Interspeech 2018
2-6 September 2018, Hyderabad

Voice Analysis Using Acoustic and Throat Microphones for Speech Therapy

Lani Rachel Mathew[1], *K. Gopakumar*[2]

[1]Mar Baselios College of Engineering, Thiruvananthapuram, Kerala, India
[2]TKM College of Engineering, Kollam, Kerala, India
lani.mathew@mbcet.ac.in, gopakumar@tkmce.ac.in

Abstract

Diagnosis of voice disorders by a speech therapist involves the process of voice recording with the patient, followed by software-aided analysis. In this paper, we propose a novel voice diagnosis system which gives voice report information based on Praat software, using voice samples from a throat microphone and an acoustic microphone, making the diagnosis near real-time, as well as robust to background noise. Results show that throat microphones give reliable Jitter and Shimmer values in ambient noise levels of 47~50 dB, while acoustic microphones show high variance in these parameters.

Index Terms: throat microphone, acoustic microphone, voice analysis, therapy, voice disorder, Praat, voice report, jitter, shimmer

1. Introduction

The vast majority of people are unaware of the strain that they burden their vocal cords with, each day of their lives. Professionals who depend on their voices – the teaching community, radio and video jockeys, sports commentators, politicians and religious leaders – risk losing their voice due to overuse or abuse of their larynx or vocal cords. The paucity of quick and dependable diagnostic tools poses an obstacle for speech therapists in identifying potential voice disorders.

In this paper, we propose a system that has the potential to be used as a quick diagnostic tool for a speech therapist. It uses an acoustic microphone in combination with a throat microphone to give voice report information immediately after recording, using open-source speech software Praat running on a Raspberry pi.

The throat microphone has been recognized to be robust to noise in many studies [1, 2]. Throat microphone based voice output has been compared with that of acoustic microphones for laryngeal pathology detection in [3]. Results indicate that acoustic microphones are better suited for overall pathological voice detection. However crucial laryngeal information can be extracted from throat microphone recordings even in noisy environments [4].

The paper is organized as follows: Section 2 describes throat microphone based voice or speech analysis. The specifications of the throat microphone used, the voice report generated using Praat, the methodology adopted, and the results obtained are detailed in this section. Section 3 is a discussion on the results obtained, and Section 4 presents the conclusions of the paper.

2. Voice analysis set-up

A simple condenser (acoustic) microphone provides the audio input to the Raspberry Pi. Python script-based recording and generation of voice analysis report based on Praat [5] is done. The throat microphone is connected to a separate laptop and voice report is generated using the Praat Voice report option - a selection of a portion of speech is made in the Praat View & Edit window, with Pulses visible.

2.1. Methodology adopted

The throat microphone used in our experimental set-up was a USB Throat microphone XCTM825L-USB [6] from Kenwah System, Singapore. Fourteen adults in the age group 20 – 25 years participated in voice recording using a condenser microphone and a throat microphone simultaneously. The experimental group involved 5 males and 9 females. None of the participants had attended a professional voice assessment prior to the experiment.

The subjects were asked to sit comfortably wearing the throat microphone with its piezoelectric element positioned over their laryngeal portion of neck. The acoustic microphone was positioned approximately 10 cm away from the subjects' mouths. Phonation of the 14 subjects were recorded in a large room where ambient noise level was of the range of 47~50 dB. For phonation samples, each subject was asked to produce the phoneme */a/* for as long as they could sustain. The maximum phonation time was set at 10 seconds [7], after which the voice report was generated automatically.

Each participant was asked to describe his/her voice and state whether he/she suffered from any specific ailment related to voice. Throat microphone sound recordings were captured using Microsoft Voice Recorder.

2.2. Results obtained

Two male participants described their voice as hoarse and husky. One female participant described her voice as normal but showed a significant number of voice breaks in the acoustic microphone waveform. These three voices are thus described as 'pathological' in this section. The rest of the participants did not mention voice-related difficulty and their voices are described as 'normal' for this work.

2.2.1. Voice report parameters for 'normal' and 'pathological' samples

Tables 1 and 2 list certain voice parameters of phonation obtained from Praat voice reports of the obtained samples for 'normal' voices and 'pathological' voices respectively. Figure 1 shows the set-up for obtaining the voice report. Normal pitch value is in the range of 80-180 Hz for males and 180-250 Hz

for females. Normative values for jitter and shimmer are 1.04% and 3.98% respectively [5].

Table 1: *Voice report parameters for participants with 'normal' voice*

ID	Sex	Pitch	VB	Jitter	Shimmer
001	F	206/ 207	5/0	0.62/ 0.48	13.36/ 2.34
002	M	102/ 102	3/0	0.66/ 0.61	9.21/ 4.4
003	F	228/ 229	0/0	0.27/ 0.71	5.25/ 8.25
004	M	114/ 115	0/0	0.51/ 0.53	5.54/ 4.60
005	M	114/ 114	0/0	0.45/ 0.45	16.76/ 1.93
007	F	193/ 194	1/0	1.12/ 0.37	16.53/ 3.20
008	F	198/ 190	0/0	0.40/ 0.37	4.55/ 2.00
011	F	294/ 293	0/0	0.30/ 0.30	8.37/ 1.29
012	F	266/ 265	0/0	0.43/ 0.29	12.57/ 1.25
013	F	258/ 258	1/0	0.66/ 0.33	17.13/ 1.70
014	F	277/ 277	0/0	0.46/ 0.34	11.04/ 1.81
Average			**0.9/0**	**0.53/ 0.43**	**10.94/ 2.98**

Table 2: *Voice report parameters for participants with 'pathological' voice*

ID	Sex	Pitch	VB	Jitter	Shimmer
006	M	104/ 104	14/0	1.56/ 2.30	7.51/ 3.98
009	F	233/ 209	9/1	0.68/ 0.36	11.69/ 1.27
010	M	218/ 201	1/0	0.57/ 0.51	13.87/ 4.75
Average			**8.0/ 0.33**	**0.94/ 1.06**	**11.02/ 3.33**

Figure 1: *Set-up for Praat-based voice analysis using acoustic and throat microphones*

In Tables 1 and 2, VB denotes the number of voice breaks during phonation. Jitter (local) and Shimmer (local) are in percentages. Median pitch is used. For these parameters, two values are given – the first represents the measured values based on the acoustic microphone, while the second represents those from the throat microphone.

3. Discussion

In Table 1, the median pitch values of the acoustic and throat microphones vary within ± 1 Hz implying that pitch estimation is robust to the environment for both types of microphones. High variance from the normative value is observed in the shimmer values when measured using acoustic microphone, implying that shimmer values are affected by environmental noise. In Table 2, ID 006 has jitter and shimmer values above the normative values, characteristic of dysphonia, and shows a high number of voice breaks in the acoustic microphone readings. ID 009 also shows high number of voice breaks in acoustic mic readings, but these are not captured in the throat microphone. ID 010 shows an abnormally high pitch value for a male, characteristic of puberphonia. IDs 009 and 010 show a variance of ± 24 Hz and 17 Hz respectively in their median pitch values.

4. Conclusions

In this study, we have designed a novel and fast method to obtain a reliable diagnosis of voice. Throat microphones, with their proximity to the point of origin of speech - the larynx, hold the key to insights into the nature and content of speech. However since throat microphones only capture acoustic vibrations from the larynx, they need to be used in tandem with acoustic microphones to aid in the diagnosis of voice disorders by speech therapists. The model holds great promise for use in early intervention programs and assessment of voice for patients having voice disorders.

5. Acknowledgements

The authors acknowledge the valuable suggestions and inputs from the faculty of National Institute of Speech and Hearing, Thiruvananthapuram.

6. References

[1] A. Shahina and B. Yegnanarayana, "Language identification in noisy environments using throat microphone signals," in Proc. Int. Conf. Intell. Sensing Inform. Process., 2005, pp. 400–403.

[2] Shengke Lin et al., "DNN-based Feature Transformation for Speech Recognition Using Throat Microphone", Proceedings of APSIPA Annual Summit and Conference, 2017, pp. 596 – 599.

[3] Jonas Minelga et al., "Comparing Throat and Acoustic Microphones for Laryngeal Pathology Detection from Human Voice", International Conference on Electrical and Control Technologies, 2014, pp. 50-53.

[4] Shahina A., "Processing Throat Microphone Speech", PhD Thesis, Dept. of CSE, IIT Madras, 2007; http://cdn.iiit.ac.in/cdn/speech.iiit.ac.in/svlpubs/phdthesis/shahina-phd-2008.pdf

[5] Praat, http://www.fon.hum.uva.nl/praat/

[6] http://www.ixradio.com/shop/the-computer-usb-throat-microphone-system-xctm825l-usb-with-in-line-remote-control.html

[7] Maslan J, Leng X, Rees C, Blalock D, Butler SG., "Maximum Phonation Time in Healthy Older Adults", Journal of voice : official journal of the Voice Foundation. 25(6), 2011, pp. 709 - 713. doi:10.1016/j.jvoice.2010.10.002.