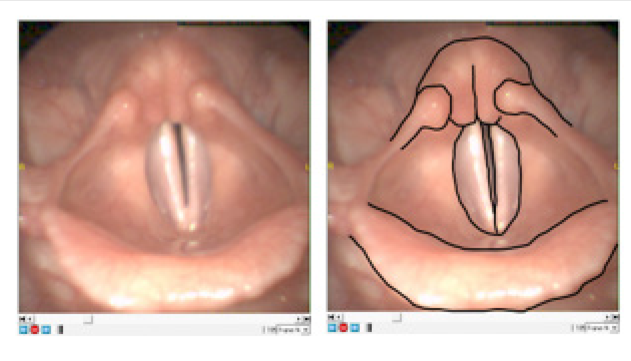
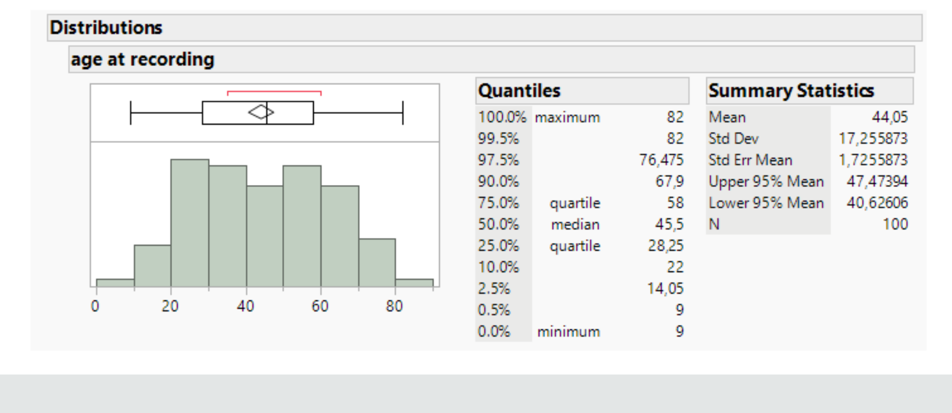


Figure 2: Clinical high-speed video with visible vocal folds and arytenoid regions (usable).



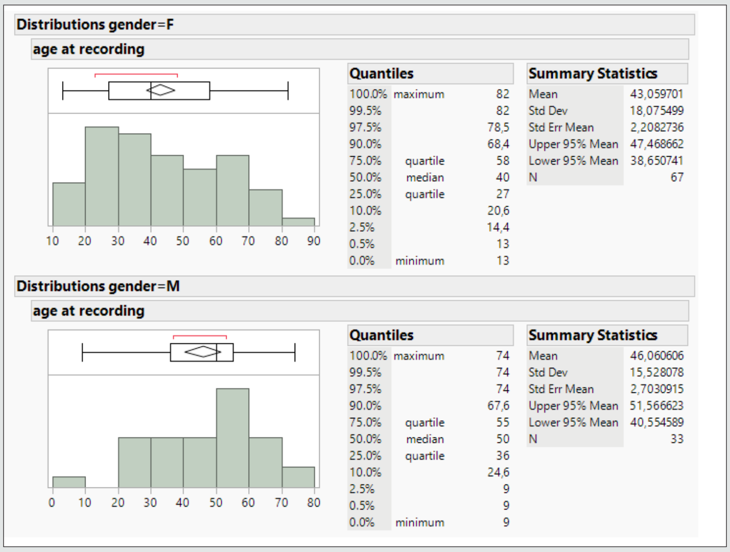
Distributions

age at recording

Quantiles		
100.0%	maximum	82
99.5%		82
97.5%		76,475
90.0%		67,9
75.0%	quartile	58
50.0%	median	45,5
25.0%	quartile	28,25
10.0%		22
2.5%		14,05
0.5%		9
0.0%	minimum	9

Summary Statistics	
Mean	44,05
Std Dev	17,255873
Std Err Mean	1,7255873
Upper 95% Mean	47,47394
Lower 95% Mean	40,62606
N	100

Figure 3: The mean age was 44 years for groups in total, range 9 - 82y, (CI 95% 40,6 – 47,5, Std 17,2) (JMP 16, 2021 SAS institute).



Distributions gender=F

age at recording

Quantiles		
100.0%	maximum	82
99.5%		82
97.5%		78,5
90.0%		68,4
75.0%	quartile	58
50.0%	median	40
25.0%	quartile	27
10.0%		20,6
2.5%		14,4
0.5%		13
0.0%	minimum	13

Summary Statistics	
Mean	43,059701
Std Dev	18,075499
Std Err Mean	2,2082736
Upper 95% Mean	47,468662
Lower 95% Mean	38,650741
N	67

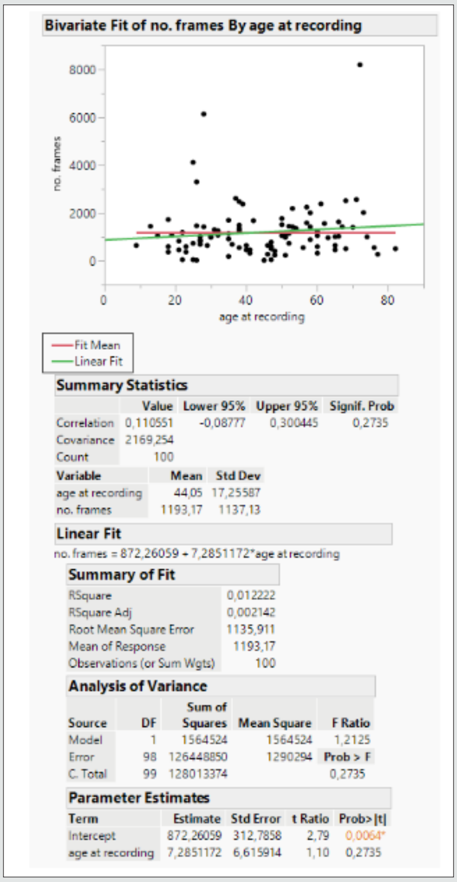
Distributions gender=M

age at recording

Quantiles		
100.0%	maximum	74
99.5%		74
97.5%		74
90.0%		67,6
75.0%	quartile	55
50.0%	median	50
25.0%	quartile	36
10.0%		24,6
2.5%		9
0.5%		9
0.0%	minimum	9

Summary Statistics	
Mean	46,060606
Std Dev	15,528078
Std Err Mean	2,7030915
Upper 95% Mean	51,566623
Lower 95% Mean	40,554589
N	33

Figure 4: The mean age for men (M) was 46 years, range 9 - 74y, (CI 95% 40,5 – 51,6, Std 15,5) (JMP 16, 2021 SAS institute). The mean age for females (F) was 43 years, range 13 - 82y, (CI 95% 38,6 – 47,5, Std 18,1) (JMP 16, 2021 SAS institute).



Bivariate Fit of no. frames By age at recording

Summary Statistics

	Value	Lower 95%	Upper 95%	Signif. Prob
Correlation	0,110551	-0,08777	0,300445	0,2735
Covariance	2169,254			
Count	100			

Variable	Mean	Std Dev
age at recording	44,05	17,25587
no. frames	1193,17	1137,13

Linear Fit

no. frames = 872,26059 + 7,2851172*age at recording

Summary of Fit

RSquare	0,012222
RSquare Adj	0,002142
Root Mean Square Error	1135,911
Mean of Response	1193,17
Observations (or Sum Wgts)	100

Analysis of Variance

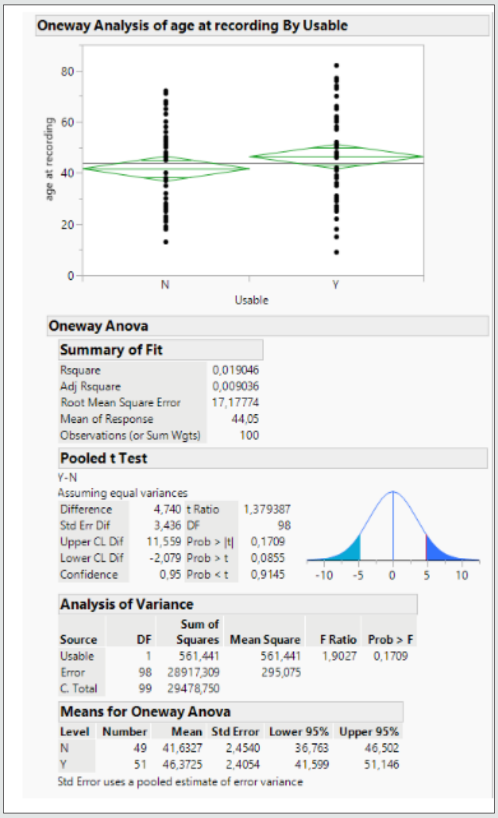
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1564524	1564524	1,2125
Error	98	126448850	1290294	Prob > F
C. Total	99	128013374		0,2735

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	872,26059	312,7858	2,79	0,0064*
age at recording	7,2851172	6,615914	1,10	0,2735

Figure 5: Bivariate fit for number of frames by age is presented, ANOVA showed no statistically significant difference for age groups and number of frames (Prob > F = 0,2735) (JMP 16, 2021 SAS institute).

Observations on high-speed video for the usable versus non usable videos with 95% confidence intervals, showed that only 51% were usable. The interesting result is that oblique angle pictures (10%) and insufficient pictures of the front of the vocal folds (14%) and arytenoids (14%) were the largest groups of the non-usable. They can be augmented by the examiner in the future. Another interesting result was that so many had persistent constricted larynx (9%). It should be noted that vibration of the vocal folds (7%) should be secured by the examiner if possible. In some cases, parts of the vocal folds were not visible (5%) the laryngoscope not being centered due to among others anatomical variance. It was also noted that in some cases the epiglottis or uvula blocked the vision (4%). In our study we included an overview of the cases where arytenoids were insufficiently visible (14%). This is of special interest in mucosal disorders of the larynx (e.g., reflux, allergy, infection, etc.) Figure 8. Rear part of the vocal folds not seen 3 % (Wald 95% ci: 0% – 6,3%), Epiglottis or uvula block the vision 4% (Wald 95% ci: 0,1% - 7,8%), Parts of the vocal folds are not seen 5% (Wald 95% ci: 0,7% - 9,3%), No vibration of vocal folds 7% (Wald 95% ci: 2% - 12%), Persistent constricted larynx 9% (Wald 95% ci: 3,4% - 14,6%), Picture taken from an oblique angle 10% (Wald 95% ci: 4,1% - 15,9%), Front part of the vocal folds not seen 14% (Wald 95% ci: 7,2% - 20,8%), Parts of the arytenoids are not seen 14% (Wald 95% ci: 7,2% - 20,8%), Indirect video endoscopy with visible vocal folds and arytenoid regions (usable) 51% (Wald 95% ci: 41,2% – 60,8%).



Oneway Analysis of age at recording By Usable

Oneway Anova

Summary of Fit

Rsquare	0,019046
Adj Rsquare	0,009036
Root Mean Square Error	17,17774
Mean of Response	44,05
Observations (or Sum Wgts)	100

Pooled t Test

Y-N

Assuming equal variances

Difference	4,740	t Ratio	1,379387
Std Err Dif	3,436	DF	98
Upper CL Dif	11,559	Prob > |t|	0,1709
Lower CL Dif	-2,079	Prob > t	0,0855
Confidence	0,95	Prob < t	0,9145

Analysis of Variance

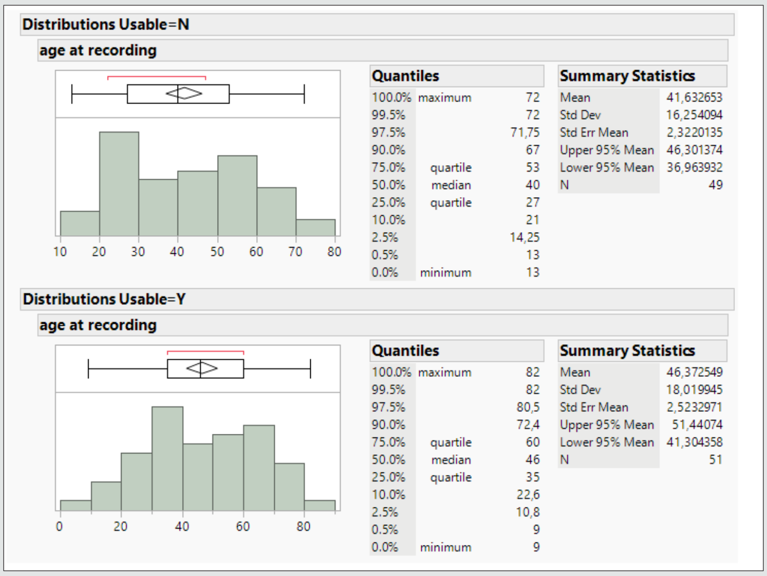
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Usable	1	561,441	561,441	1,9027	0,1709
Error	98	28917,309	295,075		
C. Total	99	29478,750			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
N	49	41,6327	2,4540	36,763	46,502
Y	51	46,3725	2,4054	41,599	51,146

Std Error uses a pooled estimate of error variance

Figure 6: One way analysis of age groups at recording by usable versus non usable video are presented and shows no statistical difference between groups (Prob > F 0,17) (JMP 16, 2021 SAS institute).



Distributions Usable=N

age at recording

Quantiles

100.0%	maximum	72
99.5%		72
97.5%		71,75
90.0%		67
75.0%	quartile	53
50.0%	median	40
25.0%	quartile	27
10.0%		21
2.5%		14,25
0.5%		13
0.0%	minimum	13

Summary Statistics

Mean	41,632653
Std Dev	16,254094
Std Err Mean	2,3220135
Upper 95% Mean	46,301374
Lower 95% Mean	36,963932
N	49

Distributions Usable=Y

age at recording

Quantiles

100.0%	maximum	82
99.5%		82
97.5%		80,5
90.0%		72,4
75.0%	quartile	60
50.0%	median	46
25.0%	quartile	35
10.0%		22,6
2.5%		10,8
0.5%		9
0.0%	minimum	9

Summary Statistics

Mean	46,372549
Std Dev	18,019945
Std Err Mean	2,5232971
Upper 95% Mean	51,44074
Lower 95% Mean	41,304358
N	51

Figure 7: Age distribution for usable (Y) versus non usable (N) videos is shown (JMP 16, 2021 SAS institute).

The statistical advice to randomise 100 films illustrates our point sufficiently. The results show that it is necessary for the examiner to focus on optimising the recordings so that they can be used for quantification and AI. In the literature search we found 3 larynx related papers in S9 out of the 13 papers in total. Ghasemzadeh and Deliyski discuss fiberoptic flexible endoscope distortions on the calibration of images acquired by the laser production system. The first one being from the wide-angle lens with higher spatial resolution in the center of the field of view, the second one being from the variation in the imaging angle. [14]. Adamain N, Naunheim M and Jowett N discus an automatic quantitative tracking of vocal fold motion from video laryngoscopy focusing on the glottal opening angles [17]. A thesis by Deng J. concludes that use of the camera framerate, spatial resolution and angle of view can all modify the resulting video of the vocal folds, and various algorithms are discussed [18]. Based on the results it must be underlined that the importance of good quality videos is prerequisite for any statistical quantitative evaluation of the vocal folds. The results show that this is probably not the case in daily clinical work.

Discussion

The background for this study was a potential collaboration on deep learning that initially required 20 normal videos, which proved to be challenging [1]. Statistically the randomisation of 100 videos of 15.732 was sufficient to find the percentage of usable videos for deep learning. 51% of the videos were usable for reproduction with deep learning. Some of the non-usable videos could be augmented by the examiner. This is the case in Figure 8 with adjustment of the laryngoscope for the pictures taken from an angle, the front and the rear of the vocal folds not seen, and the part of the arytenoids not seen. In a few cases the larynx will remain constricted, or epiglottis or uvula block the vision. It is noted statistically that the age and gender distribution for males and females are not significantly different in the study. In Figure 8 the indirect videoendoscopy with visible vocal folds and arytenoid regions (usable) were 51% (Wald 95% ci: 41,2% – 60,8%). It is noted that the examination number does not influence the probability of it being usable. A higher percentage of women were usable than men, but not enough for a statistical significance difference in this study (Tables 1 & 2). There is an ongoing discussion of video stroboscopy and high-speed videoendoscopy as for evaluation of amplitude and edge of mucosal wave and left-right phase asymmetry [19]. A possible solution has been suggested for consideration of artefacts [20]. Various deep learning software are discussed [21-23]. Since stroboscopy is a major diagnostic clinical tool for functional larynx evaluation, its use in deep learning and optical coherence tomography probably must be elucidated more in the future [24,25].

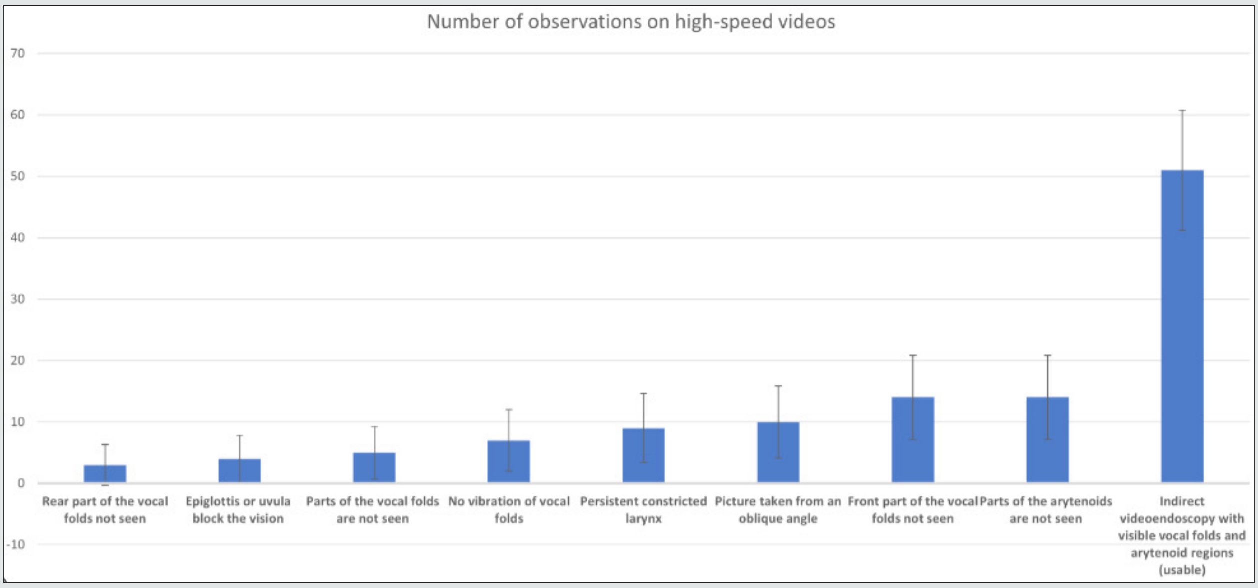


Figure 8: Bar chart of observations on high-speed videos for the usable versus non usable videos with 95% confidence intervals. What is notable is that only 51 are usable (Excel version).

Conclusion

Only 51% of high-speed videos in a clinical setting were sufficient with full pictures of the vocal folds and arytenoids. There is a need for clinicians to focus on optimization of videos while recording for use with quantitative measurements and deep learning, this is also the case for optical coherence tomography. The discussion of videostroboscopy versus high-speed video is in the favor of high-speed, since stroboscopy pictures does not include all consecutive vocal fold movements.

References

1. Fehling MK, Grosch F, Schuster ME, Schick B, Lohscheller J (2020) Fully automatic segmentation of glottis and vocal folds in endoscopic laryngeal high-speed videos using a deep Convolutional LSTM Network. PLoS ONE 15(2): e0227791.

2. Pedersen M (2019) Ultra-High-Resolution (UHR) Optical Coherence Tomography (OCT) in the Upper Airways: Aspect of Combined High-Speed Films and UHR OCT in the Larynx. International Journal of Clinical & Experimental Otolaryngology 5(1):101-105.

3. Zacharias SRC, Deliyski DD, Gerlach TT (2018) Utility of Laryngeal High-speed Videoendoscopy in Clinical Voice Assessment. Journal of Voice 32(2): 216-220.

4. Gandhi S, Bhatta S, Ganesuni D, Ghanpur AD, Saindani SJ (2021) High-speed videolaryngoscopy in early glottic carcinoma patients following transoral CO2 LASER cordectomy. European Archives of Oto-Rhino-Laryngology: and Head & Neck 278(4): 1119.

5. Yamauchi A, Yokonishi H, Imagawa H, Sakakibara K, Nito T, et al. (2016) Quantification of Vocal Fold Vibration in Various Laryngeal Disorders Using High-Speed Digital Imaging. Journal of Voice 30(2): 205-214.

6. Watanabe T, Kaneko K, Sakaguchi K, Takahashi H (2016) Vocal-fold vibration of patients with Reinke's edema observed using high-speed digital imaging. Auris Nasus Larynx 43(6): 654-657.

7. Woo P (2017) High-speed Imaging of Vocal Fold Vibration Onset Delay: Normal Versus Abnormal. Journal of Voice 31(3):307-312.

8. Schlegel P, Stingl M, Kunduk M, Kniesburges S, Bohr C, et al. (2019) Dependencies and Ill-designed Parameters Within High-speed Videoendoscopy and Acoustic Signal Analysis. Journal of Voice 33(5): 811.

9. Woo P, Baxter P (2017) Flexible Fiber-Optic High-Speed Imaging of Vocal Fold Vibration: A Preliminary Report. Journal of Voice 31(2): 175-181.

10. Popolo PS (2018) Investigation of Flexible High-Speed Video Nasolaryngoscopy. Journal of Voice 32(5): 529-537.

11. Schlegel P, Kunduk M, Stingl M, Semmler M, Döllinger M, et al. (2019) Influence of spatial camera resolution in high-speed videoendoscopy on laryngeal parameters. PLoS ONE 14(4): e0215168.

12. Döllinger M, Dubrovskiy D, Patel R (2012) Spatiotemporal analysis of vocal fold vibrations between children and adults. The Laryngoscope 122(11): 2511-2518.

13. Patel R, Donohue KD, Unnikrishnan H, Kryscio RJ (2015) Kinematic measurements of the vocal-fold displacement waveform in typical children and adult populations: quantification of high-speed endoscopic videos. Journal of Speech, Language, and Hearing Research 58(2): 227.

14. Ghasemzadeh H, Deliyski DD (2020) Non-Linear Image Distortions in Flexible Fiberoptic Endoscopes and their Effects on Calibrated Horizontal Measurements Using High-Speed Videoendoscopy. Journal of Voice 1997(20): 30331-30333.

15. Powell ME, Deliyski DD, Zeitels SM, Burns JA, Hillman RE, et al. (2020) Efficacy of Videostroboscopy and High-Speed Videoendoscopy to Obtain Functional Outcomes from Perioperative Ratings in Patients with Vocal Fold Mass Lesions. Journal of Voice 34(5): 769-782.

16. Gómez P, Kist AM, Schlegel P, Berry DA, Chhetri DK, et al. (2020) BAGLS, a multihospital Benchmark for Automatic Glottis Segmentation. Scientific Data 7(1).

17. Adamian N, Naunheim MR, Jowett N (2021) An Open-Source Computer Vision Tool for Automated Vocal Fold Tracking from Videoendoscopy. The Laryngoscope 131(1): E219.

18. Deng J (2018) Evaluation of High-Speed Videoendoscopy for Bayesian Inference on Reduced Order Vocal Fold Models 146(2): 1492.

19. Eysholdt U, Rosanowski F, Hoppe, U (2003) Vocal fold vibration irregularities caused by different types of laryngeal asymmetry. Eur Arch Otorhinolaryngol 260: 412–417.

20. Drioli C, Foresti GL (2020) Fitting a biomechanical model of the folds to high-speed video data through bayesian estimation. Informatics in Medicine Unlocked 20(100373).

21. Matava C, Pankiv E, Raisbeck S, Caldeira M, Alam F (2020) A Convolutional Neural Network for Real Time Classification, Identification, and Labelling of Vocal Cord and Tracheal Using Laryngoscopy and Bronchoscopy Video. Journal of Medical Systems 44(2): 44.

22. Ren J, Jing X, Wang J, Ren X, Xu Y, Yang Q et al. (2020) Automatic Recognition of Laryngoscopic Images Using a Deep-Learning Technique. The Laryngoscope 130(11): E686.

23. Schlegel P, Kniesburges S, Dürr S, Schützenberger A, Döllinger M (2020) Machine learning based identification of relevant parameters for functional voice disorders derived from endoscopic high-speed recordings. Scientific Reports 10(1):1-14.

24. Diaz-Cadiz M, McKenna VS, Vojtech JM, Stepp CE (2019) Adductory Vocal Fold Kinematic Trajectories During Conventional Versus High-Speed Videoendoscopy. Journal of Speech, Language, and Hearing Research 62(6): 1685.

25. Maguluri G, Mehta D, Kobler J, Park J, Iftimia N (2019) Synchronized, concurrent optical coherence tomography and videostroboscopy for monitoring vocal fold morphology and kinematics. Biomedical optics express 10(9): 4450–4461.

DOI: 10.32474/SJO.2021.06.000237