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.

© 2017 Elsevier B.V. All rights reserved.

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
0167-6393/© 2017 Elsevier B.V. All rights reserved.

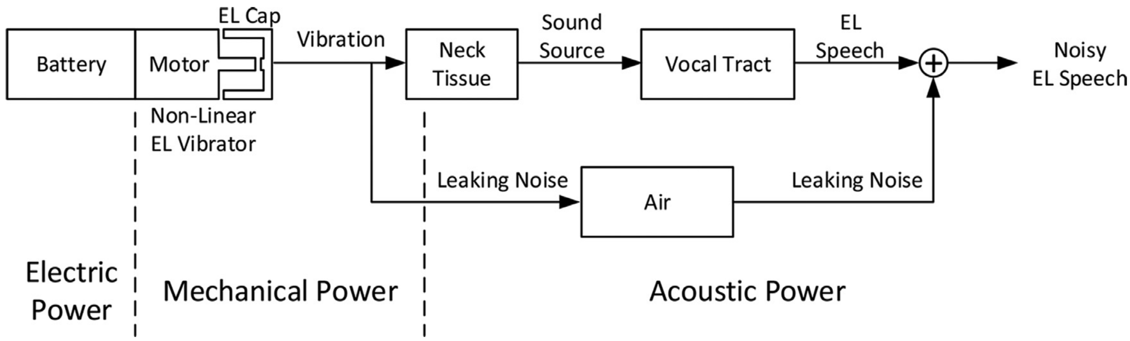

Fig. 1. Energy conversion in the electrolaryngeal speech production. Dash lines separate the process from the aspect of energy form.

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

Table 2
Electric power of the working EL.

EL cap	Frequency (Hz)	U (V)	I (A)	P_E (W)
Hard	100	7.20	0.30	2.160
	150		0.30	2.160
Soft	100		0.31	2.232
	150		0.31	2.232

noisy EL speech and the leaking noise, separately. Firstly, the subject was asked to produce a normal EL speech for the SPL measurement of the noisy EL speech. In accordance with the condition of the laryngectomized patient, the healthy speaker was requested to hold his glottis closed to eliminate inadvertent voicing and airflow from the lower respiratory tract. Then, the subject was asked to produce the same syllable with a closed mouth and nostril, the SPL value was considered the result of the leaking noise.

During the trial, the EL was placed against the neck at the common location for EL placement (Meltzner, 2003). The distance between the EL and the sound level meter was in the range of 29 cm–30 cm (averaging 29.8 cm) for all subjects. Furthermore, five Mandarin vowels (Pinyin "a," "o," "e," "i," "u" as /a/, /ɔ/, /ə/, /i/, /u/ in International Phonetic Alphabet) with a level tone were selected to study the effect of the vocal tract configuration on the EL vocal efficiency. Each vowel was repeated five times.

2.3. Calculation of EL vocal efficiencies

According to the measured values of the voltage U and the current I of the working EL, the vibration velocity $v(t)$ of the EL cap without the neck coupling, and the sound pressure level of the noisy EL speech $SPL_{\text{noisey\_EL\_speech}}$ and the leaking noise SPL_{noise}, the previously defined vocal efficiencies can be calculated step by step as follows.

Step one: *calculating the electric power of the working EL*

As a simple electric circuit, the electric power of the working EL, P_E, was the product of the voltage and current.

$$P_E = UI \tag{1}$$

The voltage remained stable and the current changed less than 0.01 A before and after the coupling with the neck. The electric power of the working EL is listed in Table 2.

Step two: *estimating the mechanical power with the neck coupling*

Because the commercial EL is a non-linear transducer (Houston et al., 1999; Wu et al., 2013), the vibration of the EL cap can be considered a damped oscillation under periodic impulse excitations. Accordingly, the mechanical power of the EL cap P_C was equal to the initial kinetic energy E_k of each cycle divided by the fundamental frequency F_0.

$$P_C = E_k/F_0 = \frac{1}{2}m_c v_0^2/F_0 \tag{2}$$

where v_0 is the initial velocity of the EL cap measured by the LDV, and m_c is the effective mass of the EL cap evaluated by a separate experiment (See Appendix). The effective mass of the hard cap and the soft cap were 0.641 g and 0.531 g, respectively.

In the case of neck coupling, the vibration of the combination system of the EL cap and the neck tissue was still a damped oscillation with an additive neck mass m_n. Based on the assumption that the excitation time of the collision between the motor and the EL cap was short enough to meet the law of momentum conservation, the initial velocity of the EL cap $v_{\text{c-coupling}}$ and the neck tissue $v_{\text{n-coupling}}$ can be estimated as

$$v_{c-\text{coupling}} = v_{n-\text{coupling}} = m_c v_0/(m_c + m_n) \tag{3}$$

Therefore, the mechanical power of the EL cap with the neck coupling $P_{\text{c-coupling}}$ and the mechanical power that transmits into the neck $P_{\text{n-coupling}}$ can be computed as

$$P_{c-\text{coupling}} = \frac{1}{2}m_c v_{c-\text{coupling}}^2/F_0; \quad P_{n-\text{coupling}} = \frac{1}{2}m_n v_{n-\text{coupling}}^2/F_0 \tag{4}$$

In this work, the neck mass m_n was set to 1.1 g and 1.9 g as a minimum and maximum value according to an estimated range of the load mass of the neck tissue reported by Houston et al. (1999).

Step three: *calculating the acoustic power of the EL speech*

In the SPL collection experiment, a basic assumption was that the noisy EL speech was the combination of the EL speech and the leaking noise. Accordingly, the sound intensity of the EL speech I_{EL} can be estimated as

$$I_{EL} = I_{\text{noisy\_EL\_speech}} - I_{\text{noise}} = I_0 10^{SPL_{\text{noisy\_EL\_speech}}/10} - I_0 10^{SPL_{\text{noise}}/10} \tag{5}$$

where $SPL_{\text{noisy\_EL\_speech}}$ and SPL_{noise} are the sound pressure levels of the noisy EL speech and the leaking noise measured in dB by the sound level meter. I_0 is the reference sound intensity of 10^{-12} W m^{-2}. Then, the acoustic power of the EL speech P_{EL} was determined as

$$P_{EL} = 4\pi r_{EL}^2 I_{EL} \tag{6}$$

where r_{EL} is the distance between the mouth and the sound level meter.

Step four: *calculating the vocal efficiencies*

According to the definitions mentioned above, the overall vocal efficiency of EL speech production (VE_{EL}), the electromechanical efficiency of the non-linear vibrator (EE_{NL}), the coupling efficiency of the EL cap and the neck (CE_{CN}), and the transmission efficiency of the neck and the vocal tract (TE_{NV}) can be computed as

$$VE_{EL} = P_{EL}/P_E \tag{7}$$

$$EE_{NL} = (P_{c-\text{coupling}} + P_{n-\text{coupling}})/P_E \tag{8}$$

$$CE_{CN} = P_{n-\text{coupling}}/(P_{c-\text{coupling}} + P_{n-\text{coupling}}) \tag{9}$$

$$TE_{NV} = P_{EL}/P_{n-\text{coupling}} \tag{10}$$

2.4. Data analysis

For each subject, 50 sets (2 caps × 5 vowels × 5 repeats) of the EL vocal efficiencies (including VE_{EL}, EE_{NL}, CE_{CN}, and TE_{NV}) were finally obtained for statistic analysis. Because both the within-subjects factor (EL cap and vowel) and the between-subjects factor (subject group) were tested, a mixed ANOVA was selected to investigate their effects on the vocal efficiencies. For the EE_{NL}, a regular ANOVA was used to compare the difference due to the EL caps and fundamental frequencies.

3. Results

3.1. Overall vocal efficiency of EL speech production

The mean values of the overall vocal efficiencies of EL speech production in different subject groups are shown in Fig 2. The average VE_{EL} of the healthy subjects was 1.77×10^{-5} using the hard cap and 2.04×10^{-5} using the soft cap, while the average VE_{EL} of the laryngectomees was 0.82×10^{-5} using the hard cap and 1.43×10^{-5} using the soft cap. A one-way mixed ANOVA showed no significant effects of the EL cap ($F = 3.455$, $p = 0.093 > 0.05$), the subject group ($F = 2.485$, $p = 0.146 > 0.05$), and their interaction effect ($F = 0.72$, $p = 0.416 > 0.05$) on the overall vocal efficiency.

Fig. 3 shows the average vocal efficiencies of different vowels in different cases. The results of a two-way mixed ANOVA showed significant differences of the VE_{EL} across vowels ($F = 12.863$,

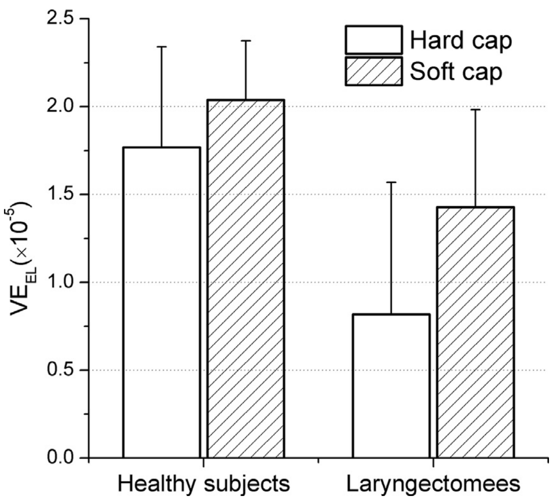

Fig. 2. Overall vocal efficiencies of the EL speech production (VE_{EL}) in different subject groups. Each bar represents the average VE_{EL} of all the group members, and error bars represent the corresponding standard deviations.

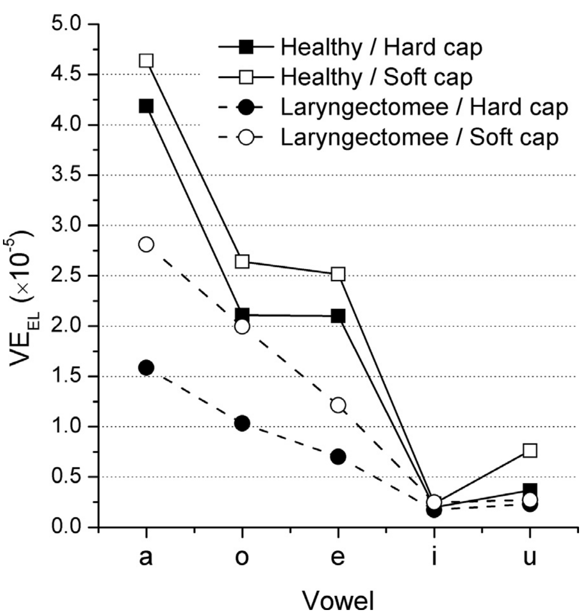

Fig. 3. Overall vocal efficiencies of the EL speech production (VE_{EL}) for different vowels. Each bar represents the average VE_{EL} of all the group members, and error bars represent the corresponding standard deviations.

$p = 0.005 < 0.05$), but no significant interaction effects of the vowel and any other factor on the VE_{EL}. Furthermore, pairwise comparisons of different vowels indicated that the VE_{EL} was the highest when producing the vowel “a” and the lowest when producing the vowel “i” and “u” ($p < 0.01$). This difference should be related to the vocal tract configuration for different vowels. Based on the vocal tract area function from magnetic resonance imaging (Story et al., 1996), the vowel “a” and “o” have a similar constricted pharynx, widened oral cavity, and large mouth opening. On the contrary, the vowel “i” and “u” have a higher area in pharyngeal region, narrow oral cavity and small mouth opening. The vowel “e” has a neural shape between the two opposite configurations.

3.2. Electromechanical efficiency and coupling efficiency

Fig. 4 shows the average electromechanical efficiencies of the non-linear vibrator with and without the neck coupling. Statistical results indicated that the EE_{NL} with the neck coupling

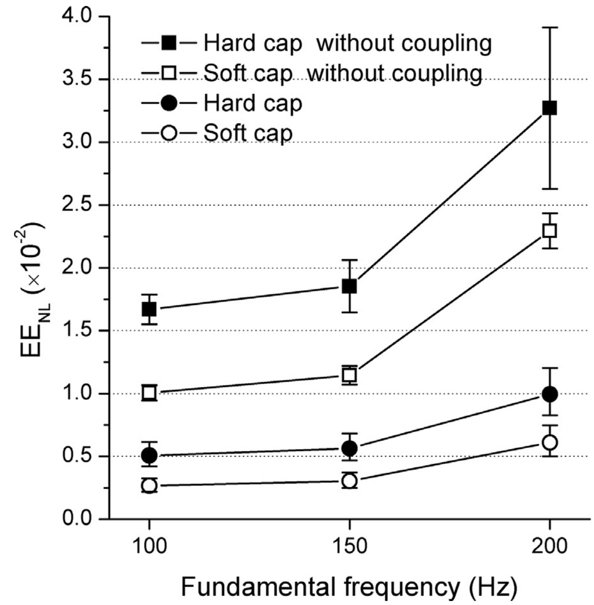

Fig. 4. Average electromechanical efficiencies of the non-linear vibrator (EE_{NL}) with and without the neck coupling. In the case without the neck coupling, each square and error bars represent the average EE_{NL} and the standard deviations of 10 repeated tests. In the neck coupling case, each circle represents the average EE_{NL} of different neck masses, and error bars show the maximum and minimum EE_{NL}.

was significantly lower than that without the neck coupling (*T*-test, $p < 0.001$). In the case of neck coupling, a two−way ANOVA showed significant main effects of EL cap ($F = 169.054$, $p < 0.001$) and fundamental frequency ($F = 134.373$, $p < 0.001$) on the EE_{NL}. For different EL caps, the EE_{NL} of the hard cap was significantly higher than that of the soft cap. For different fundamental frequencies, Tukey post-hoc test indicated that the EE_{NL} of 200 Hz was higher than those of 100 Hz and 150 Hz ($p < 0.001$), but no significant difference between those of 100 Hz and 150 Hz ($p = 0.219 > 0.05$).

Fig. 5 shows the coupling efficiencies of the EL cap and the neck (CE_{CN}) for different neck masses. Under the assumption of momentum conservation, the CE_{CN} was positively related to the load mass of the neck tissue. The results showed that the CE_{CN} was higher

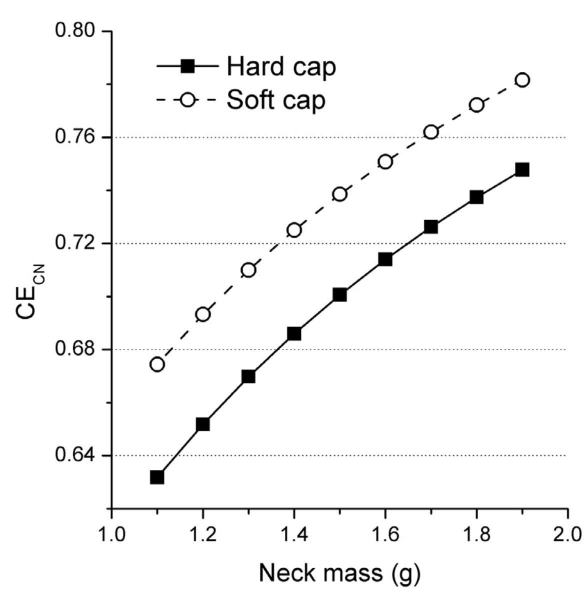

Fig. 5. Coupling efficiencies of the EL cap and the neck (CECN) for different load masses of the neck tissue.

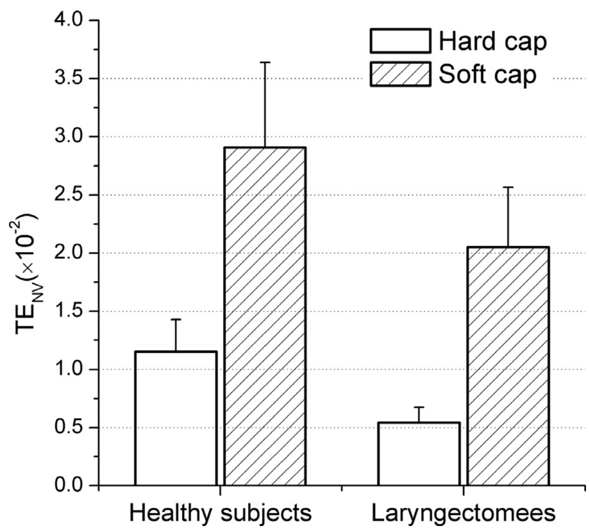

Fig. 6. Average transmission efficiencies of the neck and the vocal tract (TE_{NV}) in different subject groups. Each bar represents the average TE_{NV} of all the group members, and error bars represent the corresponding standard deviations.

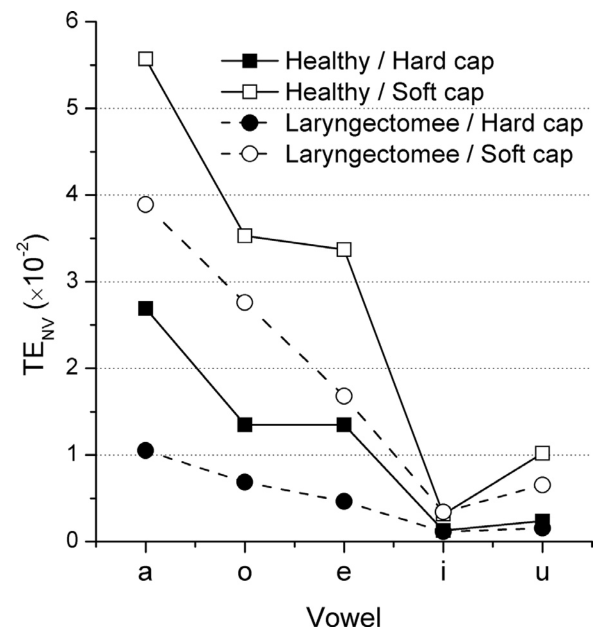

Fig. 7. Average transmission efficiencies of the neck and the vocal tract (TE_{NV}) for different vowels. Each bar represents the average TE_{NV} of all the group members, and error bars represent the corresponding standard deviations.

when the coupling neck mass was larger. For the same neck condition, the CE_{CN} of the soft cap was higher than that of the hard cap by average 3.81×10^{-2}.

3.3. Transmission efficiency

The average transmission efficiencies of the neck and the vocal tract are shown in Fig 6 in different subject groups. The average TE_{NV} of the healthy subjects was 1.15×10^{-2} using the hard cap and 2.91×10^{-2} using the soft cap, while the average TE_{NV} of the laryngectomees was 0.54×10^{-2} using the hard cap and 2.05×10^{-2} using the soft cap. A one-way mixed ANOVA indicated a significant main effect of the EL cap ($F = 5.715$, $p = 0.038 < 0.05$) on the transmission efficiency, but no difference of the TE_{NV} between the healthy subjects and the laryngectomees ($F = 4.244$, $p = 0.066 > 0.05$).

For different vowels, Fig. 7 shows the average transmission efficiencies of the neck and the vocal tract in different cases. Consistent with the results of the overall vocal efficiency, the TE_{NV} was significantly different across vowels (a two-way mixed ANOVA, $F = 9.564$, $p = 0.011 < 0.05$). The TE_{NV} of the vowel "a" was the highest while the TE_{NV} of the vowel "i" and "u" were the lowest (pairwise comparison, $p < 0.01$). However, unlike with the VE_{EL}, the within-subjects factor of the EL cap ($F = 9.246$, $p = 0.012 < 0.05$) also played main effect on the transmission efficiency. Especially, the TE_{NV} of the soft cap was higher than that of the hard cap when producing the same vowel.

4. Discussion

4.1. Efficiencies and power losses in the EL speech production

Considering the EL device and human body as an integrated system, the overall EL efficiency reflects the conversion ratio of the input energy (electric power provided by the battery) to the acoustic power of the EL speech directly radiated from the mouth. For the commercial EL of Servox Digital in this work, the electric power was approximately 2.2 W and the acoustic power was estimated in the range of 10^{-7}–10^{-3} W with the maximum volume setting. Therefore, the vocal efficiency of EL speech production using this commercial EL ranged from zero to as high as 0.1%. This result reveals a very low utilization of the battery energy for producing the useful acoustic energy. To find where the rest of the electric power goes, the process of EL speech production can be divided into three successive stages of energy transfer as shown in Fig 1.

Since the EL is an electromechanical transducer providing a mechanical sound as a substitution source for voice rehabilitation, the first stage is the power transfer from the electricity to the mechanical vibration. The available output of this stage is the mechanical vibration of the EL cap. In this work, the mechanical power of the available output was $2{\sim}8 \times 10^{-2}$ W without the neck coupling and $4{\sim}20 \times 10^{-3}$ W with the neck coupling, respectively. Hence, the efficiency of this stage, named as electromechanical efficiency of the non-linear vibrator (EE_{NL}), was on the order of 10^{-3}–10^{-2} approximately. In this stage, at least 90% of the battery power is consumed mainly in two ways. Firstly, being an electric motor, the EL device has motor losses in driving the piston, including the resistive loss, core loss, mechanical loss and aerodynamic loss. Secondly, because of physical separation between the piston and the EL cap, only a small part of the kinetic energy of the piston can be transferred to the EL cap or the combination of the EL cap and the neck tissue in a partial inelastic collision, and this energy loss is related to the coupling pressure (Meltzner et al., 2005). Therefore, most of the battery energy is lost in the vibrating motor and the inelastic collision.

The second stage is the power transfer from the EL cap to the neck tissue without changes of energy form in the neck coupling case. Because of the difficulty in directly measuring the vibration of the EL cap and the neck tissue during the EL speech production, the mechanical powers of the EL cap and the neck tissue here were estimated based on the assumption that the EL cap and the neck tissue were a rigid combination at the time of the EL motor-cap collision. Thus, the estimated coupling efficiency of the EL cap and the neck tissue (CE_{CN}) in this work was mainly related with their masses and ranged from 0.6 to 0.8 as shown in Fig 5. A higher neck mass will gain a higher CE_{CN} due to a more energy distribution in this combination object. However, in reality, the neck tissue is elastic and not closely coupled with the EL cap, resulting in elastic losses and leaking losses. The actual value of the CE_{CN} should be lower than the estimates here.

In the last stage, the vibration power is transferred through the neck tissue and the vocal tract and finally transformed into the radiated acoustic power from the mouth. The efficiency of this stage was the transmission efficiency of the neck and the vocal tract (TE_{NV}), which was on the order of 10^{-3}–10^{-1}. For the neck transfer, the low-pass characteristic of the neck tissue attenuates about 10∼30 dB energy in the frequency band lower than 1000 Hz (Wu et al., 2014), leading to an energy transmission efficiency of 10^{-3}–10^{-1}. Then, in the process of vocal tract transfer, viscous losses and wall vibration losses contribute to major energy consumption as acoustic waves propagate along the vocal tract (Titze, 1992). However, these losses are likely to be small in comparison to the neck losses. In addition, the TE_{NV} gradually decreases when the configuration of the vocal tract changes in the order of "a," "o," "e," "i," "u". In this order, the area of pharyngeal region is getting bigger and the oral cavity and mouth-opening area are getting smaller (Story et al., 1996), which results in more energy losses in the vocal tract transmission and less energy radiation to the air.

In conclusion, the three stage efficiencies indicate commonly existing energy losses in the whole process of the EL speech production, and reflect the performance of EL device and physiological features on the EL speech production. By contrast, the non-linear design of the EL transducer and the neck attenuation are the first two reasons of power losses and low perceptual intelligibility of EL speech (Meltzner et al., 2005).

4.2. Variations of EL vocal efficiencies

As mentioned above, the EL vocal efficiency varied in a very wide range from zero to 10^{-3}. This result indicates that the EL vocal efficiency should be affected by some conditions. In this work, two factors (EL cap and vowel) were investigated to reveal their potential influences on these efficiencies.

Two different EL caps were studied in this work, and the results showed significant difference in every stage efficiency, but not in the VE_{EL}. This might be due to the different effects of the EL cap on different stage efficiencies. The CE_{CN} and TE_{NV} of the soft cap were higher than those of the hard cap by about a factor of 10^{-2} and 10^{-3} respectively, whereas the EE_{NL} of the soft cap were lower than that of the hard cap by about a factor of 10^{-2}. Although there was no statistical significant difference between the EE_{NL} of 100 Hz and 150 Hz, the EE_{NL} became higher with the increase of the fundamental frequency. This can be explained by the fact that high fundamental frequencies increase the number of collisions per unit time, thus transferring more energy from the motor to the EL cap. Thus, the offset of the opposite effects leads to no significant impact on the overall efficiency. Furthermore, the effect of the EL cap mostly reflects on the EE_{NL} and CE_{CN}, which may be related to the hardness of the EL cap. The hard cap gains higher energy transfer than soft cap in the inelastic collision between the motor and the EL cap, but delivers less energy from the cap to the neck tissue due to the poor coupling of the EL cap and the neck. Therefore, the EL cap should be chosen according to the stiffness of the neck tissue.

For different vowels, the results showed significant differences in the VE_{EL} as well as in the TE_{NV}. This outcome indicates that the vowel influences the VE_{EL} by its effect on the TE_{NV}. TE_{NV} refers to the combination of the transmission efficiencies of the neck and the vocal tract. For the neck transmission, Wu et al. (2014) reported that the neck transmission characteristic was not sensitive to the vowel. So, the vowel affects the TE_{NV} mainly by changing the transmission efficiency of the vocal tract. For vowel "a", it is an open vowel with a largest mouth area for acoustic radiation. Conversely, both the vowel "i" and "u" are closed vowels with a small mouth opening area for acoustic radiation. Taking the vowel "i" as the reference, the acoustic intensity of the vowels "a", "o", "e", and "u" are higher than that of vowel "i" by average 4.78 times, 0.26 times, 0.19 times, and 1.45×10^{-9} times, respectively. Therefore, the wider mouth opens, the more acoustic power radiates.

Although both laryngectomees and healthy subjects were measured in this work, the significant difference of the vocal efficiency between these two groups is not clear because of few samples and different ages. However, the variability of physiological structures in the laryngectomized subject (such as the surgical removal of partial vocal tract, radiation-induced changes of neck stiffness, etc.) could be possible factors affecting the coupling efficiency CE_{CN} and transmission efficiency TE_{NV}. In this work, the slight lower vocal efficiency of the laryngectomees than that of the healthy subjects may be due to these structural differences.

This work was done in a fixed state of experiment (vowel production). However, a daily running EL speech consists of the vowels and consonants, and average vocal efficiency per unit time should be dependent on proportion of different syllables in the running EL speech. Since the selected vowels in this work covers a wide range of vocal tract configurations, the vocal efficiency of a running EL speech may be variable in a similar range as the vowels (0∼0.1%) according to different syllable combinations. Nevertheless, the actual vocal efficiency of the running EL speech may be lower than this estimation value because of more energy losses during non-speech interval between words and sentences.

4.3. Suggestions for improvement of EL technology

Although this work was done with just a Servox device, the definitions, measurements and findings of the vocal efficiencies may also be extended to other commercial devices (such as Trutone, Nu-Vois, etc.) because most of the current EL devices operate with the similar mechanical structure and the same process of EL speech production.

As a phonation machine, the EL device and human body cannot be considered an efficient system, so eliminating this inefficiency will transfer more energy into the meaningful acoustic energy and improve the EL speech quality.

For the EL device, a linear vibrator may offer benefits in the EL speech production compared to the current non-linear vibrator. The linear vibrator may increase the electromechanical efficiency by saving energy from the collision between the piston and the separated vibrator (EL cap), thus transmitting more energy into the vocal tract for phonation and increasing speech intensity. Some researchers have tried to design linear vibrators for EL speech production (Houston et al., 1999; Sugio et al., 2007), but there is no further report of perception evaluation of the EL speech produced using a linear vibrator.

In addition, the results show the significant impact of EL cap on the coupling efficiency, which determines the effective transfer of energy into the neck. Although there is no further research on the relationship between the EL cap and the EL speech quality, the material property of the EL cap should be compatible with the biomechanical features of the neck tissue. Also, the increase of coupling efficiency will reduce the leaking noise and improve the signal-noise ratio of the EL speech. Although the deficits in signal transmission is a longstanding problem in the EL speech (Meltzner et al., 2005), this work quantitatively evaluates the energy efficiency and losses during the EL speech production. The suggestions for improving EL technology are based on the findings in this work, but it is not an easy work and needs further studies.

5. Conclusions

As an energy conversion process, the vocal efficiency of EL speech production can be defined as the ratio of the acoustic power of the radiated EL speech from the mouth to the electric power supplied by the battery. The commonly used Servox Digital

EL converted less than 0.1% of the battery energy into the meaningful speech, indicating high power losses in the process. On one hand, the non-linear design of the EL device contributed a high power loss in the collision, thus only $10^{-3} \sim 10^{-2}$ of the battery energy was transferred to the EL cap. On the other hand, the physiological features of the neck tissue, such as the stiffness and sound transmission characteristics, also contributed to high power losses because of the energy reflection and energy attenuation. The factor of the EL cap affected the energy conversion efficiency through this way. Therefore, EL linear vibrator and compatible cap may be potential ways to raise the vocal efficiency, finally improving the EL speech quality by increasing the speech intensity and reducing the leaking noise.

Acknowledgments

The authors would like to express special appreciations to Yubo Guan for his help in the experiments and the evaluation of the EL cap mass. This work was supported by National Natural Science Foundation of China (No. 11404256, 11274250, and 61271087), and Natural Science Basic Research Plan in Shaanxi Province of China (No. 2016JQ2017), and Fundamental Research Funds for the Central Universities (No. xjj2014057).

Appendix. Effective mass evaluation of the EL cap

For a damping oscillation system, the natural period T_n is related to the mass of the system M. In the case of EL vibration, if a solid object, mass m, is attached to the EL cap, the natural period T_n is determined as

$$T_n = 2\pi\sqrt{\frac{m+m_c}{K}} \quad (A1)$$

where m_c is the effective mass of the EL cap, and K is the stiffness factor of the EL cap. After a simple transformation, the Eq. (A1) becomes

$$T_n^2 = \frac{4\pi^2}{K}m + \frac{4\pi^2}{K}m_c = am + b \quad (A2)$$

Apparently, T_n^2 is a linear function of the attached mass m. The effective mass m_c is the ratio of the intercept b to the slope a.

In this experiment, six nylon cubes were selected as the added object. Each cube was attached in the surface of each EL cap, and the corresponding vibration displacement of the system was measured by the LDV to evaluate the natural period T_n. Each cube was repeated ten times.

As shown in Fig. A1, the amplitude gradually decreases according to an exponential decay, $Ae^{-\zeta\omega_n t}$, where A is the initial amplitude, ζ is the damping ratio ($0<\zeta<1$), and ω_n is the natural

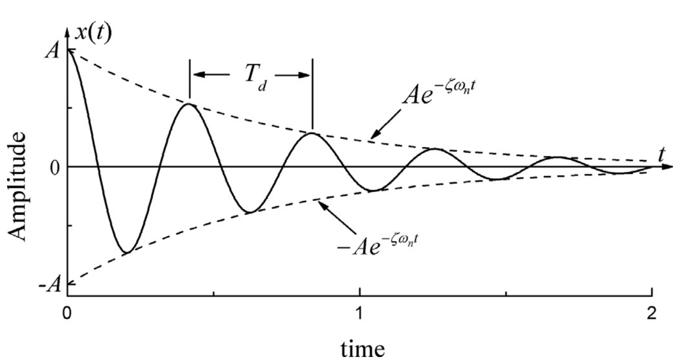

Fig. A1. Vibration displacement of a damped oscillation.

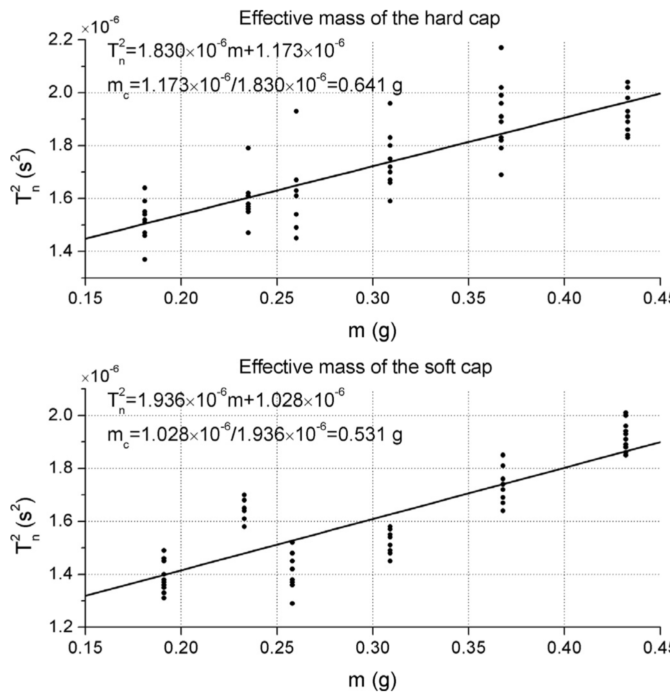

Fig. A2. Measurement of the effective mass of the hard cap (Top) and the soft cap (Bottom). Each dot represents the square of the natural period T_n^2 with each added mass m. The straight lines are linear fits to all the data points. Top: The effective mass of the hard cap is 0.641 g. Bottom: The effective mass of the soft cap is 0.531 g.

angular frequency of the system. In addition, T_d is the damped period of the system, and the relationship between T_d and ω_n is

$$\omega_n = \frac{\omega_d}{\sqrt{1-\zeta^2}} = \frac{2\pi}{T_d\sqrt{1-\zeta^2}} \quad (A3)$$

where ω_d is the damped angular frequency of the system.

First, the damping factor $n=\zeta\omega_n$ and the damped period T_d can be estimated through the exponential curve fitting of the peaks of the vibration displacement. Then, the natural angular frequency ω_n can be obtained as

$$\omega_n = \frac{\sqrt{4\pi^2 + (nT_d)^2}}{T_d} \quad (A4)$$

Finally, T_n^2 is computed as

$$T_n^2 = \frac{4\pi^2}{\omega_n^2} = \frac{4\pi^2 T_d^2}{4\pi^2 + (nT_d)^2} \quad (A5)$$

Fig. A2 shows the measured distribution of $T_n^2 \sim m$ and corresponding fitting line. The m_c of the hard cap was 0.641 g, and the m_c of the soft cap was 0.531 g. The effective mass of the hard cap was little higher than that of the soft cap.

References

Grillo, E.U., Verdolini, K., 2008. Evidence for distinguishing pressed, normal, resonant, and breathy voice qualities by laryngeal resistance and vocal efficiency in vocally trained subjects. J. Voice 22, 546–552.

Houston, K.M., Hillman, R.E., Kobler, J.B., Meltzner, G.S., 1999. Development of sound source components for a new electrolarynx speech prosthesis. In: Acoustics, Speech, and Signal Processing, IEEE International Conference on, vol. 4, pp. 2347–2350.

Howard, D.M., Lindsey, G.A., Allen, B., 1990. Toward the quantification of vocal efficiency. J. Voice 4, 205–212.

Jiang, J., Stern, J., Chen, H., Solomon, N.P., 2004. Vocal efficiency measurements in subjects with vocal polyps and nodules: a preliminary report. Ann. Otol. Rhinol. Laryngol. 113, 277–282.

Meltzner, G.S., Hillman, R.E., Heaton, J.T., Houston, K.M., Kobler, J.B., Qi, Y., 2005. Electrolaryngeal speech: the state of the art and future directions for development. Contemporary Considerations in the Treatment and Rehabilitation of Head and Neck Cancer: Voice, Speech, and Swallowing 571–590.

Meltzner, G.S., 2003. Perceptual and Acoustic Impacts of Aberrant Properties of Electolaryngeal Speech Ph.D. thesis. Harvard-MIT Division of Health Sciences and Technology, Harvard University, Cambridge, Massachusetts.
Story, B.H., Titze, I.R., Hoffman, E.A., 1996. Vocal tract area functions from magnetic resonance imaging. J. Acoust. Soc. Am. 100 (1), 537–554.
Sugio, Y., Kanetake, R., Tanaka, A., Ooe, K., 2007. Work of PZT ceramics sounder for sound source artificial larynx. In: Proceedings of SPIE - The International Society for Optical Engineering, pp. 237–242.
Titze, I.R., 1992. Vocal efficiency. J. Voice 6, 135–138.
Wu, L., Wan, C., Wang, S., Wan, M., 2013. Improvement of electrolaryngeal speech quality using a supraglottal voice source with compensation of vocal tract characteristics. IEEE Trans. Biomed. Eng 60, 1965–1974.
Wu, L., Xiao, K., Dong, J., Wang, S., Wan, M., 2014. Measurement of the sound transmission characteristics of normal neck tissue using a reflectionless uniform tube. J. Acoust. Soc. Am. 136 (1), 350–356.