PORTABLE PHOTOGLOTTOGRAPHY FOR MONITORING VOCAL FOLD VIBRATIONS IN SPEECH PRODUCTION

Yujie Chi[1], Kiyoshi Honda[1], and Jianguo Wei[1,2]

[1]College of Intelligence and Computing, Tianjin University, Tianjin, China
[2]Computer College, Qinghai Nationalities University, Xining, China
yujiechi@qq.com, khonda@sannet.ne.jp, jianguo@tju.edu.cn

ABSTRACT

Photoglottography (PGG) is an effective method to monitor vocal fold vibrations via measuring light transmission across the glottis. The difficulty in operation however limits its wide use in speech studies. This paper is to realize a portable PGG (P-PGG) module with an audio interface to record glottal and speech waveforms simultaneously with ease. Near Infrared (NIR) LEDs are driven as a light source and an extremely high-gain photodetector circuit is employed. The output PGG signal is subsequently band-pass filtered and amplified for recording. The whole system is minimized, battery powered with well-controlled heat radiation. In experiments, P-PGG, EGG and microphone are worn together by speakers. The results verify that the NIR lighting P-PGG is successful at recording complete information of glottal cycles in comparison to EGG. Thus, the P-PGG is an effective method for investigating phonation types and consonant-vowel interactions in speech.

Index Terms— Photoglottography, vocal fold vibration, Near Infrared light, portable device, speech production

1. INTRODUCTION

Photoglottography (PGG) is an optoelectronic system for studying glottal behaviors, as firstly developed by Sonesson [1], [2]. In the current common systems similar to Ohala's [5], the basic units are a pair of light emitter and detector with the former being in the supraglottic cavity and the later placed on the skin surface below the glottis. The visible light passing through the glottis is transformed into electric voltage and then amplified to be recorded as glottograms. It was first developed for the study of vibratory states of the vocal folds (VF), and the further development of PGG was also used in running speech for both VF vibration and glottal adduction/abduction activities [3]-[5].

Soon after, Sawashima developed a laryngoscopic technique by use of a fiberoptic endoscope (fiberscope), which make it possible to record glottal motion pictures and glottograms simultaneously [6]-[10]. With a merit of allowing real-time acquisition and simple analysis of the signals, PGG combined with Electroglottography (EGG) was an attractive technique for studies on normal or pathologic voice production [11], [12]. As the high-speed imaging technique becomes popular, PGG alone is not frequently used for voice studies [13]. The non-invasive EGG combined with speech recording is standard for current non-clinical studies.

The conventional PGG has been reported to have two problems. Firstly, the operation of a fiberscope must be done by a medical specialist. Secondly, laryngeal views from the fiberscope are occluded by the tongue or epiglottis during production of back vowels. To overcome the problems, a new non-invasive means of measurement was sought for with advanced optoelectronic technologies. One of the key points for realizing the technology is the therapeutic optical window of tissues for higher light penetration, which ranges from 600 to 1400 nm in wavelength. Instead of high-power visible light source, infrared red (NIR) LEDs are applicable for lighting through tissues. Also, a new high-performance photodetector circuit can be combined for a new system.

Honda and Maeda first developed a non-invasive PGG system with external NIR lighting (ePGG) [14]-[16]. A NIR LED was used to illuminate the lateral neck tissues between the hyoid bone and thyroid cartilage. To avoid ambient light noise, a pulse-driven light source was applied, which required a wideband (but low-gain) circuit for demodulation. To follow, the ePGG circuits were improved to record high-quality signals in our previous work [22]. An elaborate transimpedance amplifier with a lock-in amplifier was used for higher-gain and lower-noise performances. Besides us, Suthau and Bouvet et al. reported their versions of ePGG [17]-[19]. So far, the ePGG has been applied in analyzing consonant production in Korean and Chinese [20]-[22].

The present study reports a new portable PGG (P-PGG) with a new aim at observing vocal fold vibrations as AC components of PGG signals. This modification for the AC-coupled system is due to our experience of the ePGG regarding articulatory effects showing signal baseline fluctuation at a low frequency range. Table 1 shows a comparison across the original PGG made by Ohala, our previous ePGG, and the new portable PGG [5], [22]. The

circuits of the P-PGG are simplified and further improved for better AC performances. Experiments are easier and more reliable with the P-PGG benefiting acquisition from many speakers. Also, combined recording with EGG and audio is possible as shown in what follows.

Table 1, Comparison of PGG and ePGG and P-PGG

Types	Light source		Light detector
	Lamp	Power	Transimpedance
PGG	Incandescent lamp	120V，100W	68 KΩ (0~1.2 kHz)
External PGG	NIR LEDs (850 nm)	20 kHz Pulsed, 1~3W	30 MΩ (0~50 kHz)
Portable PGG	NIR LEDs (850 nm)	Direct current, < 1 W	60 MΩ (0.05~5 kHz)

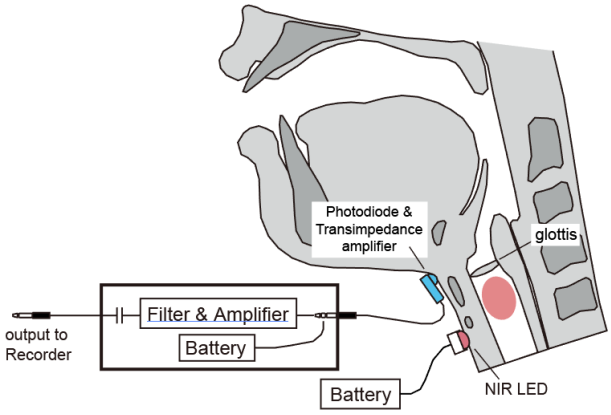


Figure 1, *Schematic diagram of Portable PGG setting*

2. METHODOLOGY FOR P-PGG

2.1. Circuit design

The circuit design for the P-PGG agrees basically with the principle of detecting weak optic signals. Fig. 1 shows a configuration for subglottal illumination with DC-driven LEDs, high-sensitivity photodiode with a transimpedance amplifier, and other filters and amplifiers. The whole system is battery powered.

High power NIR LEDs of 850-nm wavelength (OSI3XAE1C1E, OptoSupply) are selected as the light source. The LED of a surface mount package has a heat sink for power consumption within 1 W. In practice, one or more LEDs are used according to the necessity.

In our previous ePGG, a PIN photodiode BPW34 (OSRAM) was chosen because of its peak sensitivity at the 900-nm wavelength, large sensing area and high response time. Here, BPW34FAS (OSRAM) with the narrower NIR band is used to inhibit environment light.

The transillumination in P-PGG is so faint that the detectable photocurrent is as small as a nA level. Therefore, the pre-amplifier stage is crucial to transform the photocurrent into voltage at a mV level for further processing stages. As shown in the schematic diagram in Fig. 2, a transimpedance amplifier was composed with a feedback resistance (60 MΩ). The output voltage comprises the photocurrent from the photodiode multiplied by the resistance. The larger resistor causes larger noise; thus a small capacitor (0.5 pF) is added in parallel for narrowing the noise band. The operational amplifier LTC6268 (Analog Devices) has an extremely low noise and wide bandwidth up to 500 MHz. In all, the transimpedance amplifier gains the capability of transferring a current of 1 nA into a voltage of 60 mV with a signal bandwidth up to 5 kHz.

The right panel of Fig. 2 shows the print circuit board of the transimpedance amplifier circuit. All the circuit components are surface-mount on a thumbnail-size board. The board surface was sealed to protect the circuit from moisture, and the whole board is in a shielded metal enclosure.

The signals from the transimpedance amplifier are then band-pass filtered and amplified. Fig. 3 shows the band-pass filter of a cascade Sallen-Key type of high-pass and low-pass filters. The filter bandwidth is set at 50~2000 Hz to attenuate DC and high frequency noises. The following amplifier is a cascade of two inverters with a gain of 100 totally. PGG signals of vocal fold vibration are output for recording further blocking DC components.

In addition, the amplifier system is powered by two 3V button batteries. The 3.3V voltage source from a voltage regulator LP2950-3.3 supplies to the transimpedance amplifier and provides the reference voltage for the single-supply amplifier ICs. The output interface is made by a 3.5-mm phone jack with two cores to be linked to a recorder.

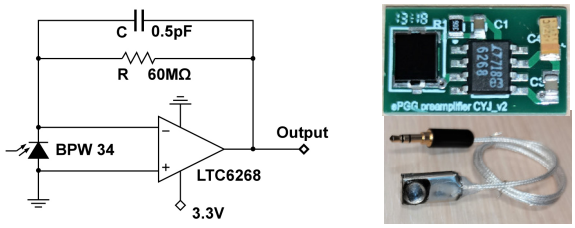


Figure 2, *Circuit diagram of the transimpedance amplifier (left) and the circuit board and sensor assembly photo (right)*

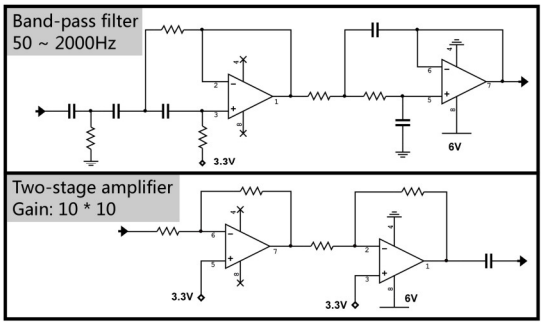


Figure 3, *Circuit diagram of the band-pass filter (top) and the two-stage amplifier (bottom)*

2.2. Experimental Setup

The experiment was conducted in a soundproof room with the device setup as shown in Fig. 4. Firstly, the electrodes of EGG were placed on each side of the larynx at the level of the glottis, and then a wearable microphone is positioned at about 5 cm from the mouth. Finally, the photodiode unit was pressed tightly on the skin at the thyroid notch, and the NIR emitter was attached below the cricoid cartilage. In this experiment, three channels of signals were recorded simultaneously with a data recorder (DAS40), including audio, P-PGG and EGG signals.

A Chinese female speaker (29 y. o., the first author) served as a subject, who pronounced single vowels /a/, /i/ and /u/ with modal, soft and hard voices, consonant-vowel syllables /pa/, /p^{h}a/, /ta/, /t^{h}a/, /ka/ and /k^{h}a/ (all in high tone), and an utterance "I come from Tianjin University" in English and Chinese. Each utterance was repeated five times. All the wave data are processed and presented with Python scripts. Pratt is also used for quick speech analyses.

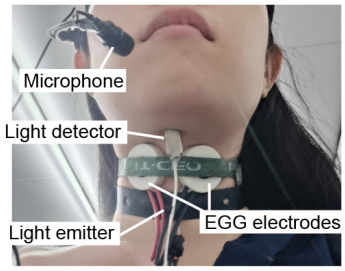


Figure 4, *Experimental setup for the combined measurement of audio, EGG and P-PGG signals*

3. RESULTS

The portable PGG (P-PGG) demonstrated stable and excellent performance in the experiments. With an oscilloscope, glottal cycles were clearly observed with a minimal level of disturbing noise. The system performance was satisfactory in amplifying the photocurrent of a nA level into a Volt level, with the signal-to-noise ratio above 20 dB. In the following, P-PGG and EGG waveforms are compared in detail.

3.1. Waveforms of single vowels

Fig. 5 shows examples of the audio, P-PGG and EGG waveforms for vowel /a/. Compared with modal voice, soft voice in Fig. 5 (b) exhibits attenuated vocal contacts during VF vibrations in EGG. Fig. 5 (c) and (d) present finer differences between EGG and P-PGG waveforms. The P-PGG demonstrates glottal area variations during the open phase of glottal cycles, often with biphasic or ripple patterns as seen in Fig. 5 (d). Fig. 5 (e) and (f) show transitions at the vowel onset and offset, where the P-PGG signals shows rises and falls in amplitude according to the glottal oscillation, which are distorted in EGG. In Fig. 5 (f), both P-PGG and audio signals show a gradual amplitude decay into silence at vowel offset, where the EGG signals no longer track such details.

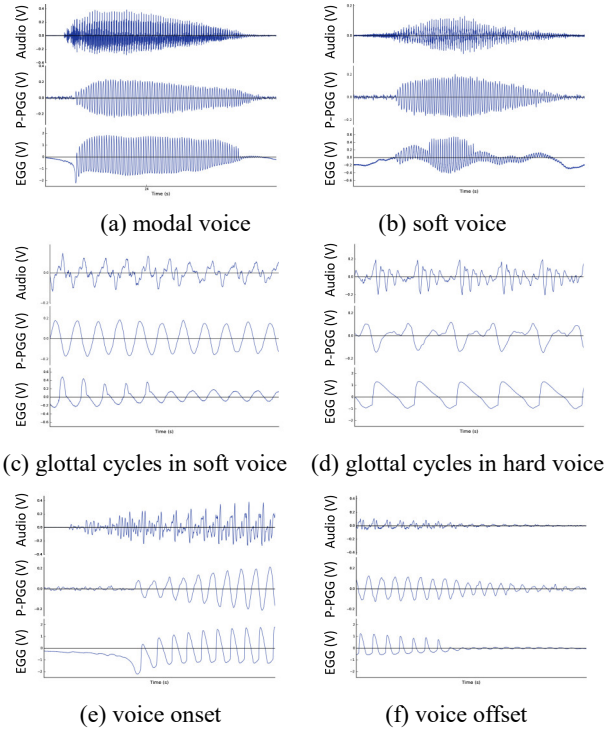


Figure 5, *Audio, P-PGG and EGG waveforms for vowel /a/*

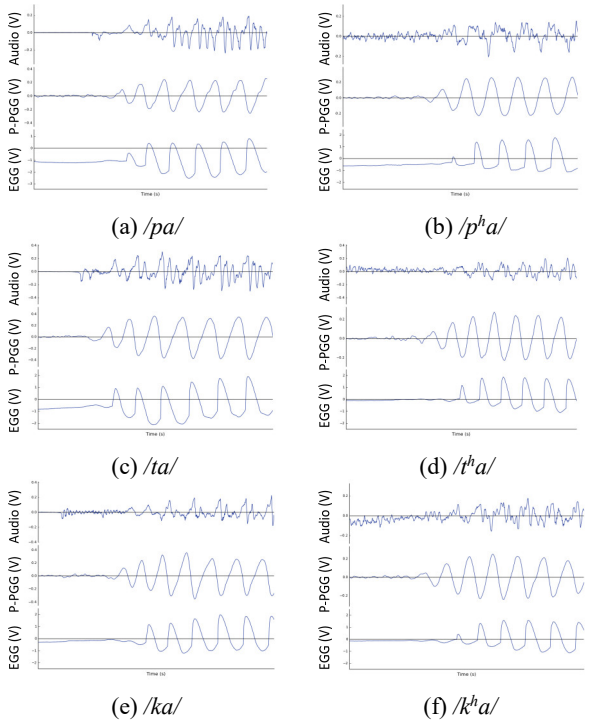


Figure 6, *Audio, P-PGG and EGG waveforms of voice onset for syllables /pa/, /p^{h}a/, /ta/, /t^{h}a/, /ka/ and /k^{h}a/*

3.2. Waveforms of consonant-vowel syllables

Vocal fold (VF) activities at voice onset were also documented by the P-PGG as distinctive information among stop consonants. As shown in Fig. 6, in voiceless aspirated stops, P-PGG signals show a few cycles before the initiation of VF contacts. This phenomenon is most obvious in the velar stops (both aspirated and unaspirated), which may be related to the interaction between the tongue root and larynx.

3.3. Waveforms of the utterance

Fig. 7 demonstrates successful measurements of P-PGG amplitude envelopes in natural speech. The spectrogram is also shown to remark utterance segments with fundamental frequency changes. For the example sentence, the syllables "come" and "Tian" are magnified and shown in Fig. 8. It is interesting to find that the amplitude of glottal oscillation increases slightly in the nasals /m/ and /n/. Thus, the P-PGG signal amplitude is uniquely different from that of audio and EGG waveforms. Yet, new evidence from more experiments is expected to explain the relationship among the P-PGG signal amplitude, phonetic types, and control of fundamental frequency.

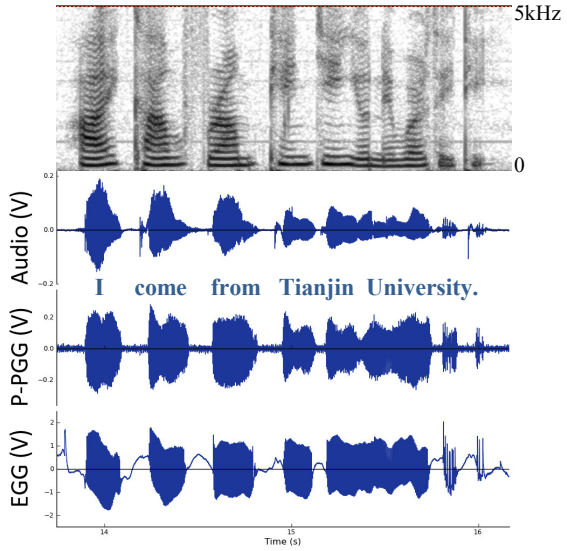
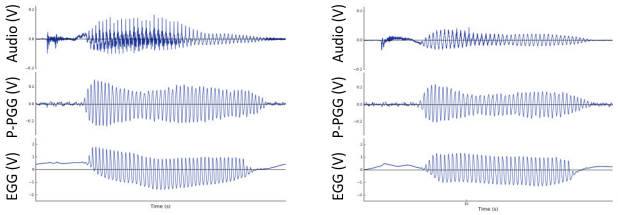


Figure 7, *Spectrogram, Audio, P-PGG and EGG waveforms for an English utterance*



(a) *"come"* (b) *"Tian"*

Figure 8, *Audio, P-PGG and EGG waveforms for the syllables "come" and "Tian"*

4. DISCUSSIONS AND CONCLUSION

The portable PGG (P-PGG) was successfully realized with extremely high-gain, low-noise circuits and minimal components for simple and stable measurements. With the P-PGG, vocal fold (VF) vibrations are conveniently monitored for a female speaker, in contrast to our common experience of low-quality ePGG signals from female speakers in the past. P-PGG data obtained from other several speakers in our ongoing examinations essentially agree with those in this report. The SNR of the waveforms may relate to individual differences of laryngeal dimensions and tissue properties of light transmission [23]-[25].

Simultaneous recordings of P-PGG and EGG signals with audio indicate the following observations:

- P-PGG waveforms correspond well with the open and close phases of glottal cycles, but they differ from the realistic glottal area function [26], [27].
- Glottal cycles in P-PGG waveforms tent to show sharp notches at glottal closure and ripples of various degrees during open periods.
- P-PGG waveforms accurately trace the complete VF vibrations from vowel onset to offset where EGG signals are ambiguous.

Potential research topics expected by the application of the P-PGG system are:

- Detailed phonetic descriptions with accuracy such as on voice onset time (VOT), onset F0, and voiceless duration.
- Vocal investigation into the relation of the P-PGG amplitude and VF contact pressure, which is directly relevant to voice quality and pathology.

To summarize, the P-PGG offers rich applications in experimental phonetics, vocal assessment, and wearable systems for monitoring speech. It is also necessary in the future to clarify the nature of the P-PGG signals to interpret signal characteristics and parameterize temporal glottal events.

Acknowledgement

This work was supported in part by National Key R&D Program of China (No. 2018YFC0806802), National Natural Science Foundation of China (No. 61876131, U1936102).

5. REFERENCES

[1] B. Sonesson, "A method for studying the vibratory movements of the vocal cords: a preliminary report," *Journal of Laryngology & Otology*, vol. 73, no. 11, pp. 732-737, 1959.

[2] B. Sonesson, "On the anatomy and vibratory patterns of the human vocal folds," (Thesis work) *Acta Oto-Laryngologica Supplement*, vol. 156, pp. 1-80, 1960.

[3] A. Malecot, and K. Peebles, "An optical device for recording glottal adduction-abduction during normal speech," *Zeitschrift fur Phonetik Sprachwissenschaft und Kommunikationsforschung*, vol. 18, no. 6, pp. 545-550, 1965.

[4] L. Lisker, A. S. Abramson, F. S. Cooper, and M. H. Schvey, "Transillumination of the Larynx in Running Speech," *Journal of the Acoustical Society of America*, vol. 45, no. 6, pp. 1544-1546, 1966.

[5] J. Ohala, "A new photo-electric glottograph,". *UCLA Working Papers in Phonetics*, vol. 4, pp. 40-52, 1966.

[6] M. Sawashima, and H. Hirose, "New laryngoscopic technique by use of fiber optics," *Journal of the Acoustical Society of America*, vol. 43, no. 1, pp. 168-169, 1968.

[7] M. Sawashima, "Movements of the larynx in articulation of Japanese consonants," *Annual Bulletin Research Institute of Logopedics and Phoniatrics*, University of Tokyo, vol. 2, pp. 11-20, 1968.

[8] J. Harden, "Comparison of glottal area changes as measured from ultrahigh-speed photographs and photoelectric glottographs," *Journal of Speech and Hearing Research*, vol. 18, no. 4, pp. 728, 1975.

[9] T. Baer, A. Löfqvist, and N. S McGarr, "Laryngeal vibrations: A comparison between high ‐ speed filming and glottographic techniques," *Journal of the Acoustical Society of America,* vol. 73, no. 4, pp. 1304-1308, 1983.

[10] W. Habermann, J. Jiang, E. Lin, and D. G. Hanson, "Correlation between glottal area and photoglottographic signal in normal subjects," *Acta Oto-laryngologica*, vol. 120, no. 6, pp. 778-782, 2000.

[11] J. J. Jiang, S. Tang, M. Dalal, C.-h. Wu, and D. G. Hanson, "Integrated analyzer and classifier of glottographic signals," *IEEE Transactions on Rehabilitation Engineering*, vol. 6, no. 2, pp. 227-234, 1998.

[12] B. R Gerratt,.D. G. Hanson, and G. S. Berke, "Laryngeal configuration associated with glottography," *American Journal of Otolaryngology*, vol. 9, no. 4, pp. 173-179, 1988.

[13] D. D. Deliyski, and R. E Hillman, "State of the art laryngeal imaging: research and clinical implications," *Current Opinion in Otolaryngology & Head and Neck Surgery*, vol. 18, no. 3, pp. 147-152, 2010.

[14] K. Honda, and S. Maeda, "Glottal opening and airflow pattern during production of voiceless fricatives: a new non ‐ invasive instrumentation," *Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 3738, 2008.

[15] K. Honda, and S. Maeda, "Non-invasive photoelectro-glottography method and device," U.S. Patent 2010/0256503, Oct. 7, 2010.

[16] J. Vaissière, K. Honda, A. Amelot, S. Maeda, and L. Crevier-Buchman, "Multisensor platform for speech physiology research in a phonetics laboratory," *Journal of the Phonetic Society of Japan*, vol. 14, no. 2, pp. 65-77, 2010.

[17] E. Suthau, P. Birkholz, A. Mainka, and A. P. Simpson, "Non-invasive photoglottography for use in the lab and the field," *Speech Communication; 12. ITG Symposium*, Paderborn, Germany, 2016

[18] A. Bouvet, A. Van Hirtum, X. Pelorson, S. Maeda, K. Honda, and A. Amelot, "Calibration of external lighting and sensing photoglottograph," *10th International Workshop Models and Analysis of the Vocal Emissions for Biomedical Applications (MAVEBA)*, Florence, Italy, 2017.

[19] A. Bouvet, A. Amelot, X. Pelorson, S. Maeda, and A. V. Hirtum, "External lighting and sensing photoglottography: Characterization and MSePGG algorithm," *Biomedical Signal Processing and Control*, vol. 51, pp. 318-327, 2019

[20] H. Kim, K. Honda, and S. Maeda, "ePGG, airflow and acoustic data on glottal opening in Korean plosives," *Proc. 18th Int. Conf. of Phonetics Sciences (ICPhS)*, Glasgow, UK, 2015.

[21] H. Kim, S. Maeda, K. Honda, and L. Crevier-Buchman, "The Mechanism and Representation of Korean Three-Way Phonation Contrast: External Photoglottography, Intra-Oral Air Pressure, Airflow, and Acoustic Data," *Phonetica*, vol. 75, no. 1, pp. 57-84, 2018.

[22] Y. Chi, K. Honda, and J. Wei, "Glottographic and Aerodynamic Analysis on Consonant Aspiration and Onset F0 in Mandarin Chinese," *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Brighton, United Kingdom, pp. 6480-6484, 2019.

[23] H. E. Eckel, and C. Sittel, "Morphometry of the larynx in horizontal sections," *American Journal of Otolaryngology*, vol. 16, no. 1, pp. 40-48, 1995.

[24] H. Eckel, C. Sittel, P. Zorowka, and A. Jerke, "Dimensions of the laryngeal framework in adults," *Surgical and Radiologic Anatomy*, vol. 16, no. 1, pp. 31-36, 1994.

[25] I. R. Titze, "Physiologic and acoustic differences between male and female voices," *Journal of the Acoustical Society of America*, vol. 85, no. 4, pp. 1699-1707, 1989.

[26] A. N. Bashkatov,. E. A. Genina, V. I. Kochubey, and V. V. Tuchin, "Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000 nm," *Journal of Physics D: Applied Physics*, vol. 38, no. 15, pp. 2543, 2005.

[27] A. N. Bashkatov,. E. A. Genina, and V. V. Tuchin, "Optical properties of skin, subcutaneous and muscle tissue: A review," *Journal of Innovation in Optical Health Science*, vol. 4, no. 1, pp. 9-38, 2011.