IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 March 2023, accepted 22 March 2023, date of publication 27 March 2023, date of current version 30 March 2023.

*Digital Object Identifier 10.1109/ACCESS.2023.3262122*

RESEARCH ARTICLE

# A Voice-Producing System With Naturalness and Variable Multi-Frequency Vocalization for Patients Who Have Undergone Laryngectomy

**TZU-YEN HUANG[1], BOR-SHING LIN[2], (Senior Member, IEEE), CHING-FENG LIEN[3], WING-HEI VIOLA YU[4], YUN-YANG PENG[5], AND BOR-SHYH LIN[5,6], (Senior Member, IEEE)**

[1]Department of Otorhinolaryngology–Head and Neck Surgery, International Thyroid Surgery Center, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung 80708, Taiwan
[2]Department of Computer Science and Information Engineering, National Taipei University, New Taipei 237303, Taiwan
[3]Department of Otolaryngology–Head and Neck Surgery, E-DA Hospital, I-Shou University, Kaohsiung 82445, Taiwan
[4]Department of Otolaryngology, Kaohsiung Medical University Hospital, Kaohsiung 80756, Taiwan
[5]Institute of Imaging and Biomedical Photonics, National Yang Ming Chiao Tung University, Tainan 71150, Taiwan
[6]Department of Medical Research, Chi Mei Medical Center, Tainan 71004, Taiwan

Corresponding author: Bor-Shyh Lin (borshyhlin@gmail.com)

This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST 110-2221-E-A49-096-MY3, Grant MOST 109-2221-E-305-001-MY2, Grant MOST 110-2314-B-305-001, and Grant MOST 111-2628-B-037-007; in part by the Grants of Kaohsiung Medical University Hospital, Kaohsiung Medical University (KMUH111-1R48); in part by the E-DA Hospital, I-Shou University (EDAHP106073); and in part by the Faculty Group Research Funding Sponsorship by National Taipei University (2022-NTPU-ORDA-03).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of E-DA Hospital, Kaohsiung, Taiwan, under Approval No. EMRP15104N.

**ABSTRACT** Patients with advanced laryngeal cancer usually mist undergo total laryngectomy, which causes communication difficulties. Several voice-producing devices have been developed to assist such patients in restoring their ability to speak. However, these devices produce monotonous sounds and require hand-held operation. A novel electronic voice-producing system was thus proposed to assist patients with laryngectomy in improving their speech performance. By imitating normal human speech, the proposed system enables the adjustment of the fundamental frequency of generated sound according to changes in lung pressure through exhalation; this enables different tones to be produced. Moreover, the proposed system is wearable and thus convenient to use. According to the experimental results, the proposed system provides changes in fundamental frequency corresponding to different tones and sentences in human speech, and outperforms existing devices in speech performance. A novel voice-producing system was proposed and implemented to assist patients who had undergone laryngectomy to resume basic communication activities.

**INDEX TERMS** Laryngeal cancer, total laryngectomy, communication ability, voice-producing devices, wearable design.

## I. INTRODUCTION

According to the 2013 Global Cancer Incidence and Mortality report of the World Health Organization, the annual incidence of laryngeal cancer worldwide is approximately 130,000, and the main risk factors for laryngeal cancer are tobacco smoking and alcohol consumption [1]. Laryngeal cancer causes pain and difficulty swallowing, and it hinders speech and social activities [2], [3]. Laryngectomy is one of the main treatments for patients with advanced laryngeal cancer. However, it results in a substantial change in anatomical structure. If the larynx is surgically removed through tracheostomy and the end of the vocal tract is shortened, the patient's ability to speak is compromised. Patients who

The associate editor coordinating the review of this manuscript and approving it for publication was Venkata Rajesh Pamula.

undergo laryngectomy lose the ability to speak [4]. Therefore, it is important to help patients with laryngeal cancer engage in basic communication after total laryngectomy.

Several approaches have been proposed to help patients communicate verbally after total laryngectomy, such as tracheoesophageal speech, standard esophageal speech, pneumatic artificial larynx speech, and electrolarynx speech. Tracheoesophageal speech and esophageal speech rely on the manipulation of pulmonary arteries [5]. For tracheoesophageal speech, a small device composed of silicone is inserted into an opening between the trachea and esophagus formed by surgery, allowing air from the lungs to produce sound through pharyngoesophageal shunt airflow. However, the quality of speech produced using the tracheoesophageal approach is not easy clear, and patients must receive vocalization training from experienced clinicians [6]. In the tracheoesophageal speech approach, the production of pharyngoesophageal speech is mainly based on aerodynamic factors, and some voice features, such as pitch and amplitude, cannot be adjusted by patients [7].

Standard esophageal speech is produced by using the esophagus as an air supply and the pharyngoesophageal segment located above the esophagus as a sound source [8]. This method does not require any equipment for voice generation. Compared with other approaches, in the esophageal speech approach, a longer training time is required, and the voices generated sound more natural. It is one of the most popular approach for speech restoration after total laryngectomy [9]. However, compared with normal speech, esophageal speech has a lower fundamental frequency and smaller amplitude [10], and individuals using esophageal speech may develop conditions such as gastroesophageal reflux [11].

A pneumatic artificial larynx is a noninvasive prosthesis that assists in vocalization. It comprises a sound conducting tube, a vibration chamber, an air entry tube, and a stoma cover. The stoma cover is placed tightly over the stoma, and the other end of the tuning tube is placed in the oral cavity. When an individual uses this prosthesis for speech, air, which is exhaled from a tracheal incision, vibrates through the reed of the vibration chamber to generate sound [12], [13]. A person must have the use of a hand for pneumatic artificial larynx operation, and the generated voice usually exhibits low-frequency noise, slight pitch changes, and poor sound quality; the speech produced is similar to that of Donald Duck [14]. Moreover, the tuning tube is easily clogged by saliva when placed in the mouth [15].

The electrolarynx is the most commonly used voice prosthesis. From the 1990s to the early 2000s, more than 50% of patients who underwent laryngectomy used the electrolarynx as their main communication method [16]. The electrolarynx is a hand-held, battery-powered electromechanical transducer that can produce mechanical sound [17]. During vocalization, the electrolarynx is placed on the neck near the glottis, and the vibrating electronic sound source is transmitted through the neck to the oral cavity and pharyngeal cavity. The user modulates this sound using their tongue, lips, and jaws [18]. However, the lack of natural fundamental frequency changes causes the device to emit unnatural and monotonous sounds [19], [20].

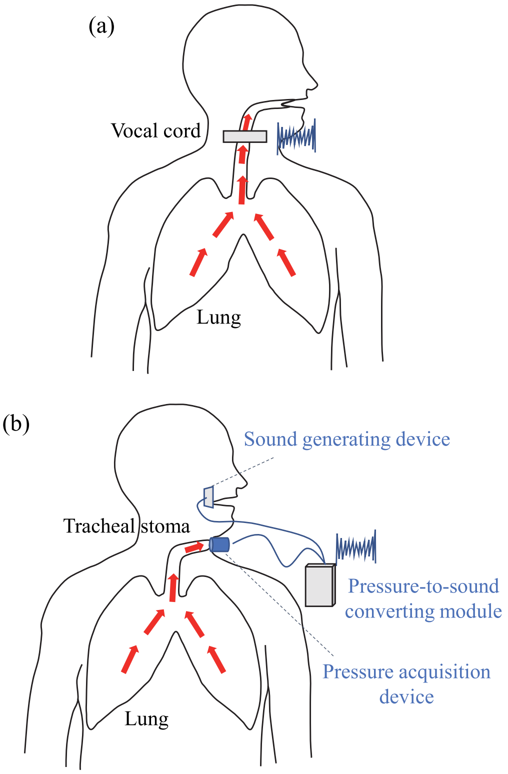

**FIGURE 1.** Basic mechanisms of (a) human speech and (b) the proposed voice production system.

To overcome problems of monotonous sound and limited amplitude, a novel electronic voice-producing system for patients with laryngectomy was proposed in this study. The proposed device is wearable and is convenient for everyday use. The device generates sounds in multiple frequencies, such as fundamental and harmonic frequencies; similar to the mechanism for normal human speech, when the proposed device is used, frequencies are changed according to lung pressure. The user can modulate the frequency and amplitude of the voice generated by altering their lung pressure and mouth shape. Finally, a clinical experiment and a questionnaire were designed to validate the vocalization performance of the proposed voice-producing system. The experimental results revealed that the four-tone speech sound and Mandarin Chinese sentence generated by the proposed voice-producing system were similar to those produced by human vocal cords. Moreover, the device outperformed other voice-producing devices in terms of vocalization quality.

## II. METHODS

### A. BASIC ARCHITECTURE OF THE PROPOSED ELECTRONIC VOICE-PRODUCING SYSTEM

The design of the proposed electronic voice-producing system is based on the mechanism underlying human vocalization, as illustrated in Fig. 1 (a). In laryngeal speech, air is

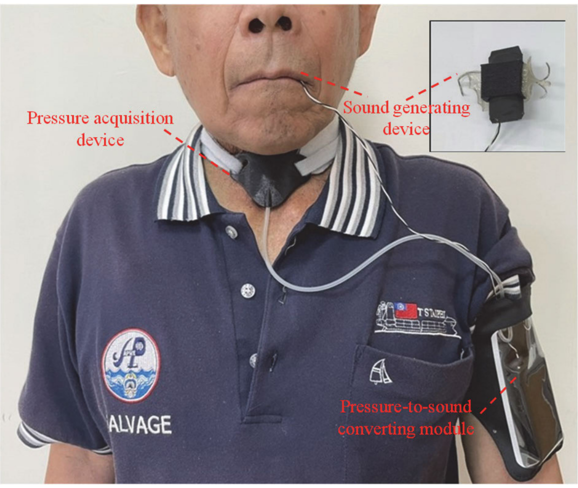

**FIGURE 2.** **Photograph of a person wearing the proposed voice production system.**

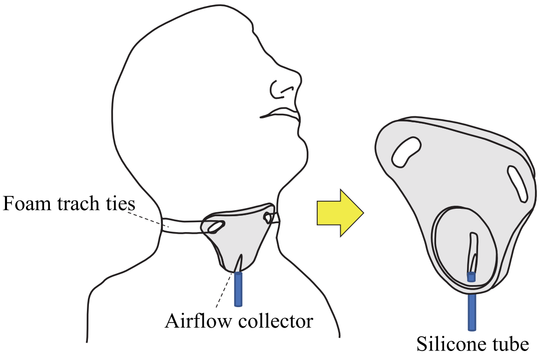

**FIGURE 3.** **Basic structure of the pressure acquisition device.**

exhaled from their lungs and a change in pressure causes the vocal cords to vibrate and generate sound. Patients whose vocal cords have been resected through total laryngectomy cannot generate sound by exhaling air from their lungs through their vocal cords. In these patients, a tracheal stoma is formed after surgery. From the tracheal stoma, air exhaled from the lungs can be easily accessed. Our system, which is designed to mimic the sound of human vocal cords, is illustrated in Fig. 1 (b). The proposed system comprises several parts, including a pressure acquisition device, a pressure-to-sound conversion module, and a sound generating device (Fig. 2). The pressure acquisition device is designed to be placed on the tracheal stoma of a person who has undergone total laryngectomy to acquire the airflow pressure produced through exhalation from the lungs. The pressure-to-sound conversion module is deigned to convert the acquired airflow pressure to a multifrequency signal. The sound generating device is placed on the upper jaw of the oral cavity to generate multifrequency sounds. Finally, the user can modulate the frequency and amplitude of the sound generated by altering their lung pressure and mouth shape.

### B. DESIGN OF THE PRESSURE ACQUISITION DEVICE

The basic structure of the pressure acquisition device is presented in Fig. 3. It contains an airflow collector, foam trach ties, and a silicone tube. The airflow collector is designed to effectively collect air pressure exhaled from the lungs, and it allows for sufficient air circulation for the user to breathe normally. It is composed of thermoplastic polyurethane (TPU), and it is tough and has high ductility. The device was made using a 3D printer. Foam trach ties (Posey 8197L, TIDI Products, Neenah, WI, USA) are used to stably fix the pressure acquisition device to the neck. The collected airflow is transmitted to the pressure-to-sound conversion module through a silicone tube.

### C. DESIGN OF THE PRESSURE-TO-SOUND CONVERSION MODULE

For the proposed system, an algorithm for converting air pressure to multifrequency sound was designed to convert the air pressure from exhalation into a multifrequency signal. The human voice mainly comprises a fundamental frequency and many harmonic frequencies. The fundamental frequency is the major contributor to human speech. Harmonic frequencies are generally located at the multiples of the fundamental frequency, and their contribution decreases as the frequency increases. To generate a multifrequency sound that is similar to the human voice, a sinusoidal model with multiple frequencies was used, and the generated multifrequency sound $S(t)$ can be expressed as

$$S(t) = sin\,(2\pi * F_0 * t) + \sum\nolimits_{k=1}^{l} \frac{1}{2k^2} sin(2\pi * F_0 * 2^k * t), \tag{1}$$

where $F_0$ denotes the fundamental frequency and $l$ is the number of harmonic frequencies. Moreover, the sum of $S(t)$ for different $F_0$ values is 0. The fundamental frequency bands for male and female speech are 100–146 and 188–221 Hz, respectively [21], [22]. Here, the range of the detected air pressure was divided into 50 levels in this study.

A block diagram of the pressure-to-sound conversion module is presented in Fig. 4; the module mainly comprises a microprocessor, pressure sensor, front-end amplifier, and speaker-driving circuit. The pressure sensor (HSCMRRN100MGAA3, Honeywell, Charlotte, NC, USA), which has a driving voltage of 3.3 V and can detect a pressure at a range of 0–100 mbar, is used to convert lung pressure signals into electrical signals. The acquired lung pressure signal is amplified by the front-end amplifier circuit. The gain of the front-end amplifier circuit is set to 12, and the frequency band of the band-pass filter in this amplifier circuit is set to 0.1–9 Hz. The preprocessed lung pressure signal is digitized by a 12-bit analog-to-digital converter (sampling rate: 80 Hz) in the module's microprocessor. According to the received lung pressure signal, a sound signal with specific amplitudes and frequencies is generated in the microprocessor through the pressure-to-sound conversion algorithm. This signal is then sent to a digital-to-analog converter, which has a sampling rate of 16 kHz. The generated multifrequency signal is then input into the speaker-driving circuit to drive the external sound generating device. In the speaker-driving

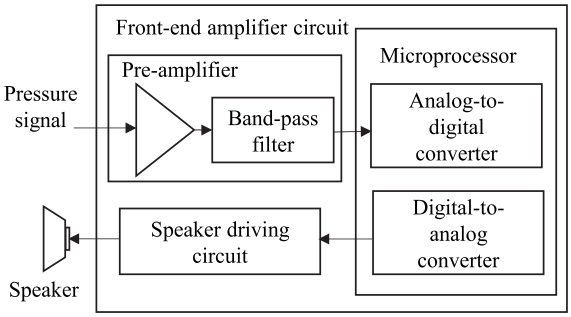

**FIGURE 4.** Block diagram of the press-to-sound conversion module.

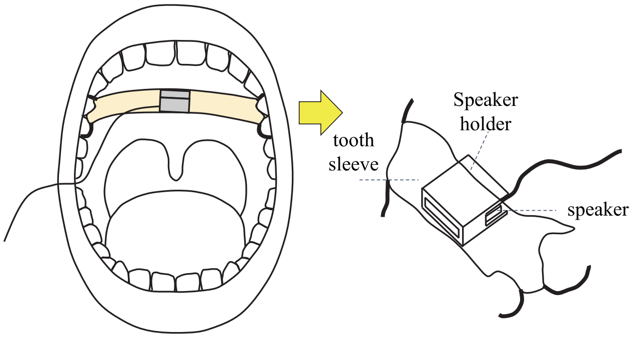

**FIGURE 5.** Basic structure of the sound generating device.

circuit, an audio power amplitude chip (LM386, National Semiconductor, Santa Clara, CA, USA) is used to provide power to a sound driving amplifier, which has a gain of 20 dB. The designed pressure-to-sound conversion module can be fixed on the arm by using an armband pocket (One Size Fits All, Kalenji, Lille, France).

### *D. DESIGN OF THE SOUND GENERATING DEVICE*

The basic structure of the sound generating device is presented in Fig. 5. It comprises a speaker holder, set of speakers (SW350804-1, DB Unlimited, Dayton, OH, USA), and a tooth sleeve. The speaker holder, which was manufactured through 3D printing (CR-10 V3, Creality, Shenzhen, China), is designed to form a mold and fix the speaker to the mold, allowing the mold and speaker to be fixed to the tooth sleeve. A user can use the tooth sleeve to fix the sound generating device to the inside of their mouth, and the springs of the tooth sleeve hold the molars, allowing the sound generation device to be fixed to the inside of upper jaw.

## III. EXPERIMENTAL RESULTS

### *A. CHARACTERISTICS OF VOICE TONES GENERATED BY DIFFERENT DEVICES*

This section provides a comparison of voices generated by human vocal cords, a pneumatic artificial larynx (One Size Fits All, Association of Laryngectomees, Atlanta, GA, USA), an electrolarynx (Provox TruTone EMOTE, Atos Medical, Bagneux, France), and the proposed system. A digital voice recorder (R328, BESTA, Taipei City, Taiwan) was used to record the four tones of Mandarin Chinese speech generated by different devices. The four tones in Mandarin Chinese speech are described as follows: the first tone is high and level, the second tone rises moderately, the third tone falls

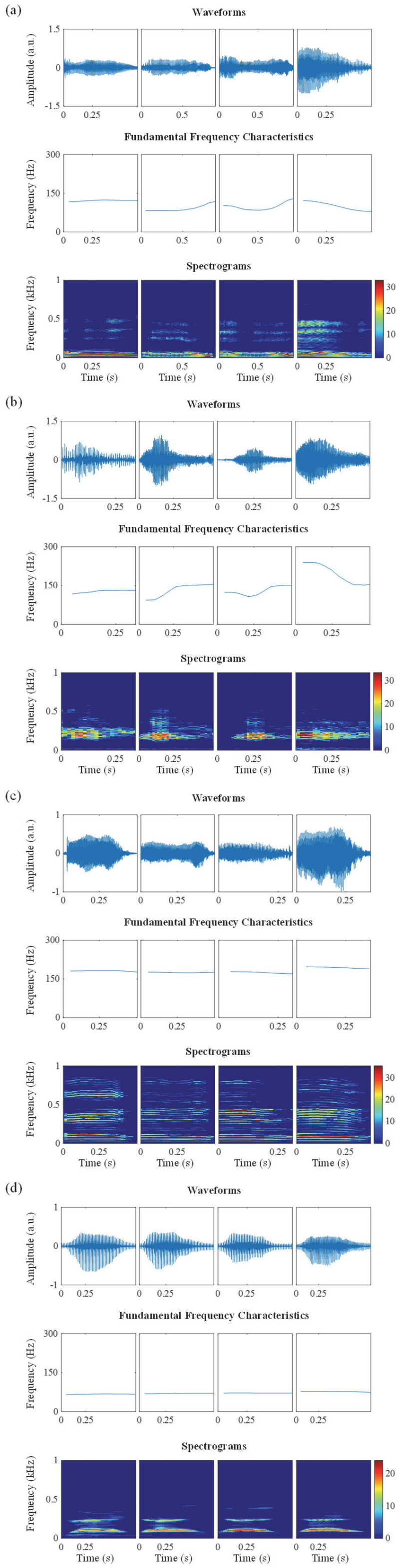

**FIGURE 6.** Voice characteristics corresponding to the a sound in four tones generated by (a) real human vocal cords, (b) the proposed system, (c) a pneumatic artificial larynx, and (d) an electrolarynx.

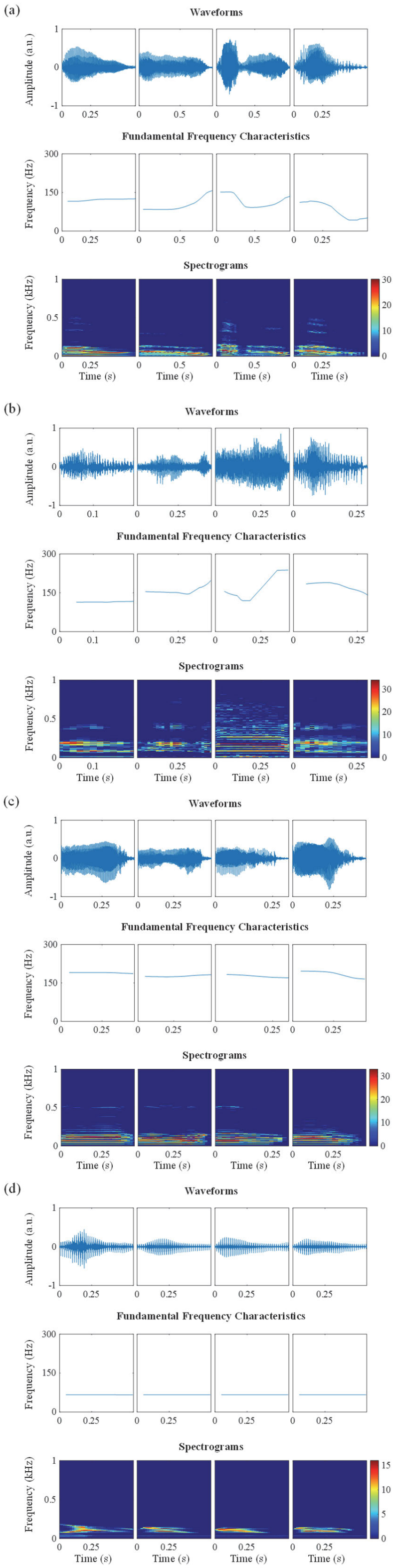

**FIGURE 7.** Voice characteristics corresponding to the o sound in four tones generated by (a) real human vocal cords, (b) the proposed system, (c) a pneumatic artificial larynx, and (d) an electrolarynx.

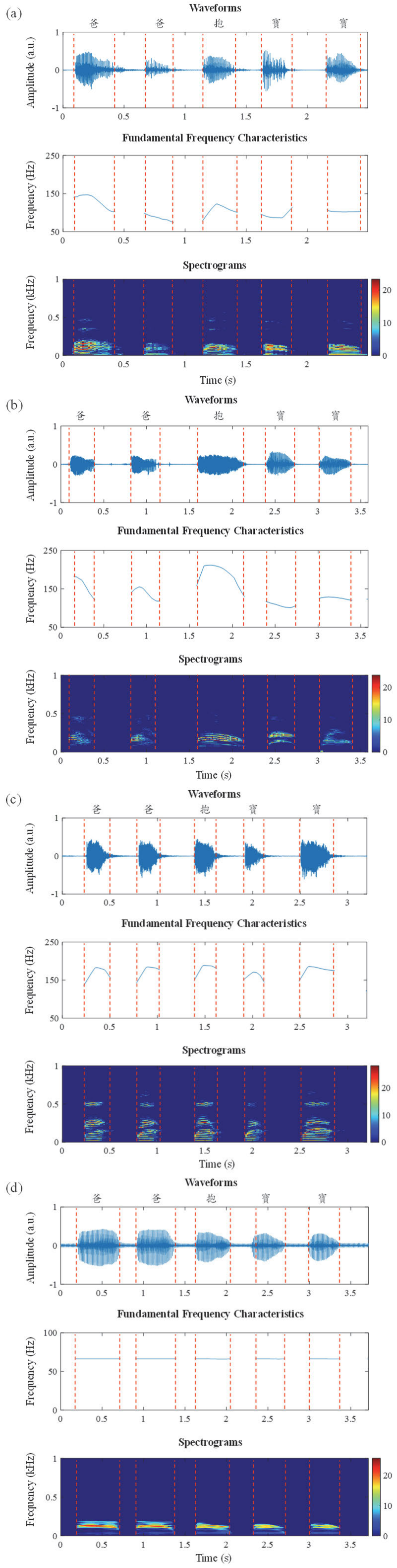

**FIGURE 8.** Voice characteristics during the generation of the sentence “爸爸抱寶寶” by (a) real human vocal cords, (b) the proposed system, (c) the pneumatic artificial larynx, and (d) the electrolarynx.

and then rises again, and the fourth tone starts out high but drops sharply to the bottom of the tonal range. The features of the aforementioned four tones can be regarded as the fundamental frequency as a function of time [23], [24]. Using different devices, the participants were instructed to vocalize vowels ($a$, $e$, $i$, $o$, and $u$) and the *ma* phoneme using four tones.

The study protocol was approved by the Institutional Review Board (approval number: EMRP15104N) of E-DA Hospital, Kaohsiung, Taiwan, and all patients provide informed consent prior to participation. Figs. 6 and 7 present the voice characteristics, such as waveforms, fundamental frequency, and spectrograms, corresponding to the *a* and *o* sounds generated by different devices. The fundamental frequency of the voices generated by the proposed system was similar to that generated by real human vocal cords; variations in fundamental frequency were obvious and fit the features of four-tone Mandarin Chinese speech. In terms of fundamental frequency, the pneumatic artificial larynx only generated a flat tone. Moreover, the spectrograms of the speech generated by the pneumatic artificial larynx contained a wide frequency band of strong harmonics and some strong low-frequency noise. To use an electrolarynx, a user touches the skin on their throat. A buzzing noise is generated, and the fundamental frequency of a voice generated by an electrolarynx is monotonous.

### B. CHARACTERISTICS OF SENTENCES GENERATED USING DIFFERENT DEVICES

The characteristics of Mandarin Chinese sentences generated using different devices was investigated. The participants were instructed to vocalize Mandarin Chinese sentences using different devices, such as “三輪車跑得快,” “爸爸抱寶寶,” and “紅茶很好喝.” The English translations of the above three sentences are respectively “the tricycle runs fast”, “dad holds the baby”, and “the black tea is delicious”. Fig. 8 presents the waveforms, fundamental frequency, and spectrograms corresponding to the Mandarin Chinese sentence “爸爸抱寶寶” generated by different devices. The experimental results revealed that the change in fundamental frequency generated by the proposed system was similar to that generated by real human vocal cords; the change in fundamental frequency was obvious and fit the features of the four Chinese Mandarin tones. However, the fundamental frequency generated by the pneumatic artificial larynx was constant and slight owing to exhalation by the participant. The participant could not change the fundamental frequency of the pneumatic artificial larynx through training because this device has the drawback of only outputting a constant frequency during vibrations. The fundamental frequency of the sound generated by the electrolarynx is monotonous.

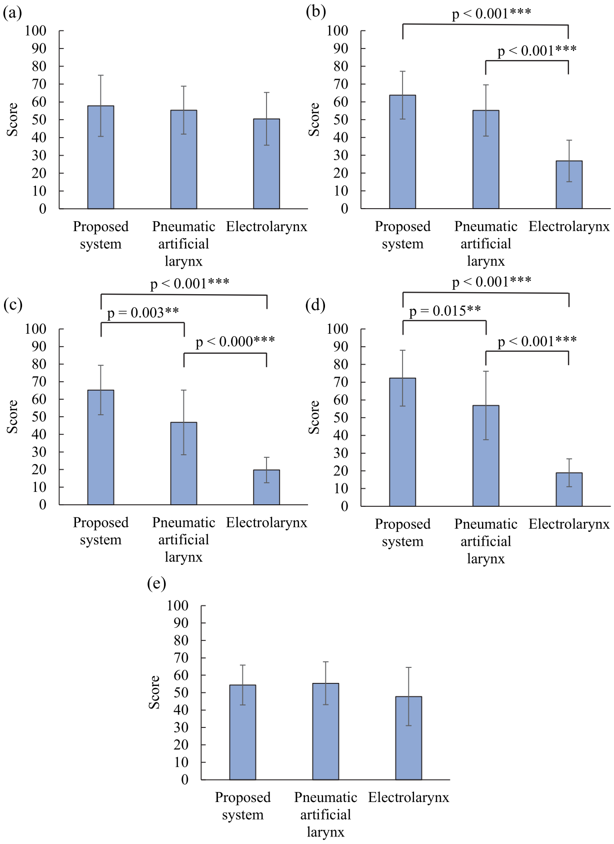

**FIGURE 9.** **Speech performance of different devices in terms (a) speech intelligibility, (b) speech tone, (c) naturalness, (d) noise, and (e) speech rate.**

### C. COMPARISON OF THE SPEECH PERFORMANCE OF DIFFERENT DEVICES

The speech performance (in terms of speech intelligibility, speech tone, naturalness, noise, and speech rate) of different devices was evaluated through questionnaires. Speech intelligibility is related to speech comprehension. Speech tone is related to frequency change. Naturalness indicates the similarity to speech produced by human vocal cords. Noise indicates the effect of external noise. Speech rate represents the similarity to the speed of human speech. In total, 17 listener participants participated in the experiment. The instructor first introduced the five aforementioned speech performance metrics and the different voice-producing devices to the participants. Six Mandarin Chinese sentences, namely “我看見花園有八隻蝴蝶,” “姊姊聞到廚房有六道菜,” “他聽到房間有人在唱歌,” “爸爸吃了桌上一盒餅乾,” “你看見操場上四顆足球,” “媽媽拿走信箱七封信,” were produced using different devices by the patient. The English translations of the above six sentences are respectively “I saw eight butterflies in the garden”, “sister smelled six dishes in the kitchen”, “he heard someone singing in the room”, “dad ate a box of cookies on the table”, “you saw four footballs on the playground”, and “mother took seven letters from the mailbox”. The instructor randomly selected and played recordings of these sentences generated by the patient using each voice-producing device. Subsequently, the participants were instructed to objectively evaluate these recordings and complete a questionnaire designed to score such recordings in terms of the aforementioned speech performance metrics. Fig. 9 presents the speech performance levels corresponding to different devices. A *t* test was used to determine significant differences ($p < 0.05$). The speech intelligibility of the proposed system was better than that of the pneumatic artificial larynx and electrolarynx. The speech performance in terms of speech tone, naturalness, and noise of the proposed system was significantly better

**TABLE 1.** Characteristics and performance of proposed system and competing vocalization devices for comparison.

| | **Proposed system** | **Pneumatic artificial larynx** [14] | **Pneumatic bionic voice with a statistical approach** [25], [26] | **Electrolarynx** [27], [28], [29] | **Intra-oral electrolarynx with fingertip switch** [31] | **Electrolarynx with sEMG and electromagnetic mechanisms** [4] |
|---|---|---|---|---|---|---|
| **Frequency range of sound (Hz)** | 100 ~ 250 | 150 ~ 185 | 80 ~ 140 | 67 | 125 | 100 |
| **Usage mode** | Wearable mode | Hand-held mode | Hand-held mode | Hand-held mode | Wearable mode | Hands-free mode |
| **Size (mm)** | 86 × 54 × 14.5 | 260 × 40 × 40 | - | 108 × 35 × 35 | 120 × 80 × 25 | 37 × 18 × 5 |
| **Power supply** | Battery (3.3V) | - | - | Battery (9V) | Battery (9V) | - |
| **Vocalization approach** | Conversion from exhaled air pressure to variable multi-frequency sound | Conversion from exhaled air pressure to monotonous sound | Conversion from exhaled air pressure to monotonous sound | Vibration of throat skin | Monotonous sound generator controlled via infrared communication. | A new electro-magnetic transducer as artificial excitation source |
| **Advantages** | Easy operation, low cost, naturalness and variable multi-frequency vocalization | Easy operation, and low cost | Easy operation, and low cost | Easy operation | Easy operation | Easy operation, sEMG to automatically adjust the fundamental frequency |
| **Limitations** | Requirement of training procedure | Monotonous vocalization, conspicuous operation, requirement of training procedure | Monotonous vocalization, conspicuous operation, requirement of training procedure | Monotonous vocalization, conspicuous operation, and higher cost | Monotonous vocalization | Requires peripheral hardware and has not yet been integrated into a small device that can be worn on the body |

than that of the pneumatic artificial larynx and electrolarynx. Moreover, the speech performance in terms of naturalness and noise of the pneumatic artificial larynx was significantly better than that of the electrolarynx. Differences between the speech rate of the proposed system, pneumatic artificial larynx, and electrolarynx were nonsignificant. In summary, the proposed system has superior speech performance relative to the other devices, and the electrolarynx had the poorest performance of the tested voice-producing devices.

## IV. DISCUSSION

The experimental results demonstrate that for four-tone speech sounds and Mandarin Chinese sentences, the fundamental frequency change in speech sounds generated by the proposed system is similar to that of human vocal cords. Through a suitable training procedure, users can use the proposed system to induce multifrequency changes by altering their lung pressure through exhalation. The pneumatic artificial larynx and electrolarynx cannot generate fundamental frequency changes corresponding to the four-tone speech sounds because of device design limitations. The pneumatic artificial larynx produces speech sounds through reed vibration, and the speech is produced when the frequency exceeds a certain level. The fundamental frequency generated by the pneumatic artificial larynx is generally constant (over 150 Hz); it changes slightly change during changes in lung airflow during exhalation. Moreover, the fundamental frequency of the electrolarynx is constant. The electrolarynx is a hand-held device, which reduces its convenience of use. All the aforementioned voice-producing devices allow patients who have undergone laryngectomy to engage in rudimentary vocalization. The speech tone of the proposed system is markedly better than that of other devices because of its ability to produce different tones. The slight change in fundamental frequency of the pneumatic artificial larynx results from changes in lung airflow; such changes cannot be easily controlled by the user. The electrolarynx can only generate a constant vibration frequency; therefore, its speech tone performance is inferior to that of other devices. The naturalness of the proposed system is also better than that of other devices because naturalness is directly related to speech tone. The pneumatic artificial larynx easily generates speech with obvious low-frequency noise, and the electrolarynx usually generates a noticeable buzzing noise. However, the differences between the speech rates of different devices were nonsignificant.

Several voice-assisted speech systems have been developed, including the pneumatic artificial larynx, conventional neck-type electrolarynx, silent electrolarynx, and electrolaryngeal speech with intraoral vibrator. Table 1 presents the characteristics of these devices for comparison. In 1859, the pneumatic artificial larynx was proposed. Exhaled air from the lungs vibrates the rubber reed inside the device to generate sound, and the generated sound is then adjusted through changes in mouth shape. The advantages of this device are its ease of use and loud volume. However, it must be operated using a hand, and the sound generated is a monotone. Moreover, after prolonged use, the elasticity of the spring leaf changes, causing the sound emitted to become more unnatural [14]. Ahmadi et al. proposed a statistical approach for source excitation generation to implement pneumatic bionic voice [25], [26]. In theory, it will be better than the traditional pneumatic artificial larynx, but the author has not tested it on patients. In 1957, the conventional neck-type electrolarynx was first proposed and applied [27]. The main components of an electrolarynx include a transducer, a transducer driving circuit, and a battery pack. An electrolarynx provides a periodic mechanical vibration to the neck or mouth skin to replace the vibration of the vocal cords and construct speech. Its advantages include ease of operation [28]. However, this device must be placed on the neck or mouth using a hand. It generates only monotones, and speech intelligibility is poor [29]. In 2010, the silent electrolarynx approach was introduced; it involves the use of an inaudible murmur to generate speech through voice conversion. This inaudible murmur is produced using an electret condenser microphone that is covered with urethane elastomer for adherence to the skin [30]. In 2005, electrolaryngeal speech through an intraoral vibrator device was proposed in Japan. The device comprises a wireless miniature fingertip switch, a denture-base intraoral vibrator, and a controller. The denture-based intraoral vibrator is fixed on the teeth as a replacement for the human vocal cords, and the vibrator is connected to the controller through a wire. The wireless miniature fingertip switch sends binary commands to the controller to adjust speech sounds (e.g., accent), and the generated sound is adjusted by through changes in mouth shape. The device does not require the use of a hand, but the generated sound is monotonous [31]. The fundamental frequency generated in the electrolarynx is usually fixed, so it is also the reason for the unpleasant sound of the electrolarynx. In 2014, Ahmadi et al. proposed to use the analysis of surface electromyography (sEMG) signals to learn the vocalization of healthy people and build models [32]. The system can process the sEMG signal through the trained Gaussian mixture model to adjust the fundamental frequency of speech to improve the user's pronunciation. In 2016, Fuchs et al improved the system of [32] to propose an electrolarynx with electromagnetic mechanisms [4]. They proposed a new electro-magnetic transducer as artificial excitation source and provided a proof-of-concept for this kind of transducer. Through this design, the waveform shape of the excitation signal can be effectively and automatically changed so that it can produce an effect similar to natural sound. But the whole system is a bit bulky and still not suitable for wearing on the human body. Different from the aforementioned voice-producing devices that require hand-held operation, the proposed system is wearable, making it convenient for everyday use. Moreover, the fundamental frequency range of the proposed system varies from 100 Hz to 250 Hz; this frequency is controlled through air pressure to provide favorable speech performance. The cost of the proposed system is low. Similar to other voice-producing devices, the operation of the proposed system also requires simple training.

## V. CONCLUSION

A novel electronic voice-producing system was proposed and implemented to assist the patients who have undergone laryngectomy to resume basic communication activities. The proposed system mimics human pronunciation. Specifically, fundamental frequency changes can be controlled according to changes in exhaled lung pressure to form different tones, and sounds can then be reconstructed by changing the mouth shape, tongue position, and generated multifrequency sound. As expected, the experimental results revealed that the fundamental frequency changes of the voice generated by the proposed system for four tones and different Mandarin Chinese sentences were similar to those of human vocal cords. Moreover, the questionnaire revealed that the proposed system outperformed the pneumatic artificial larynx and electrolarynx in terms of most speech performance metrics. However, the speech rate was still insufficient, and users of the proposed system require some (albeit rudimentary) training. Different from other voice-producing devices that require hand-held operation and generate monotonous sounds, the proposed system can provide more natural speech and is convenient for everyday use. Therefore, the system is promising as a device that enables communication (especially in tonal languages) for patients who have undergone laryngectomy.

## REFERENCES

[1] A. Pereira da Silva, T. Feliciano, S. Vaz Freitas, S. Esteves, and C. Almeida e Sousa, "Quality of life in patients submitted to total laryngectomy," *J. Voice*, vol. 29, no. 3, pp. 382–388, May 2015.

[2] M. P. J. Offerman, J. F. A. Pruyn, M. F. de Boer, J. J. V. Busschbach, and R. J. Baatenburg de Jong, "Psychosocial consequences for partners of patients after total laryngectomy and for the relationship between patients and partners," *Oral Oncol.*, vol. 51, no. 4, pp. 389–398, Apr. 2015.

[3] T. D. Woodard, A. Oplatek, and G. J. Petruzzelli, "Life after total laryngectomy: A measure of long-term survival, function, and quality of life," *Arch. Otolaryngol. Head Neck Surg.*, vol. 133, no. 6, pp. 526–532, 2007.

[4] A. K. Fuchs, M. Hagmuller, and G. Kubin, "The new bionic electro-larynx speech system," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 5, pp. 952–961, Aug. 2016.

[5] T. A. Bohnenkamp, T. Stowell, J. Hesse, and S. Wright, "Speech breathing in speakers who use an electrolarynx," *J. Commun. Disorders*, vol. 43, no. 3, pp. 199–211, May 2010.

[6] J. P. Searl, M. A. Carpenter, and C. L. Banta, "Intelligibility of stops and fricatives in tracheoesophageal speech," *J. Commun. Disorders*, vol. 34, no. 4, pp. 305–321, Jul. 2001.

[7] J. Searl, "Whispering by individuals using tracheoesophageal speech," *J. Voice*, vol. 32, no. 1, pp. 1–13, Jan. 2018.

[8] H. Liu, M. Wan, S. Wang, X. Wang, and C. Lu, "Acoustic characteristics of Mandarin esophageal speech," *J. Acoust. Soc. Amer.*, vol. 118, no. 2, pp. 1016–1025, Aug. 2005.

[9] H. Doi, T. Toda, K. Nakamura, H. Saruwatari, and K. Shikano, "Alaryngeal speech enhancement based on one-to-many eigenvoice conversion," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 1, pp. 172–183, Jan. 2014.

[10] H. Doi, K. Nakamura, T. Toda, H. Saruwatari, and K. Shikano, "Esophageal speech enhancement based on statistical voice conversion with Gaussian mixture models," *IEICE Trans. Inf. Syst.*, vols. E93–D, no. 9, pp. 2472–2482, 2010.

[11] J. G. Mathis, G. A. Lehman, J. C. Shanks, E. D. Blom, and R. L. Brunelle, "Effect of gastroesophageal reflux on esophageal speech," *J. Clin. Gastroenterol.*, vol. 5, no. 6, pp. 503–507, 1983.

[12] J. J. Xu, X. Chen, M. P. Lu, and M. Z. Qiao, "Perceptual evaluation and acoustic analysis of pneumatic artificial larynx," *Otolaryngology–Head Neck Surgery*, vol. 141, no. 6, pp. 776–780, Dec. 2009.

[13] J. B. Moon, J. W. Folkins, A. E. Smith, and E. S. Luschei, "Air pressure regulation during speech production," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 54–63, 1993.

[14] F. Ahmadi, F. Noorian, D. Novakovic, and A. van Schaik, "A pneumatic bionic voice prosthesis—Pre-clinical trials of controlling the voice onset and offset," *PLoS ONE*, vol. 13, no. 2, Feb. 2018, Art. no. e0192257.

[15] S. Bien, A. Rinaldo, C. E. Silver, J. J. Fagan, L. W. Pratt, C. Tarnowska, E. Towpik, N. Weir, B. J. Folz, and A. Ferlito, "History of voice rehabilitation following laryngectomy," *Laryngoscope*, vol. 118, no. 3, pp. 453–458, Mar. 2008.

[16] S. R. Cox and P. C. Doyle, "The influence of electrolarynx use on postlaryngectomy voice-related quality of life," *Otolaryngology–Head Neck Surgery*, vol. 150, no. 6, pp. 1005–1009, Jun. 2014.

[17] L. Wu, C. Wan, S. Wang, and M. Wan, "Improvement of electrolaryngeal speech quality using a supraglottal voice source with compensation of vocal tract characteristics," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 7, pp. 1965–1974, Jul. 2013.

[18] H. Liu and M. L. Ng, "Electrolarynx in voice rehabilitation," *Auris Nasus Larynx*, vol. 34, no. 3, pp. 327–332, Sep. 2007.

[19] Y. Saikachi, K. N. Stevens, and R. E. Hillman, "Development and perceptual evaluation of amplitude-based F0 control in electrolarynx speech," *J. Speech, Lang., Hearing Res.*, vol. 52, no. 5, pp. 1360–1369, Oct. 2009.

[20] K. F. Nagle, T. L. Eadie, D. R. Wright, and Y. A. Sumida, "Effect of fundamental frequency on judgments of electrolaryngeal speech," *Amer. J. Speech-Language Pathol.*, vol. 21, no. 2, pp. 154–166, May 2012.

[21] S. Barreda, "Investigating the use of formant frequencies in listener judgments of speaker size," *J. Phonetics*, vol. 55, pp. 1–18, Mar. 2016.

[22] K. Pisanski and D. Rendall, "The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness," *J. Acoust. Soc. Amer.*, vol. 129, no. 4, pp. 2201–2212, Apr. 2011.

[23] Y.-C. Hao, "Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers," *J. Phonetics*, vol. 40, no. 2, pp. 269–279, Mar. 2012.

[24] W. Li, Q. Zhaopeng, F. Yijun, and N. Haijun, "Design and preliminary evaluation of electrolarynx with F0 control based on capacitive touch technology," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 3, pp. 629–636, Mar. 2018.

[25] F. Ahmadi and T. Toda, "Designing a pneumatic bionic voice prosthesis—A statistical approach for source excitation generation," in *Proc. Interspeech*, Hyderabad, India, Sep. 2018, pp. 3142–3146.

[26] F. Ahmadi, K. Kobayashi, and T. Toda, "Development of a real-time bionic voice generation system based on statistical excitation prediction," in *Proc. ASSETS*, Pittsburgh, PA, USA, 2019, pp. 655–657.

[27] L. D. Lowry, "Artificial larynges: A review and development of a prototype self-contained intra-oral artificial larynx," *Laryngoscope*, vol. 91, no. 8, pp. 1332–1355, 1981.

[28] K. Xiao, S. Wang, M. Wan, and L. Wu, "Radiated noise suppression for electrolarynx speech based on multiband time-domain amplitude modulation," *IEEE/ACM Trans. Audio, Speech, Lang., Process.*, vol. 26, no. 9, pp. 1585–1593, Sep. 2018.

[29] E. A. Goldstein, J. T. Heaton, C. E. Stepp, and R. E. Hillman, "Training effects on speech production using a hands-free electromyographically controlled electrolarynx," *J. Speech, Lang., Hearing Res.*, vol. 50, no. 2, pp. 335–351, Apr. 2007.

[30] T. Hirahara, M. Otani, S. Shimizu, T. Toda, K. Nakamura, Y. Nakajima, and K. Shikano, "Silent-speech enhancement using body-conducted vocal-tract resonance signals," *Speech Commun.*, vol. 52, no. 4, pp. 301–313, Apr. 2010.

[31] H. Takahashi, M. Nakao, Y. Kikuchi, and K. Kaga, "Alaryngeal speech aid using an intra-oral electrolarynx and a miniature fingertip switch," *Auris Nasus Larynx*, vol. 32, no. 2, pp. 157–162, Jun. 2005.

[32] F. Ahmadi, M. A. Ribeiro, and M. Halaki, "Surface electromyography of neck strap muscles for estimating the intended pitch of a bionic voice source," in *Proc. BioCAS*, Lausanne, Switzerland, 2014, pp. 37–40.

**TZU-YEN HUANG** received the M.D. degree from National Taiwan University, Taipei, Taiwan, in 2009. He is currently pursuing the Ph.D. degree with the Department of Biological Science and Technology, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. He had postgraduate training with the Department of Otorhinolaryngology–Head and Neck Surgery, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan, where he is currently an Attending Physician. He focuses his clinical research on thyroid and parathyroid surgery, intraoperative neuromonitoring, and post-thyroid procedures dysphonia.

**BOR-SHING LIN** (Senior Member, IEEE) received the B.S. degree in electrical engineering from National Cheng Kung University, Taiwan, in 1997, and the M.S. and Ph.D. degrees in electrical engineering from National Taiwan University, Taiwan, in 1999 and 2006, respectively. He is currently a Distinguished Professor with the Department of Computer Science and Information Engineering, National Taipei University, Taiwan, where he is also the Director of the Computer and Information Center. His research interests include the areas of smart medicine, embedded systems, wearable systems, biomedical signal processing, biomedical image processing, and portable biomedical electronic system design.

**CHING-FENG LIEN** received the M.B.B.S. degree from Kaohsiung Medical University, Taiwan, in 2003, and the Ph.D. degree from the Institute of Imaging and Biomedical Photonics, National Chiao Tung University, Taiwan, in 2020. He is currently an Attending Doctor and the Director of the Division of Rhinology, Department of Otolaryngology–Head and Neck Surgery, E-DA Hospital. His research interests include the development of biomedical devices, sinonasal diseases, and head and neck surgery.

**WING-HEI VIOLA YU** received the B.A. degree in speech and hearing science from the Department of Speech Pathology and Audiology, University of Iowa, USA, in 1994, and the M.A. degree in communication disorders from Truman State University (formerly Northeast Missouri State University), USA, in 1996. She has been a speech-language pathologist in the USA and Taiwan. She is currently a Speech-Language Pathologist with the Department of Otolaryngology, Kaohsiung Medical University Hospital, Taiwan. She is also the Chief of Directors of the Kaohsiung Speech-Language Pathologists Union.

**YUN-YANG PENG** received the M.S. degree from the Institute of Imaging and Biomedical Photonics, National Yang Ming Chiao Tung University, Taiwan. His research interests include the areas of biomedical circuits and systems, biomedical signal processing, and biomedical optoelectronics.

**BOR-SHYH LIN** (Senior Member, IEEE) received the B.S. degree from National Chiao Tung University, Hsinchu, Taiwan, in 1997, and the M.S. and Ph.D. degrees from the Institute of Electrical Engineering, National Taiwan University, Taipei, Taiwan, in 1999 and 2006, respectively. He is currently a Distinguished Professor and the Director of the Institute of Imaging and Biomedical Photonics, National Yang Ming Chiao Tung University. His current research interests include biomedical circuits and systems, biomedical signal processing, and biosensors. He is a fellow of the Institution of Engineering and Technology (IET), U.K.

• • •