Speech Communication 89 (2017) 17–24

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

Vocal efficiency of electrolaryngeal speech production

Liang Wu, Ke Xiao, Supin Wang, Mingxi Wan*

The Key Laboratory of Biomedical Information Engineering of the Ministry of Education, Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi' an 710049, People's Republic of China

ARTICLE INFO

Article history:
Received 11 January 2016
Revised 7 February 2017
Accepted 8 February 2017
Available online 22 February 2017

Keywords:
Electrolaryngeal speech
Energy transfer
Vocal efficiency

ABSTRACT

From the perspective of efficiency, this article studied the energy transfer and conversion in the process of electrolaryngeal (EL) speech production. An overall vocal efficiency of EL speech production was defined as the ratio of the acoustic power of the EL speech to the electric power supplied by the battery. The measurements of a commercial EL showed that the actual utilization efficiency of the battery energy was no more than 0.1%. The energy transfer process was divided into three successive stages. The corresponding efficiencies of these stages were defined and estimated to analyze potential power losses and possible impact of two factors (EL cap and vowel) on the vocal efficiency. It was concluded that the non-linear transducer of the EL device and the physiological features of the neck tissue were the main reasons for the high power losses and low vocal efficiency. Furthermore, both EL cap and phonation vowel showed significant effects on the EL vocal efficiency. Thus, improvement of EL linear vibrator and compatible cap will be beneficial to raising the vocal efficiency and improving the EL speech quality.

1. Introduction

Normal speech production is an energy conversion process from the aerodynamic power provided by the pulmonary system to the acoustic power radiated from the mouth. Vocal efficiency (VE), defined as the ratio of the acoustic power to the sub-glottal power, is widely used to quantitatively evaluate the functional status of the larynx and vocal tract system (Howard et al., 1990; Titze, 1992; Jiang et al., 2004; Grillo et al., 2008).

For alaryngeal speech, electrolaryngeal (EL) speech production is also an energy conversion process but different from normal speech production. First, the sound source of EL speech is a mechanical vibration of a vibrator head driven by a piston that is connected to an electric motor. Then, this vibration transmits through the neck tissue and vocal tract to produce EL speech at the lips. In this process, the electric power of the battery is transformed into the mechanical power of the EL vibrator and then into the acoustic power of the EL speech.

Regarding energy conversion, the vocal efficiency of EL speech production is the integrative performance of the EL device, human neck tissue, and vocal tract, which must affect the quality of EL speech communication. For example, the high noise level in the EL speech resulting from the leakage of EL vibration should be due to a poor coupling between the EL vibrator and neck tissue. The low intensity of EL speech may be related to the high losses of energy in the speech production process, such as source generation of the EL device and sound transmission through neck tissue and vocal tract, etc. (Meltzner et al., 2003; 2005) reported that these two aspects are noted in the top acoustic aberrant properties reducing the perceptual intelligibility of EL speech, so reducing the EL self-noise and correcting for the low frequency energy will enhance the EL speech quality.

Therefore, it is necessary to quantitatively measure the vocal efficiency of EL speech production. The EL vocal efficiency may reveal the influence of EL speech production on the perception-related acoustic features (such as intensity, leaking noise, and energy distribution in frequency domain, etc.), which will benefit to the improvement of EL technology. However, the definition and calculation of the vocal efficiency of normal speech production cannot be directly applied to the case of EL speech production because of different energy forms and physiological structures of the two vocal systems. So far as we know, there are few reports about the vocal efficiency of EL speech production.

In this work, we defined and measured some vocal efficiencies of electrolaryngeal speech production, including an overall vocal efficiency as well as the efficiencies of three successive stages in the process. Then, the values measured from healthy subjects and laryngectomees and differences of these efficiencies were analyzed to investigate the potential power losses and possible influence of two factors (EL cap and vowel) on EL vocal efficiency. Finally, some suggestions of raising vocal efficiency were mentioned for the improvement of EL technology and speech quality.

* Corresponding author.
E-mail address: mxwan@mail.xjtu.edu.cn (M. Wan).

http://dx.doi.org/10.1016/j.specom.2017.02.002

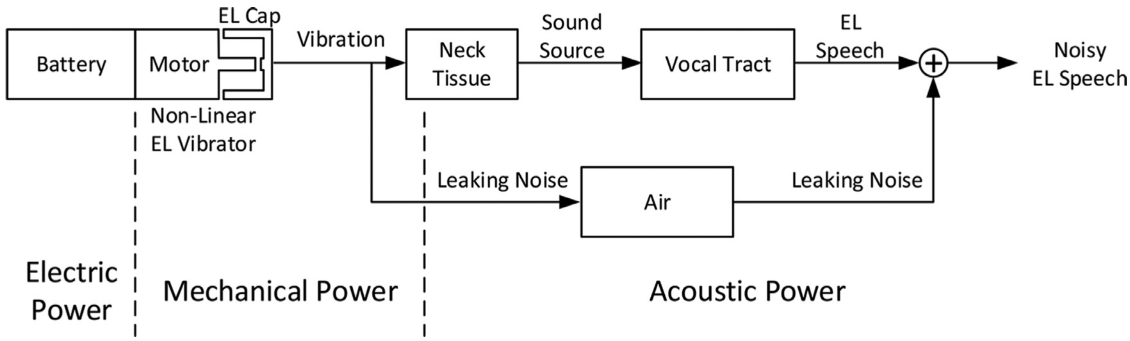

Fig. 1. Energy conversion in the electrolaryngeal speech production. Dash lines separate the process from the aspect of energy form.

2. Methods and experiments

2.1. Definitions of EL vocal efficiencies

Fig. 1 shows the entire process of electrolaryngeal speech production. Overall vocal efficiency of EL speech production (VE_{EL}) was defined as the ratio of the acoustic power of the EL speech to the electric power supplied by the battery. From the perspective of energy transfer, three successive stage efficiencies in the process were defined as follows.

(1) Electromechanical efficiency of the non-linear vibrator (EE_{NL}) was defined as the ratio of the mechanical power of the output vibration to the electric power of the battery.
(2) Coupling efficiency of the EL cap and the neck (CE_{CN}) was defined as the ratio of the mechanical power that transmits into the neck to the mechanical power of the output vibration of the EL cap.
(3) Transmission efficiency of the neck and the vocal tract (TE_{NV}) was defined as the ratio of the acoustic power of EL speech to the mechanical power that transmits into the neck.

It is important to note that the EL speech in these efficiencies referred to the radiated sound from the mouth rather than the noisy EL speech with the leaking noise, because only the radiated speech from the mouth is the meaningful part in the noisy EL speech and the effective output of the EL system.

2.2. Experimental procedures

Nine healthy subjects (five males and four females) and four laryngectomees (two males and two females) participated in the following experiments. The average ages of the healthy subjects and the laryngectomees were 26.1 ± 2.4 and 62.7 ± 6.8 years old, respectively. All the laryngectomees suffered a late stage of glottic carcinoma and had undergone total laryngectomy with bilateral neck dissection and radiation therapy for at least 17 years as shown in Table 1. All subjects were native Mandarin Chinese speakers familiar with using a commercial EL.

A wide-used commercial EL (Servox Digital, Servona, Germany) was selected in this study. The Servox Digital EL had two different sound caps for soft and hard tones. The two plastic caps have the same size (3 cm radius) but different hardness (Vikers hardness, 13.03 MPa for hard cap and 10.81 MPa for soft cap, measured with 10 g force and 10 s holding time). Therefore, both caps were tested to investigate the influence of the vibrator on the EL vocal efficiency. In accordance with actual use, different fundamental frequencies were chosen for male EL users (100 Hz) and female EL users (150 Hz). In addition, the volume was set to the maximum level to make sure that energy input of the motor was consistent in different fundamental frequencies. This assumption was confirmed by the measurement of the vibration amplitudes of the piston driven by the motor, which was 2.7 mm for 100 Hz and 2.6 mm for 150 Hz.

Two separate experiments were carried out in a soundproof room to collect the necessary data to calculate the vocal efficiencies of EL speech production.

2.2.1. Collection of mechanical vibration of the non-linear vibrator without the neck coupling

A laser Doppler vibrometer (LDV) system (Polytec, controller model OFV-5000 & laser head model OFV-534, Waldbronn, Germany) was used to measure the vibration velocity signal of the commercial EL cap. During the trial, the commercial EL and the LDV were fixed head-to-head with the laser point at the center of the EL cap. The EL was refitted to be driven by a DC regulated power supply (Zhongce Electronics, model DF1731SLL3A, Ningbo, China) which can display the real-time working current of the EL system.

In this experiment, three different fundamental frequencies (100 Hz, 150 Hz, and 200 Hz) were tested to see how the frequency affects the electromechanical efficiency of the non-linear vibrator. In each trial, the voltage and current of the working EL were recorded, and the velocity data of the EL cap was captured using a computer with a Gage PCI card (Model CompuScope 12502, Gage Inc., Lockport, USA) at a sampling frequency of 50 kHz. Each EL case was repeated ten times.

2.2.2. Collection of sound pressure level of the noisy EL speech and the leaking noise

In this experiment, each subject was asked to produce a sustained vowel for 5 seconds. A sound level meter (Model HT-8352, HCJYET, Guangzhou, China) mounted 20 cm in front of the mouth was used to record the sound pressure level (SPL in dB) of the

Table 1
Detailed information of the laryngectomized speakers.

Subjects	Gender	Age	Laryngectomy	Years after surgery	Radiation	Neck dissection
S1	Male	59	Total	22	Yes	Bilateral
S2	Male	68	Total	30	Yes	Bilateral
S3	Female	69	Total	28	Yes	Bilateral
S4	Female	55	Total	17	Yes	Bilateral