



Digital otoscopy videos versus composite images: A reader study to compare the accuracy of ENT physicians

Journal:	<i>The Laryngoscope</i>
Manuscript ID	Draft
Wiley - Manuscript type:	Original Reports
Date Submitted by the Author:	n/a
Complete List of Authors:	Binol, Hamidullah; Wake Forest School of Medicine, Niazi, Muhammad Khalid Khan; Wake Forest School of Medicine Essig, Garth; the Ohio State University, Otolaryngology-Head and Neck Surgery Shah, Jay; Case Western Reserve Univ, ENT Mattingly, Jameson; The Ohio State University, Department of Otolaryngology Harris, Michael; Medical College of Wisconsin, Otolaryngology and Communication Sciences Elmaraghy, Charles; Nationwide Children's Hospital, Teknos, Theodoros; Case Comprehensive Cancer Center, Department of Otolaryngology; Taj-Schaal, Nazhat; The Ohio State University, Department of Internal Medicine Yu, Lianbo; the Ohio State University, Department of Biomedical Informatics Gurcan, Metin; Wake Forest School of Medicine Moberly, Aaron; Ohio State University, Otolaryngology
Keywords - Combo:	Otitis media/ chronic otitis media/conductive hearing loss < Pediatric otology < Pediatrics, Middle ear < Otology/Neurotology, otitis media (OM) < Otology/Neurotology

SCHOLARONE™
Manuscripts

1
2
3 **Title:** Digital otoscopy videos versus composite images:
4 A reader study to compare the accuracy of ENT physicians
5

6 **Running Title:** Otoscopy videos vs composite images
7

8
9 **Name(s) of the author(s):** Hamidullah Binol, PhD¹, M. Khalid Khan Niazi, PhD¹, Garth
10 Essig, MD², Jay Shah, MD³, Jameson K. Mattingly, MD², Michael S. Harris, MD⁴, Charles
11 Elmaraghy, MD², Theodoros Teknos, MD⁵, Nazhat Taj-Schaal, MD⁶, Lianbo Yu, PhD⁷,
12 Metin N. Gurcan, PhD¹, Aaron C. Moberly, MD²
13

14
15 **Affiliations of author(s):**

16 ¹ Center for Biomedical Informatics, Wake Forest School of Medicine, Winston-Salem,
17 NC, USA

18 ² Department of Otolaryngology, Ohio State University, OH, USA

19 ³ Case Western Reserve University School of Medicine, OH, USA

20 ⁴ Otolaryngology and Comm. Sciences - Froedtert Hospital, WI, USA

21 ⁵ University Hospitals Seidman Cancer Center, OH, USA

22 ⁶ Department of Internal Medicine, Ohio State University College of Medicine, OH, USA

23 ⁷ Department of Biomedical Informatics, Ohio State University, OH, USA
24
25

26 **Corresponding author:** Hamidullah Binol

27 **Postal address:** 486 N. Patterson Avenue, Winston-Salem, NC 27101, USA

28 **Mobile number:** (786) 442-5600

29 **Email address:** hbinol@wakehealth.edu

30 **Fax:** (336) 716-0239

31 **ORCID:** <https://orcid.org/0000-0003-2537-8280>
32
33

34 **Funding:** The project described was supported in part by Award R21 DC016972 (PIs:
35 Gurcan, Moberly) from National Institute on Deafness and Other Communication
36 Disorders. The content is solely the responsibility of the authors and does not necessarily
37 represent the official views of the National Institute on Deafness and Other Communication
38 Disorders or the National Institutes of Health.
39
40

41 **Conflicts of interest:** Authors ACM, GE, and CE are shareholders in Otologic
42 Technologies. Authors ACM and MNG are paid consultants and serve on the Board of
43 Directors for Otologic Technologies.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

Objectives: With the increasing emphasis on developing effective telemedicine approaches in Otolaryngology, this study explored whether a single composite image stitched from a digital otoscopy video provides acceptable diagnostic information to make an accurate diagnosis, as compared with that provided by the full video.

Methods: Five Ear, Nose, and Throat (ENT) physicians reviewed the same set of 78 digital otoscope eardrum videos from four eardrum conditions: normal, effusion, retraction, and tympanosclerosis, along with the composite images generated by a *SelectStitch* method that selectively uses video frames with computer-assisted selection, as well as a *Stitch* method that incorporates all the video frames. Participants provided a diagnosis for each item along with a rating of diagnostic confidence. Diagnostic accuracy for each pathology of *SelectStitch* was compared with accuracy when reviewing the entire video clip and when reviewing the *Stitch* image.

Results: There were no significant differences in diagnostic accuracy for physicians reviewing *SelectStitch* images and full video clips, but both provided better diagnostic accuracy than *Stitch* images. The inter-reader agreement was moderate.

Conclusion: Equal to using full video clips, composite images of eardrums generated by *SelectStitch* provided sufficient information for ENTs to make the correct diagnoses for most pathologies. These findings suggest that use of a composite eardrum image may be sufficient for telemedicine approaches to ear diagnosis, eliminating the need for storage and transmission of large video files, along with future applications for improved documentation in electronic medical record systems, patient/family counseling, and clinical training.

1
2
3 **Level of Evidence:** Level 3
4

5 **Keywords:** Computer-assisted Diagnosis, Eardrum, Image stitching, Otoscope,
6
7
8 Telemedicine.
9

10 **Total words:** 2993
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Introduction

Clinical examination of the eardrum (tympanic membrane – TM) through handheld otoscopy is the most common diagnostic approach for TM pathologies¹. Interpretation of the often brief glimpse of the TM obtained through the small viewing window requires extensive experience. With a growing need to develop effective telemedicine systems, which recently gained widespread attention during the COVID-19 pandemic², novel methods to perform telemedicine otoscopy are needed. Previous studies have shown that telemedicine review of images is sufficiently accurate to use in burn³⁻⁵ and trauma care^{6,7}. For otoscopy, one telemedicine approach is to apply digital otoscopy, during which a short video examination of the TM is recorded, which is reviewed by a telemedicine physician. Although there are no studies directly comparing otoscopic diagnoses based on a video clip as compared with a single digital image, our previous work led us to use videos⁸. There, diagnostic performance was compared between otoscopic single images and in-office microscopy. We included only images that were of sufficient focus/lighting, representing relatively ideal imaging conditions. Other authors have also noted insufficient image quality in a large percentage of their otoscopic still image databases, and/or the broad variability inherent across still images^{9,10}.

The use of digital otoscopic video clips could overcome limitations imposed in real clinical settings: collecting a string of frames in a video could capture at least a few useful frames with sufficient focus and lighting, even in the setting of partially obstructing cerumen or a moving child. However, a major downside of video clips is that they require a large amount of storage space. A typical otoscopic video clip is $1440 \times 1080 \times 24$ bits/pixels per frame, with between 200 and 1000 frames, contrasted with a single frame

1
2
3 for a still image. Transferring large video clips, even after compression, can provide a
4 major barrier to care in settings where internet bandwidth is limited¹¹⁻¹⁵. Moreover, storage
5 of videos within electronic medical records is not currently streamlined, and replaying
6 videos for patients/families for counseling purposes is tedious.
7
8
9

10
11
12 Along similar lines, previous studies have been done to detect ear abnormalities by
13 computer-based methods, requiring either manually extracting a single image from videos
14 or capturing a single image with minimal glare/obstruction^{9,10,16-19}. However, manually
15 selecting a frame from a video is time-consuming and subject to high inter- and intra-reader
16 variability²⁰⁻²². A more sophisticated computerized method that creates a “composite”
17 otoscopic image from a video should lead to a more useful final image, since a typical
18 video comprises at least 200 frames, more than one of which may contribute information
19 to the diagnostician.
20
21
22
23
24
25
26
27
28
29

30
31 We previously reported a computer-aided otoscopic frame selecting and stitching
32 framework called *SelectStitch*²³, in which a semantic segmentation-based framework
33 automatically selects meaningful frames containing portions of the TM from videos,
34 reducing irrelevant frames (e.g., those heavily blurred or having excessive cerumen). We
35 then conducted a reader study with three Ear, Nose, and Throat (ENT) physicians who
36 reviewed these composite *SelectStitch* images and compared them to composite images
37 generated using the entire video (i.e., without frame selection, called *Stitch*) in terms of
38 diagnostic decisions. Figure 1 provides an overview of *SelectStitch* and *Stitch*. We found
39 that *SelectStitch* improved the diagnostic quality of composite images relative to *Stitch*.
40 However, that study did not address several remaining questions important for the
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 translation of this approach to the clinic and particularly to telemedicine settings, which
4
5 are addressed in the current study:
6

7
8 *(a) When reviewing SelectStitch composite images, what is the accuracy of diagnosis? To*
9
10 answer this question, five ENTs reviewed 78 composite images and provided diagnoses,
11
12 compared with a “true” diagnosis. For adult patients, the “true” diagnosis was based on
13
14 digital otoscopy, supplemented with clinical microscopy as well as audiology testing
15
16 (hearing testing and/or tympanometry). For pediatric patients, the “true” diagnosis was
17
18 based on digital otoscopy, supplemented with microscopy in the operating room during
19
20 placement of pressure equalization tubes. We also aimed to determine which pathologies
21
22 were easiest and hardest to diagnose.
23
24

25
26 *(b) Is the accuracy of diagnosis for SelectStitch composite images different from Stitch*
27
28 *images and from videos?* Because the *Stitch* technique generates composite images using
29
30 all frames of a video, including redundant frames and frames of poor quality, we predicted
31
32 that the diagnostic accuracy for *SelectStitch* images would be superior to the accuracy for
33
34 *Stitch*. More importantly, we predicted that the diagnostic accuracy for *SelectStitch* would
35
36 be equivalent to the accuracy for full video clips.
37
38

39
40 *(c) How does the level of confidence of ENTs for each diagnostic tool (SelectStitch, Stitch,*
41
42 *and video) relate to diagnostic ability?* To answer this question, the five ENTs rated their
43
44 level of confidence in making diagnoses for each type of pathology in each diagnostic tool
45
46 condition.
47
48

49
50 *(d) What is the inter-reader variability of ENTs on diagnosing with the diagnostic tools?*
51
52 As with any medical application, we expected that there would be inter-reader variability
53
54 among ENTs, but that agreement would generally be relatively high.
55
56
57

Materials and Methods

A database of high-resolution digital adult and pediatric videos, captured via a digital otoscope from ENT clinics and operating rooms, as well as in a primary care Medicine/Pediatrics setting, was created after Institutional Review Board (IRB) approval⁸. A high definition (HD) video otoscope (JEDMED Horus+ HD Video Otoscope, St. Louis, MO) was utilized²³. The video frames were 1440 by 1080 pixels and were recorded in a MPEG 4 file format. In this study, 78 video clips from the database were used, selected if they only had one single diagnostic label associated. These videos consisted of 20 normal ears, 20 with middle ear effusions (serous or mucoid), 20 with TM retractions, and 18 with tympanosclerosis (i.e., myringosclerosis). Videos were excluded if they had low light throughout the video and/or if they did not contain a clear view of at least part of the TM. Where possible, we selected pediatric and adult videos as balanced (e.g., 10 pediatrics and 10 adults), except for tympanosclerosis for which there were 10 adult and eight pediatric videos.

An online diagnostic assessment tool was designed using SurveyMonkey, an online survey software. The video clips were hosted on Vimeo, and the composite images were uploaded to imgbox. An example of a question from our online survey is shown in Figure 2. Each sample (*Stitch* or *SelectStitch* composite image or video) was displayed on the screen, and the reader was asked to state the diagnosis (or normality). The order of presentation (video first or composite image first) was randomized to each clinician separated by four weeks (see Figure 3). If a reader viewed the video of a sample in the first survey, he/she read the *Stitch* and *SelectStitch* composite images (also in a randomized

1
2
3 order) in the second survey. The cases from adult and pediatric patients were also mixed
4
5 in each survey.
6

7
8 At the completion of each survey, readers were asked to rate their degree of
9
10 confidence in making each type of diagnosis, on a scale of 1 to 5, in which 1 indicated no
11
12 confidence while 5 indicated extreme confidence. Five ENTs (authors ACM, GE, JS, JKM,
13
14 and MSH; three neurotologists, one comprehensive otolaryngologist, and one pediatric
15
16 otolaryngologist) were invited by email to complete the online assessment, and all
17
18 completed the assessment after completing written informed consent.
19
20

21 22 *Statistical Analyses* 23

24
25 Two different scoring strategies were applied to the survey answers. Although each
26
27 sample had only one true diagnostic label, we did not restrict the readers regarding the
28
29 number of diagnostic answers they could provide for each sample. Answers were scored
30
31 using two different strategies. In Score-1, we scored the answers according to whether the
32
33 reviewer provided the correct diagnostic answer as well as how many answers were given
34
35 (e.g., the true label was effusion but the reviewer provided two diagnoses: effusion and
36
37 tympanosclerosis). To compute accuracy using Score-1, an answer-weighting strategy was
38
39 used: $\text{proportion} = \frac{\delta}{N_A}$ where N_A was the number of answers provided by the reviewer and
40
41 δ was the binary output of answers, where the item received a 1 if any of the answers were
42
43 correct and 0 otherwise. For example, if a reader selected two answers ($N_A = 2$) and one of
44
45 them was correct ($\delta = 1$), then the proportion (in percentage) for that particular sample
46
47 would be 50%. It should be noted that this is not “accuracy” in a traditional sense. In
48
49 contrast, for Score-2, an answer was accepted as correct if any diagnosis in the response
50
51
52
53
54
55
56
57
58
59
60

1
2
3 matched the true label. Score-2 was computed as the percent of responses that contained a
4
5 correct diagnosis, which is a relatively lenient approach to accuracy.
6

7
8 To study differences in scoring among diagnostic tools (video clips, *Stitch*,
9
10 *SelectStitch*) and among the five ENT doctors, ordinal logistic regression was applied to
11
12 both Score-1 and Score-2. Similar analysis was performed for studying the association
13
14 between confidence level in scoring and scores. Wald tests were performed for comparison
15
16 between diagnostic tools. Bonferroni method was used for multiple comparisons (e.g., $\alpha =$
17
18 0.016 when adjusting for three comparisons). Kendall's concordance was calculated to
19
20 assess inter-reader agreement for each diagnostic tool, where a concordance of 0 suggests
21
22 no inter-reader agreement and a concordance of 1 suggests perfect agreement.
23
24
25
26

27 **Results**

28 Question (a): What was the accuracy of diagnosis for SelectStitch?

29
30
31
32
33
34 As shown in Table 1, the overall proportions of diagnostic accuracy for *SelectStitch*
35
36 images among ENTs varied between 46 and 62% for Score-1. The easiest and hardest
37
38 categories to diagnose, respectively, were Tympanosclerosis (mean \pm std: 69% \pm 9) and
39
40 Retraction (mean \pm std: 39% \pm 7).
41
42

43
44 For Score-2, also shown in Table 1, the average accuracy rates of ENT doctors for
45
46 *SelectStitch* suggested that the easiest category to diagnose was again Tympanosclerosis
47
48 (mean \pm std: 84% \pm 7) and the hardest to diagnose was Retraction (mean \pm std: 56% \pm 12).
49
50 Overall Score-2 accuracies among ENTs varied between 57 and 74%.
51
52
53
54
55
56
57

1
2
3 Question (b): Did the accuracy of diagnosis differ among *SelectStitch*, *Stitch*, and video
4
5
6 clips?

7
8 Similar tables of diagnostic accuracy are shown in the Appendix for *Stitch*
9
10 (Appendix Table A) and video clips (Appendix Table B). For Score-1, overall, there was a
11
12 significant difference in score among the three diagnostic tools at p value < 0.0001 (F value
13
14 = 44.42). For paired comparisons, there was no significant difference between the video
15
16 method and *SelectStitch* at p value = 0.9736; there was a significant difference between
17
18 video method and *Stitch* (*Stitch* method scored less) at p value < 0.0001; and there was a
19
20 significant difference between *SelectStitch* and *Stitch* (*Stitch* scored less) at p value <
21
22 0.0001 (Table 2). For Score-2, overall, there was also a significant difference in score
23
24 among the three diagnostic tools at p value < 0.0001 (F value = 51.06). For paired
25
26 comparisons, there was no significant difference between diagnostic accuracy for video
27
28 clips and *SelectStitch* at p value = 0.9391; there was a significant difference between video
29
30 and *Stitch* (*Stitch* scored less) at p value < 0.0001; and there was a significant difference
31
32 between *SelectStitch* and the *Stitch* (*Stitch* scored less) at p value < 0.0001 (see Table 2).
33
34 In summary, for both Score-1 and Score-2, diagnostic accuracy was equivalent for
35
36 *SelectStitch* composite images and video clips, both of which were better than for *Stitch*
37
38 images.
39
40
41
42
43
44
45
46

47 Question (c): How did level of confidence of ENTs for each diagnostic tool relate to their
48
49 diagnostic ability?

50
51 Associations between confidence level and Score-1 were examined for each
52
53 diagnostic tool. For *Stitch*, this association was not significant (t value = -2.4) after
54
55
56
57

1
2
3 Bonferroni correction (p value = 0.0168; α used for Bonferroni correction = 0.0167). For
4
5 video clips, this association was positive and significant (t value = 3 and p value = 0.0027).
6
7
8 Finally, for *SelectStitch*, this association was positive and significant (t value = 2.87 and p
9
10 value = 0.0041). Next, we compared the magnitude of association between confidence and
11
12 Score-1 among the three diagnostic tools. Results demonstrated no significant difference
13
14 in this association between video clips and *SelectStitch* (p value = 0.9944). In contrast, the
15
16 association between confidence level and Score-1 for video clips was significantly higher
17
18 than that for *Stitch* (p value < 0.0001), and the association for *SelectStitch* was significantly
19
20 higher than that for *Stitch* (p value < 0.0001). Similar findings were demonstrated for the
21
22 associations between confidence level and Score-2 of each diagnostic tool. Specifically,
23
24 for *Stitch*, the association was non-significant (t value = -2.3) after Bonferroni correction
25
26 (p value = 0.0214; α used for Bonferroni correction = 0.0167). For video clips, the
27
28 association was positive and significant (t value = 3.33 and p value = 0.0009). Finally, for
29
30 *SelectStitch*, the association was positive and significant (t value = 3.19 and p value =
31
32 0.0015). Again, the magnitude of associations of confidence level and Score-2 were
33
34 compared among the three diagnostic tools. There was no significant difference between
35
36 video clips and *SelectStitch* (p value = 0.9936). However, again, the association was
37
38 significantly higher for video clips than for *Stitch* (p value < 0.0001), and the association
39
40 was significantly higher for *SelectStitch* than for *Stitch* (p value < 0.0001). In summary,
41
42 the associations between diagnostic accuracy (using both Score-1 and Score-2) and
43
44 confidence level were significant only for *SelectStitch* and video clips, and these
45
46 associations were of similar magnitude.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Question (d): What was the inter-reader variability of ENTs with each diagnostic tool?
4

5 Kendall's concordance was used for assessing inter-reader agreement. For Score-
6 1, concordance was 0.5778 for *Stitch*, 0.4096 for video clips, and 0.4779 for *SelectStitch*.
7
8 For Score-2, concordance was 0.584 for *Stitch*, 0.4312 for video clips, and 0.3529 for
9
10 *SelectStitch*. These Kendall's concordance values are all moderate in magnitude.
11
12
13

14
15
16
17 **Discussion**
18

19 Telemedicine approaches have recently been highlighted during the COVID-19
20 pandemic. Even before then, telemedicine started to gain increasing attention in
21
22 Otolaryngology²⁴⁻²⁶. Otoscopy is well-suited to the telemedicine approach^{27,28}, as long as
23
24 a sufficient image of the TM can be obtained. One way to optimize a sufficient image is to
25
26 collect a short video clip of the examination. However, this results in a digital file that is
27
28 relatively large, posing a barrier to both storage and transfer, especially in remote settings
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
27,29. We hypothesized that computer-assisted creation of a composite image would
maintain equivalent diagnostic utility and physician confidence during diagnosis.

Results demonstrated that the accuracies of ENTs in making diagnoses from
SelectStitch images were equivalent to those made when reviewing the full videos,
regardless of how diagnostic accuracy was determined (the stringent Score-1 vs the lenient
Score-2). The overall average accuracies of ENTs (specifically for Score-2) were from 57
to 74%, a range that is similar to our previous study⁸. However, diagnostic accuracy
depended largely on the type of pathology. For example, experts were 84% accurate in
diagnosing tympanosclerosis, which has some distinguishing features (i.e., discrete areas
of white plaque). In contrast, accuracy was lowest for the diagnosis of TM retraction, which

1
2
3 can be a fairly subtle finding. Nonetheless, the most important finding of this study was
4 that there were no significant differences in diagnostic accuracy between *SelectStitch*
5 composite images and the full video clips. In contrast, *Stitch* composite images, which were
6 constructed using all available frames of a given video, led to much poorer diagnostic
7 accuracy than either *SelectStitch* or full video clips. This is a highly significant finding,
8 because it suggests that single *SelectStitch* images provide details that are of equal
9 diagnostic value to full video clips for expert reviewers.
10
11
12
13
14
15
16
17
18

19 Additionally, diagnostic accuracy and diagnostic confidence level were associated
20 for both the *SelectStitch* and full video clips, while no association was found for *Stitch*
21 images, which is interesting in light of previous work that demonstrated an overall weak
22 relationship between diagnostic accuracy and confidence in ear experts ⁸. This finding is
23 important, because treatment decisions are often impacted by level of diagnostic
24 confidence of the clinician. For example, a physician may need to feel confident of
25 providing a diagnosis of “normal” in order to choose not to prescribe antibiotics for a
26 patient presenting with otalgia. Moreover, inter-reader agreement in this study was
27 generally only moderate in magnitude, providing further motivation for the need to develop
28 methods to improve the objectivity of making ear diagnoses ^{16,17}.
29
30
31
32
33
34
35
36
37
38
39
40
41

42 This study has several limitations. First, only a subset of pathologies was included
43 in the survey, while several important ear pathologies were excluded, such as acute otitis
44 media. This was a result of small numbers of videos of some pathologies in our current
45 database. Also, only videos of relatively high quality/lighting were included. Another
46 limitation is that within-reader agreement was not evaluated. Lastly, each reader used
47 his/her own computer monitor to evaluate the images and videos. Those monitors were
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 likely of different makes, models, and resolutions. All of these factors could contribute to
4
5 differences in diagnostic abilities. On the other hand, our approach was likely ecologically
6
7 valid; in various telemedicine settings, a variety of monitors will be used.
8
9

10 Although the emphasis of this study was to provide support for the value of
11
12 *SelectStitch* composite images in potential telemedicine settings, there are other scenarios
13
14 for which an otoscopic composite image is likely preferable over a video clip. For example,
15
16 current electronic medical record systems are more amenable to inclusion of photo-
17
18 documentation in patient charts, as compared with video examinations. Additionally, the
19
20 ability to show a patient or parent a simple composite image of a TM would improve
21
22 counseling, such as in providing visual confirmation of a normal ear in a child with otalgia,
23
24 which may help decrease over-prescription of oral antibiotics.
25
26
27
28
29

30 **Conclusion**

31
32
33 Results of this study demonstrated that computer-aided *SelectStitch* composite
34
35 images provide equivalent visual information as digital otoscopic video clips for ear
36
37 experts to make diagnoses of different types of pathologies. Diagnostic accuracy was also
38
39 found to be associated with diagnostic confidence level, and inter-reader agreement was
40
41 moderate. Future studies will be required to evaluate a more diverse set of ear pathologies,
42
43 as well as using videos collected under less ideal focus and lighting conditions.
44
45
46
47
48
49
50
51
52
53
54
55
56
57

1
2
3 **Acknowledgments**
4

5 The authors would like to thank Emily Luo and Benjamin Liu for curating videos and
6
7 online surveys for this study.
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

References

1. Cole LK. Ooscopic evaluation of the ear canal. *The Veterinary clinics of North America Small animal practice*. 2004;34(2):397-410.
2. Chauhan V, Galwankar S, Arquilla B, et al. Novel coronavirus (COVID-19): Leveraging telemedicine to optimize care while minimizing exposures and viral transmission. *Journal of Emergencies, Trauma, and Shock*. 2020;13(1):20.
3. Holt B, Faraklas I, Theurer L, Cochran A, Saffle JR. Telemedicine use among burn centers in the United States: a survey. *Journal of burn care & research*. 2012;33(1):157-162.
4. Wallace D, Jones S, Milroy C, Pickford M. Telemedicine for acute plastic surgical trauma and burns. *Journal of plastic, reconstructive & aesthetic surgery*. 2008;61(1):31-36.
5. Reiband HK, Lundin K, Alsbjørn B, Sørensen AM, Rasmussen LS. Optimization of burn referrals. *Burns*. 2014;40(3):397-401.
6. Chan F, Whitehall J, Hayes L, et al. Minimum requirements for remote realtime fetal tele-ultrasound consultation. *Journal of Telemedicine and Telecare*. 1999;5(3):171-176.
7. Baruffaldi F, Mattioli P, Toni A, Klutke P, Englmeier K. Low-cost ISDN videoconferencing equipment for orthopaedic second opinions. *Journal of Telemedicine and Telecare*. 1999;5(1_suppl):37-38.
8. Moberly AC, Zhang M, Yu L, et al. Digital otoscopy versus microscopy: How correct and confident are ear experts in their diagnoses? *Journal of telemedicine and telecare*. 2018;24(7):453-459.
9. Myburgh HC, Van Zijl WH, Swanepoel D, Hellström S, Laurent C. Otitis media diagnosis for developing countries using tympanic membrane image-analysis. *EBioMedicine*. 2016;5:156-160.
10. Kuruvilla A, Shaikh N, Hoberman A, Kovačević J. Automated diagnosis of otitis media: vocabulary and grammar. *Journal of Biomedical Imaging*. 2013;2013:27.
11. Nejad ES, Majma MR, Izadpanahi B, Natanzi SBH, Navaei HR. Infrastructure of data centers for transferring big data traffic: A survey research. Paper presented at: 2015 International Congress on Technology, Communication and Knowledge (ICTCK)2015.
12. Syed-Abdul S, Scholl J, Chen CC, et al. Telemedicine utilization to support the management of the burns treatment involving patient pathways in both developed and developing countries: a case study. *Journal of Burn Care & Research*. 2012;33(4):e207-e212.
13. Atiyeh B, Dibo S, Janom H. Telemedicine and burns: an overview. *Annals of burns and fire disasters*. 2014;27(2):87.
14. Gardiner S, Hartzell TL. Telemedicine and plastic surgery: a review of its applications, limitations and legal pitfalls. *Journal of Plastic, Reconstructive & Aesthetic Surgery*. 2012;65(3):e47-e53.
15. Daniel Chaves Viquez K, Arandjelovic O, Blaikie A, Ae Hwang I. Synthesising wider field images from narrow-field retinal video acquired using a low-cost direct ophthalmoscope (Arlight) attached to a smartphone. Paper presented at:

- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10
 - 11
 - 12
 - 13
 - 14
 - 15
 - 16
 - 17
 - 18
 - 19
 - 20
 - 21
 - 22
 - 23
 - 24
 - 25
 - 26
 - 27
 - 28
 - 29
 - 30
 - 31
 - 32
 - 33
 - 34
 - 35
 - 36
 - 37
 - 38
 - 39
 - 40
 - 41
 - 42
 - 43
 - 44
 - 45
 - 46
 - 47
 - 48
 - 49
 - 50
 - 51
 - 52
 - 53
 - 54
 - 55
 - 56
 - 57
 - 58
 - 59
 - 60
- Proceedings of the IEEE International Conference on Computer Vision Workshops2017.
16. Senaras C, Moberly AC, Teknos T, et al. Autoscope: automated otoscopy image analysis to diagnose ear pathology and use of clinically motivated eardrum features. Paper presented at: Medical Imaging 2017: Computer-Aided Diagnosis2017.
 17. Senaras C, Moberly AC, Teknos T, et al. Detection of eardrum abnormalities using ensemble deep learning approaches. Paper presented at: Medical Imaging 2018: Computer-Aided Diagnosis2018.
 18. Moein M, Davarpanah M, Montazeri MA, Ataei M. Classifying ear disorders using support vector machines. Paper presented at: 2010 Second International Conference on Computational Intelligence and Natural Computing2010.
 19. Binol H, Moberly AC, Niazi MKK, et al. Decision fusion on image analysis and tympanometry to detect eardrum abnormalities. Paper presented at: Medical Imaging 2020: Computer-Aided Diagnosis2020.
 20. Jeffay K, Zhang HJ. *Readings in multimedia computing and networking*. Elsevier; 2001.
 21. Han B, Hamm J, Sim J. Personalized video summarization with human in the loop. Paper presented at: 2011 IEEE Workshop on Applications of Computer Vision (WACV)2011.
 22. Gygli M, Grabner H, Riemenschneider H, Van Gool L. Creating summaries from user videos. Paper presented at: European conference on computer vision2014.
 23. Binol H, Moberly AC, Niazi MKK, et al. SelectStitch: Automated Frame Segmentation and Stitching to Create Composite Images from Otoscope Video Clips. *IEEE Journal of Biomedical and Health Informatics (Under review)*.
 24. Seim NB, Philips RH, Matrka LA, et al. Developing a synchronous otolaryngology telemedicine Clinic: Prospective study to assess fidelity and diagnostic concordance. *The Laryngoscope*. 2018;128(5):1068-1074.
 25. McCool RR, Davies L. Where does telemedicine fit into otolaryngology? An assessment of telemedicine eligibility among otolaryngology diagnoses. *Otolaryngology–Head and Neck Surgery*. 2018;158(4):641-644.
 26. Maurrasse SE, Rastatter JC, Hoff SR, Billings KR, Valika TS. Telemedicine During the COVID-19 Pandemic: A Pediatric Otolaryngology Perspective. *Otolaryngology–Head and Neck Surgery*. 2020:0194599820931827.
 27. Short AB. Efficacy of digital otoscopy in telemedicine. 2017.
 28. Meng X, Dai Z, Hang C, Wang Y. Smartphone-enabled wireless otoscope-assisted online telemedicine during the COVID-19 outbreak. *American Journal of Otolaryngology*. 2020.
 29. Kokesh J, Ferguson AS, Patricoski C, et al. Digital images for postsurgical follow-up of tympanostomy tubes in remote Alaska. *Otolaryngology—Head and Neck Surgery*. 2008;139(1):87-93.

Table 1. Proportion of correct diagnosis in percentages for each diagnostic category for each ENT physician (I through V) using Score-1 (S1) and Score-2 (S2) (%) using *SelectStitch*.

Diagnostic Categories	ENT-I		ENT-II		ENT-III		ENT-IV		ENT-V		mean (SD)	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
Normal	55	55	55	55	65	65	85	95	55	55	63 (13)	65 (17)
Effusion	78	90	52	90	64	75	16	25	25	35	47 (26)	63 (31)
Retraction	38	55	45	65	29	40	48	70	37	50	39 (7)	56 (12)
Tympanosclerosis	78	94	57	78	67	78	78	83	65	89	69 (9)	84 (7)
Average (mean (SD))	62 (19)	74 (21)	52 (5)	72 (15)	56 (18)	65 (17)	57 (32)	68 (31)	46 (18)	57 (23)		

Table 2. Comparisons of accuracy for the different diagnostic tools for Score-1 (S1) and Score-2 (S2). V = Video; S = *Stitch*; SS = *SelectStitch*.

Label	Estimate		Standard Error		DF		t Value		Pr > t	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
V vs S	1.1857	1.3279	0.1424	0.1524	1160	1163	8.33	8.71	<.0001	<.0001
SS vs S	1.1901	1.3396	0.1427	0.1526	1160	1163	8.34	8.78	<.0001	<.0001
SS vs V	0.0044	0.0117	0.1331	0.1529	1160	1163	0.03	0.08	0.9736	0.9391

Figure Captions

Figure 1. The process of *Stitch* and *SelectStitch*. In comparison to *Stitch*, *SelectStitch* possesses a deep learning based semantic segmentation step to reduce irrelevant frames from video sequences as described in ²³. These excluded frames include parts of the video with low quality (e.g., those heavily blurred or having an excessive amount of cerumen).

Figure 2. An example question from the online diagnostic survey. The readers are asked to make a diagnosis of the disease either from the video or a composite image (either produced by *Stitch* or *SelectStitch*). Readers can pick one or more of the choices. If their diagnosis is not included in any of the categories, they can pick the Other Category and enter their choice (e.g. monomeric TM). Readers are also asked their diagnostic confidence level using the Likert Scale with 5 being “extremely confident.”

Figure 3. Summary of the rounds of the otoscope diagnosis survey for each reader (ENT-I through ENT-V). The order of the *Stitch* and *SelectStitch* composite images of the same sample were mixed in each survey. The cases from adult and pediatric patients were also mixed in each evaluation set.

Appendix

Appendix Table A. Proportion of correct diagnosis in percentages for each diagnostic category for each ENT physician (I through V) using Score-1 (S1) and Score-2 (S2) (%) using *Stitch*.

Diagnostic Categories	ENT-I		ENT-II		ENT-III		ENT-IV		ENT-V		mean (SD)	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
Normal	15	15	5	5	10	10	33	35	20	20	17 (11)	17 (11)
Effusion	68	70	31	40	55	55	10	10	13	15	35 (26)	38 (26)
Retraction	39	45	38	45	23	25	36	40	20	35	31 (9)	38 (8)
Tympanosclerosis	64	67	28	33	38	39	49	50	43	50	44 (13)	48 (13)
Average (mean (SD))	46 (24)	49 (25)	25 (14)	31 (18)	31 (19)	32 (19)	31 (16)	33 (17)	23 (13)	30 (16)		

Appendix Table B. Proportion of correct diagnosis in percentages for each diagnostic category for each ENT physician (I through V) using Score-1 (S1) and Score-2 (S2) (%) using video clips.

Diagnostic Categories	ENT-I		ENT-II		ENT-III		ENT-IV		ENT-V		mean (SD)	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
Normal	60	60	55	55	70	70	93	95	29	30	61 (23)	62 (24)
Effusion	84	90	68	85	70	75	23	25	16	30	52 (31)	61 (31)
Retraction	48	55	53	70	35	45	59	60	49	70	49 (9)	60 (11)
Tympanosclerosis	83	89	65	78	72	48	80	83	76	94	75 (7)	78 (18)
Average (mean (SD))	68 (18)	73 (19)	60 (7)	72 (13)	62 (18)	67 (15)	64 (31)	65 (31)	42 (26)	55 (31)		

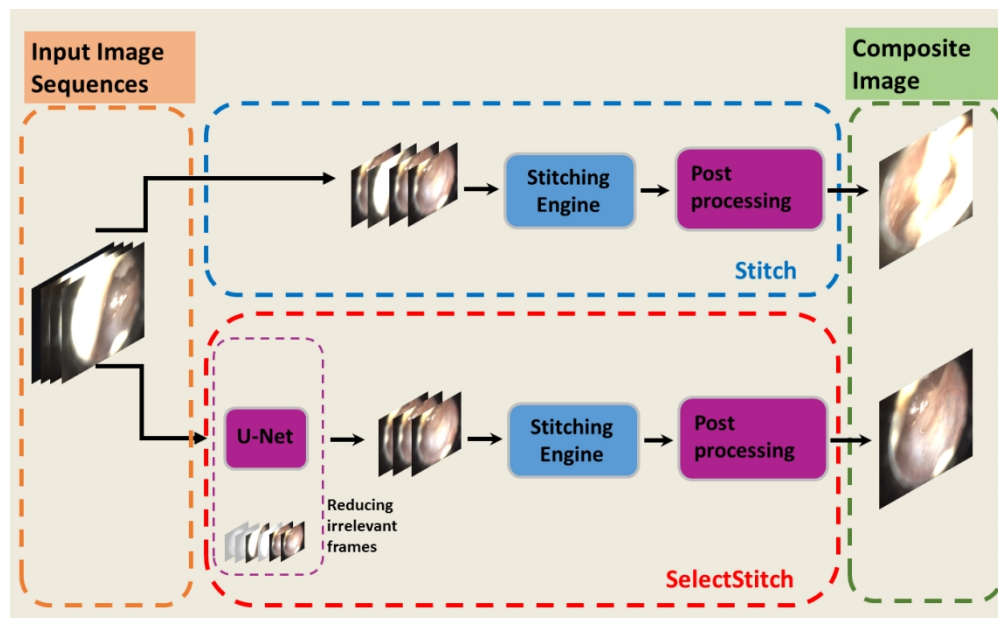
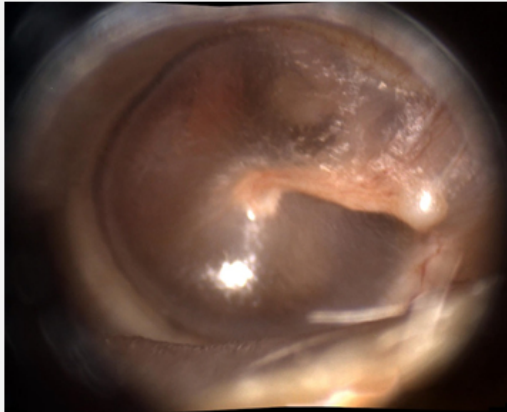


Figure 1. The process of Stitch and SelectStitch. In comparison to Stitch, SelectStitch possesses a deep learning based semantic segmentation step to reduce irrelevant frames from video sequences as described in 23. These excluded frames include parts of the video with low quality (e.g., those heavily blurred or having an excessive amount of cerumen).

368x225mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



* Disease

<input type="checkbox"/> AOM	<input type="checkbox"/> Retraction
<input type="checkbox"/> Effusion	<input type="checkbox"/> Tympanosclerosis
<input type="checkbox"/> Obstructing Cerumen	<input type="checkbox"/> Tube
<input type="checkbox"/> Perforation	<input type="checkbox"/> Normal
<input type="checkbox"/> Other (please specify)	

* Confidence Level

<input type="radio"/> (5) Extremely confident	<input type="radio"/> (2) Not so confident
<input type="radio"/> (4) Very confident	<input type="radio"/> (1) Not at all confident
<input type="radio"/> (3) Somehow confident	

Figure 2. An example question from the online diagnostic survey. The readers are asked to make a diagnosis of the disease either from the video or a composite image (either produced by Stitch or SelectStitch). Readers can pick one or more of the choices. If their diagnosis is not included in any of the categories, they can pick the Other Category and enter their choice (e.g. monomeric TM). Readers are also asked their diagnostic confidence level using the Likert Scale with 5 being "extremely confident."

185x210mm (96 x 96 DPI)

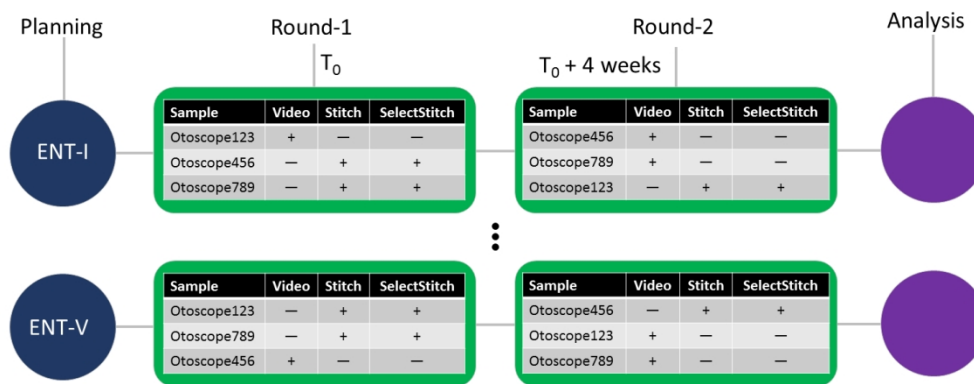


Figure 3. Summary of the rounds of the otoscope diagnosis survey for each reader (ENT-I through ENT-V). The order of the Stitch and SelectStitch composite images of the same sample were mixed in each survey. The cases from adult and pediatric patients were also mixed in each evaluation set.

406x177mm (96 x 96 DPI)