Measurement of the sound transmission characteristics of normal neck tissue using a reflectionless uniform tube

Liang Wu, Ke Xiao, Jiaqi Dong, Supin Wang, and Mingxi Wan[a]
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

(Received 26 December 2013; revised 6 May 2014; accepted 19 May 2014)

Understanding the sound transmission of the neck tissue is necessary and important in areas such as vocal function evaluation and electrolarynx improvement. In this paper, a simple method using a reflectionless tube was proposed to measure the neck frequency response function (NFRF) of ten normal subjects (five males and five females) during Mandarin vowel production. The NFRFs across different subjects producing different vowels were measured at different neck positions and compared to confirm the effectiveness of the method, and determine the NFRF variations in normal subjects. The results showed that the proposed method offered an easy and effective way to obtain an accurate NFRF. For normal subjects, the neck tissue can be treated as a low-pass filter, with a maximum gain at 310 Hz and a roll-off at a slope of −8.4 dB/octave, flattening out above 2000 Hz. The measurement position on the neck did not influence the shape of the NFRF, but did change the overall gains of the NFRF. In addition, there was a significant gender difference in NFRFs at the low frequencies. Finally, some potential applications of this method and the results are suggested.

[http://dx.doi.org/10.1121/1.4883355]

PACS number(s): 43.70.Jt, 43.70.Aj [ZZ] Pages: 350–356

I. INTRODUCTION

During normal speech production, the neck tissue transmits vocal fold vibration from the vocal tract to the skin surface. The surface vibration can be measured by accelerometer sensors or contact microphones for long-term monitoring of vocal function[1–6] or speech detection and recognition.[7–12] During electrolarynx (EL) speech production, the mechanical sound sources must be transmitted through the neck tissue to excite the vocal tract acoustically. Thus, the sound transmission characteristic of the neck tissue is an important factor that needs to be well studied.

Several researchers have tried to measure the neck frequency response function (NFRF), and reported that the neck tissue acts as a low-pass filter.[9,13,14] Norton and Bernstein computed the NFRF as the spectrum of the output signal from the mouth (with the formants removed) divided by the spectrum of the input signal of a shaker vibrating against the neck.[13] As an extension of the Norton and Bernstein study, Meltzner *et al*. measured the NFRF as the ratio of the spectrum of the estimated volume velocity that excites the vocal tract to the spectrum of the acceleration measured at the neck, using a complicated algorithm to remove the airborne noise, near-field lip radiation characteristic, and vocal tract transfer function.[14] The Meltzner approach could measure a more accurate NFRF than Norton and Bernstein's, and the results showed that the neck tissue transmitted low-frequency sound energy better than high-frequency sound energy.

However, the Meltzner approach still has some weaknesses and limitations.[14] First, the sound directly radiated from the shaker significantly corrupts the pressure signal measured at the lips for certain frequencies. Second, the estimation of vocal tract transfer function is still a difficult task due to the presence of the neck impulse response and the back cavity. Third, the estimation of the near-field lip radiation characteristic will be influenced by the experimental environment. As a whole, the Meltzner approach is a little complicated because of the additional estimations of the airborne sound and lip radiation characteristic, and the accuracy of the estimated NFRF is limited by these estimations, especially the radiated sound and vocal tract transfer function. In addition, one purpose of Meltzner's work is to improve the EL driving signal by compensating for the individual's NFRF. Meltzner measured the NFRFs of ten laryngectomized patients, but only four normal subjects (two males and two females). So, any differences between normal males and females might not be accurately identified with the small sample size.

Therefore, we propose an easy and accurate way to measure the NFRF using a reflectionless uniform tube. A reflectionless uniform tube described by Sondhi is a promising technique for recovering glottal waveforms.[15] The reflectionless tube acting as a pseudo-infinite termination of the vocal tract can significantly cancel out the vocal tract contribution to approximate the glottal volume velocity directly.[16] The advantages of this method are that it is physically simple and highly resistant to noise from the surrounding environment.[17] Hence, similar to the Meltzner approach, we used the reflectionless uniform tube, instead of the inverse filtering method, to estimate the volume velocity that excites the vocal tract, thus avoiding the influence of radiated sound from the shaker and the estimation error of the vocal tract transfer function.

In this work, we measured the NFRF of ten normal subjects during Chinese vowel production. The results for different vowels, measured at different positions on the neck, and different subgroups (male and female), were compared and analyzed

---

[a] Author to whom correspondence should be addressed. Electronic mail: mxwan@mail.xjtu.edu.cn

0001-4966/2014/136(1)/350/7/$30.00

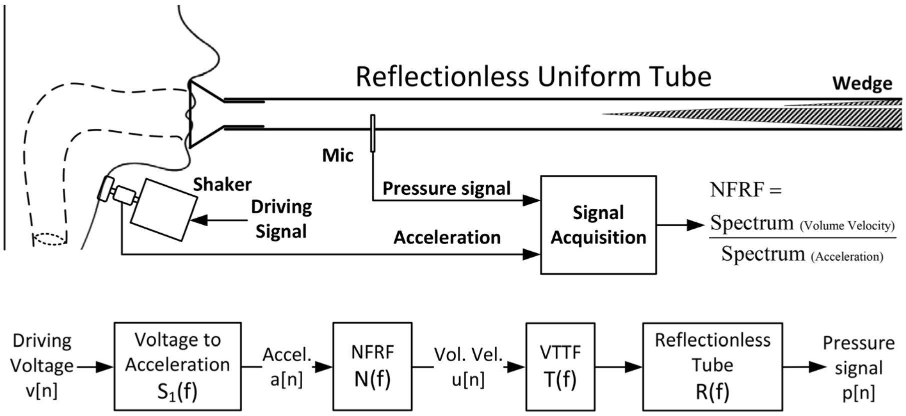

FIG. 1. (Top) The schematic diagram of the experimental procedures. (Bottom) The signal flow chart in the experimental process.

to verify the effectiveness of the method and investigate NFRF variations. Finally, some potential applications of this method and the NFRF results are discussed briefly.

II. METHODS

As in the Meltzner study, the NFRF measured in this work was also defined as the ratio of the spectrum of the estimated volume velocity that excites the vocal tract to the spectrum of the acceleration measured at the neck. But, the experimental procedures and signal flow were much simpler and easier, as shown in Fig. 1.

A. Experimental procedures

In the experiment, the subject was seated in front of a reflectionless uniform tube, placing his mouth upon a plastic mouthpiece connected with the tube. The reflectionless tube was made of steel, 190 cm long with an inner diameter of 2.5 cm. The mouthpiece was tightly coupled to one end of the tube and detachable for easy cleaning. At the other end, a 90 cm long conical wedge of polyurethane foam was inserted to terminate the tube and minimize acoustic reflection. An electret microphone (Knowles, Model WP-25993-D63, Itasca, IL) was fitted in a small hole, 30 cm from the mouthpiece end, to measure the pressure signal in the tube. The microphone had a wide response sensitivity of −55 dB (relative to 1 V/0.1 Pa) from 100 Hz to 10 000 Hz.

A Brüel & Kjær shaker (model 4810, Skodsborgvej, Denmark), driven by broadband random noise, was placed against the neck to provide vibration. The shaker was attached with an impedance head (KISTLER 8770A5, Winterthur, Switzerland) to measure the acceleration and a 3 cm diameter metal cap to provide enough contact area with the neck.

Ten normal speakers (five males and five females) participated in this study. The subjects, aged from 23 to 29 years (averaging 25.8 ± 2.6 yr for the males and 25.6 ± 2.1 yr for the females), had no reported history of speech language problems. All subjects were native Mandarin Chinese speakers. Five Mandarin vowels (Pinyin "a," "o," "e," "i," "u") were selected to study the NFRF variation across vowels. In each trial, the subject was instructed to configure the vocal tract for one vowel for 2–4 s, without phonating and while leaving the glottis open. Each vowel was repeated six times.

In addition, three different shaker locations were tested to evaluate the NFRF variations in neck positions. Position 1 was on the thyroid cartilage over the vocal folds, and position 2 was 2 cm lateral to position 1. Position 3 was about 2 cm superior to position 2, which is the common location for EL placement among laryngectomy patients.[18]

All recordings were performed in the speech laboratory of Xi'an Jiaotong University. A data acquisition system (BioPac MP 150, Goleta, CA) was used to collect the pressure signal from the microphone and the acceleration from the impedance head simultaneously. All of the signals were digitized at a 44 100 Hz sampling rate with a 16-bit quantization.

B. Data analysis

As shown in Fig. 1, assuming that the neck tissue is a linear time-invariant (LTI) system, the NFRF, *N(f)*, can be computed as

$$N(f) = \frac{\Phi_{au}(f)}{\Phi_{aa}(f)} = \frac{\mathcal{F}[\phi_{au}[n]]}{\mathcal{F}[\phi_{aa}[n]]}, \tag{1}$$

where $\Phi_{au}(f)$ is the cross-spectral density of the volume velocity, $u[n]$, and the acceleration, $a[n]$, and $\Phi_{aa}(f)$ is the power spectral density of $a[n]$, which can be obtained by the Fourier transform of the cross-correlation function, $\phi_{au}[n]$, and autocorrelation function, $\phi_{aa}[n]$, respectively.

In this method, the acceleration, $a[n]$, was measured directly through the impedance head. Due to the effect of the reflectionless tube on reducing the vocal tract resonance, the volume velocity, $u[n]$, can be approximated as the pressure signal, $p[n]$, measured by the microphone within the tube. As a result, the neck frequency transfer function can be directly computed as

$$N(f) = \frac{\Phi_{au}(f)}{\Phi_{aa}(f)} = \frac{\Phi_{ap}(f)}{\Phi_{aa}(f)} = \frac{\mathcal{F}[\phi_{ap}[n]]}{\mathcal{F}[\phi_{aa}[n]]}, \tag{2}$$

where the cross-correlation, $\phi_{ap}[n]$, and autocorrelation, $\phi_{aa}[n]$, were determined as

$$\phi_{ap}[n] = \sum_{m=-\infty}^{+\infty} a[m+n]p[m],$$

$$\phi_{aa}[n] = \sum_{m=-\infty}^{+\infty} a[m+n]a[m]. \tag{3}$$

In order to satisfy the assumption of the LTI system, the coherence function, $C_{au}(f)$, between the acceleration, $a[n]$, and the volume velocity, $u[n]$, was computed for data selection. Only data with >90% of the coherence functions >0.8 in the frequency range of 100–5000 Hz were selected for the NFRF analysis,

$$C_{au}(f) = \frac{|\Phi_{au}(f)|^2}{\Phi_{aa}(f) \cdot \Phi_{uu}(f)}. \tag{4}$$

First, the selected waveforms were filtered by a six-order Butterworth high-pass filter with a cutoff of 50 Hz to eliminate some tube resonances below 50 Hz.[17] Then, the waveforms were low-pass filtered at 6 kHz and down-sampled to 14 kHz. Finally, the NFRF was computed according to Eqs. (2) and (3), using a 512-point Gaussian window in the Fourier transform.

There were 150 NFRFs (10 subjects × 3 positions × 5 vowels) measured, and each NFRF was the average result of six trials. In order to verify the feasibility of the method, the NFRFs of different vowels were compared and analyzed. The NFRF for each position was determined as the average NFRF of five vowels. Furthermore, the average NFRFs of different subject groups (male and female groups) were also computed for the comparison of gender differences.

C. Performance of the reflectionless tube

To verify the reflectionless performance of the tube, an experiment was set up to measure its frequency response. The tube was connected directly to the B&K shaker, which was driven by broadband random noise. Then, the shaker vibration acceleration and the sound pressure in the tube were collected simultaneously. At last, the frequency response was computed as the ratio of the spectrum of the pressure signal to the spectrum of the vibration signal. As shown in Fig. 2, the magnitude frequency response curve is essentially flat (±1.27 dB) across a wide range from 100 to 5000 Hz. Therefore, the tube used in this work can be considered anechoic, and the NFRF will not be affected by the tube resonance.

Additionally, to verify the performance of the reflectionless tube in minimizing the effects of vocal tract resonances, the reflectionless tube was tested and adjusted in a preliminary experiment. One subject was asked to produce Mandarin vowel "e" in his normal voice. Then, glottal volume velocity was estimated by the microphone in the reflectionless tube. Figure 3 shows the estimated glottal waveform and corresponding magnitude spectrum. It is clear that there are no visible formant features of the vocal tract in these plots. Therefore, the vocal tract resonance was well eliminated by the reflectionless tube, making no contribution to the measured NFRF.

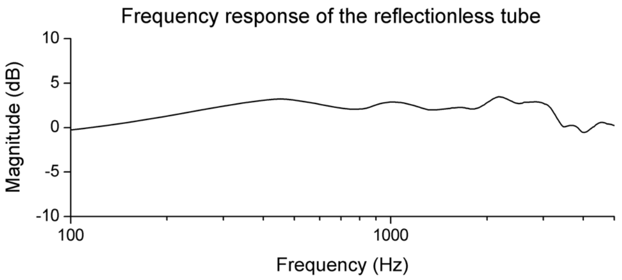

FIG. 2. The magnitude frequency response of the reflectionless tube. The maximum and minimum magnitudes were 3.4877 dB and −1.8187 dB, respectively.

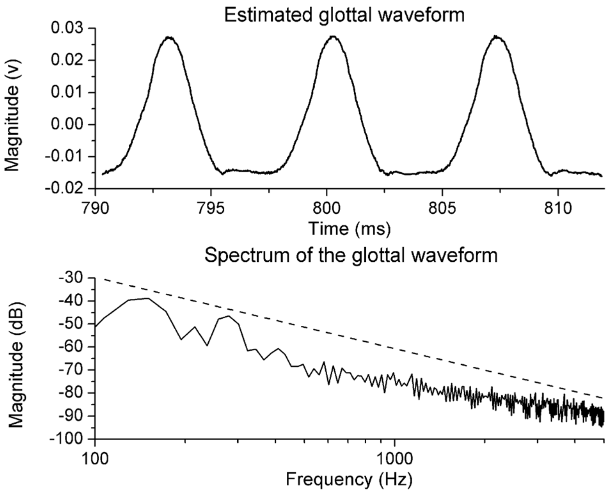

FIG. 3. The glottal waveform (Top) and magnitude spectrum (Bottom) measured by the reflectionless tube during normal speech production of Mandarin vowel "e."

Furthermore, to verify the resistance of the tube to environmental noise, average root-mean-square (RMS) values of the microphone signals were measured in three different cases. First, the pressure signal from microphone was recorded without the shaker vibration and speech production. The average RMS sound pressure level was −67.1 dB, representing the noise floor. Second, the open end of the tube was out of contact with the closed mouth, and the shaker against the neck was driven by broadband random noise, keeping 10 cm away from the open end. The average RMS power in this situation was −43.1 dB. Third, the subject connected his closed mouth with the open end of the tube, holding the vibrated shaker against the neck. Then, the average RMS power was −66.3 dB, which was 23.2 dB lower than that of case 2, and was very close to the noise floor. Therefore, it can be proved that there was no direct airborne transmission of sound from shaker to microphone.

III. RESULTS

A. General description of the NFRF

Figure 4 shows the mean value and standard deviation of the NFRFs of all subjects. The overall shape of the NFRF was similar to a low-pass filter, and the trend in the range of 100–5000 Hz can be described with three sections. In the first section, from 100 Hz to 310.1 ± 44.5 Hz, the curve was relatively stable with a little rising to a maximum gain. In the second section (S2), the NFRF curve rolled off at a slope of −8.4 ± 1.7 dB/octave from 310.1 Hz to average 1955.2 Hz. Then, in the last section (S3), the NFRF flattened out until 5000 Hz. In addition, the mean standard deviation was 2.7 ± 1.2 dB, reflecting the individual differences in the NFRFs.

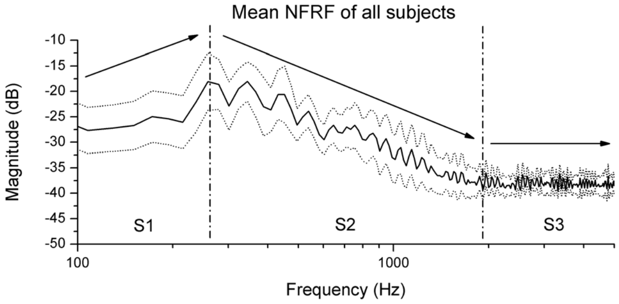

FIG. 4. The average NFRF (solid line) and standard deviation (dotted line) of all subjects. The arrows illustrate the trend of the NFRF curve.

B. Variation with different vowels

According to Sondhi's research, the tube performance of eliminating the vocal tract transfer function was dependent on the shape of vocal tract.[15] When the tract is held in a neutral position, the effect is almost entirely eliminated by the reflectionless tube. The Mandarin vowel "e" was a Chinese schwa, pronounced like the neutral vowel, [ə], in the International Phonetic Alphabet. So, in this case, the NFRF of vowel "e" was supposed to be the closest to the actual value because of the most accurate estimation of the volume velocity.[15]

Two representative results of a male and female subject are shown in Fig. 5 to illustrate the intervowel variation of the estimated NFRFs. Averaged from 100 to 5000 Hz, the standard deviation of the NFRFs was 2.7 ± 0.8 dB for the male example and 1.5 ± 0.8 dB for the female example.

In addition, the average NFRFs and the root mean square errors (RMSEs) of different vowels for all subjects were shown in Fig. 6. For each subject at each position, the RMSE was computed as the root mean square error of the NFRF curve between each vowel and the mean value. The results indicated that the NFRFs varied slightly across vowels except for vowel "i" (analysis of variance, $P < 0.05$). The vowel "e" had the minimum RMSE of 2.1 ± 0.5 dB, while the vowel "i" got the maximum RMSE of 4.9 ± 1.3 dB.

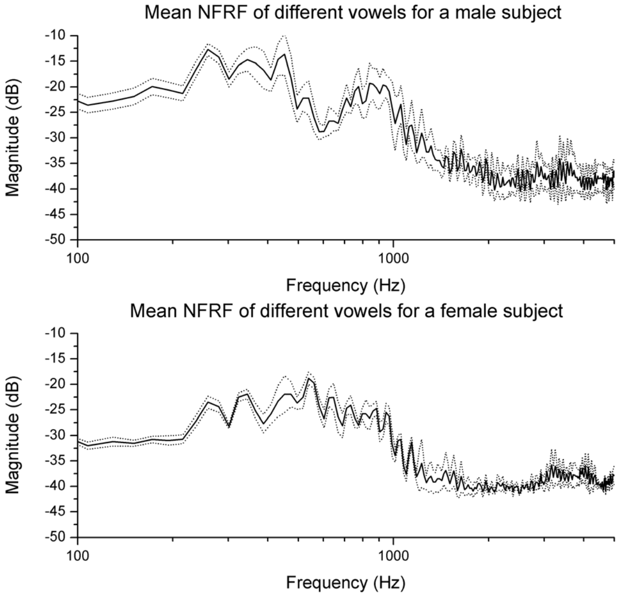

FIG. 5. The average NFRFs of a male subject (top) and a female subject (bottom) across different vowels. The solid line represents the mean value of the NFRFs of different vowels. The dotted line represents the standard deviation at each frequency.

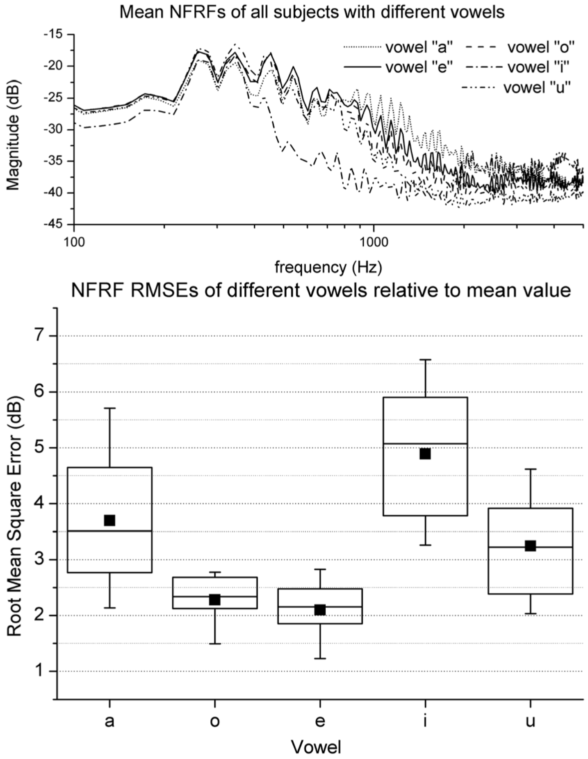

FIG. 6. The average NFRFs (top) and the root mean square errors (bottom) of different vowels for all subjects. Box plots show the median, interquartile, and minimum-maximum values; solid squares show the mean values.

C. Variation with different positions

For all subjects, there was no significant difference in the shape of the NFRF curve across shaker locations. However, the magnitude of the NFRF curve varied significantly with neck position for some subjects. Two examples of the NFRFs in Fig. 7 show the two types of the variations with different positions. For the top example, the average magnitude error between the NFRFs at any two positions was <2 dB. While for the bottom example, the NFRF magnitude at position 3 was on average 3.3 dB and 6.1 dB higher than that at positions 1 and 2, respectively. Furthermore, the average NFRFs of all subjects at different positions are plotted in Fig. 8. The NFRF at position 2 was not different from that at position 1, while the NFRF magnitude at position 3 was significantly 1.37 dB higher than those at positions 1 and 2.

D. Variation with sex

Figure 9 plots the average NFRFs for different subject groups. In each group, all subjects produced similar NFRFs with only a small difference. For male subjects, the standard deviation ranged from 1.0 dB at 2713.2 Hz to 6.8 dB at 236.9 Hz, with a mean of 3.1 dB. For female subjects, the average standard deviation was just 1.6 dB, with a minimum of 0.4 dB at 1485.8 Hz and a maximum of 3.0 dB at 2519.4 Hz.

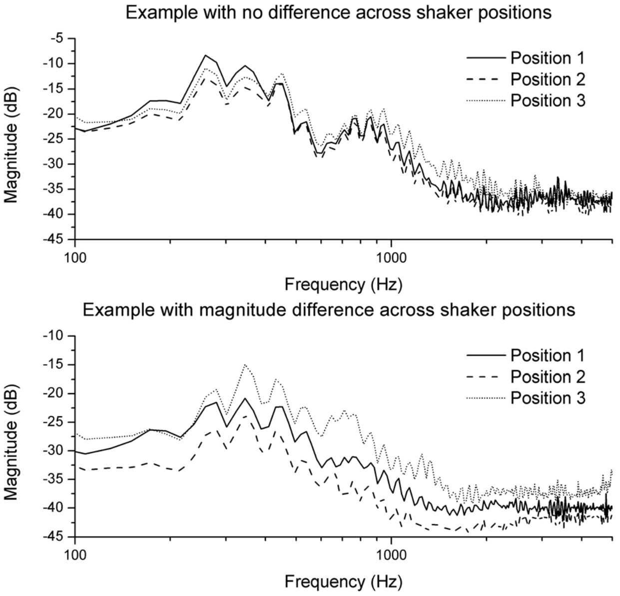

FIG. 7. The NFRFs of different shaker positions for two representative examples. (Top) An example shows no difference in the NFRFs with different positions. (Bottom) An example shows gain difference in the NFRFs with different positions.

However, the difference in the shape and magnitude of the NFRFs between male and female groups was significant. As shown in the bottom panel of Fig. 9, the frequency of the maximum gain for the male group was 292.8 ± 47.1 Hz, which was a little lower than 327.3 ± 38.5 dB for the female group. And in the frequency range of 100–300 Hz, the average magnitude of the male group was 5.9 dB higher than that of the female group. In S2, the descending slopes of male and female groups were -9.7 ± 1.1 dB/octave and -7.1 ± 1.0 dB/octave, respectively. Although the magnitude of the S3 curves for both the subject groups was -38 dB, the starting frequency of S3 for the male group (1761.2 Hz) was much lower than that of the female group (2149.2 Hz).

IV. DISCUSSION

This research proposed an easy and accurate way to measure the frequency response function of the neck tissue. So, the first purpose was to confirm the effectiveness of the method through comparison with former studies. Then, the NFRFs of normal subjects with producing Mandarin vowels were measured to reveal the sound transmission characteristics of the neck tissue. Finally, this work was expected to be useful for the research of normal and EL speech production.

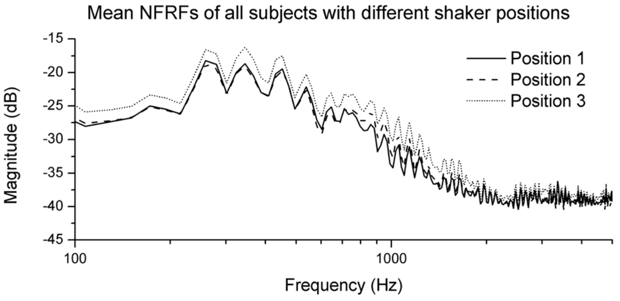

FIG. 8. The average NFRFs of all subjects with different shaker positions.

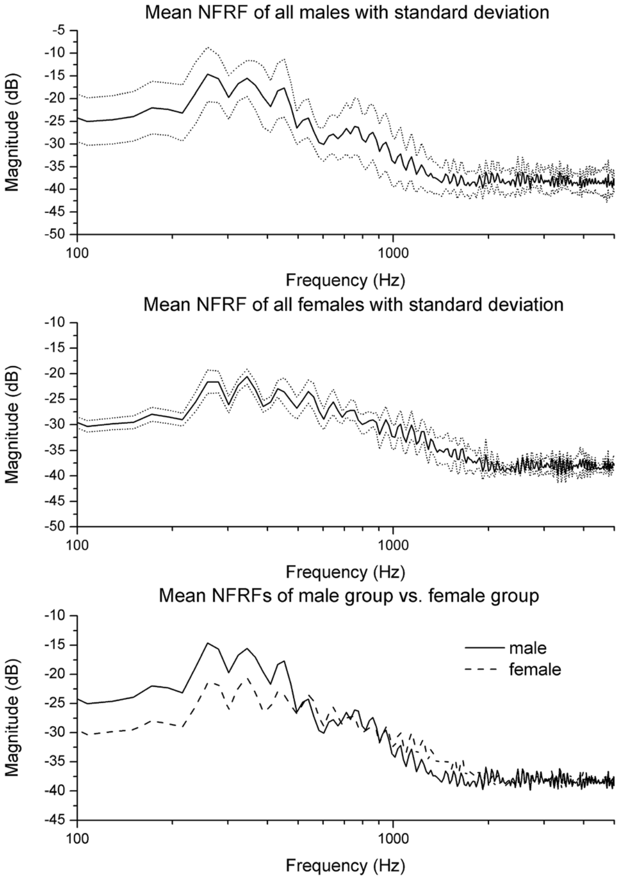

FIG. 9. The average NFRFs (solid line) and standard deviations (dotted line) for the male group (top) and female group (middle). (Bottom) The comparison of the average NFRFs between the male group (solid line) and female group (dashed line).

A. Effectiveness of the method

The NFRF shown in Fig. 4 was obtained from ten normal subjects. Considering that the individual differences were small, the mean curve represented the general frequency response function of the normal neck tissue, which played a major role of a low-pass filter in the sound transmission. These results are extremely similar with the previous findings reported by other researchers, especially Meltzner *et al.*[9,14] Their NFRF was also shaped like a low-pass filter, with a maximum gain ∼320.8 Hz and a following roll-off section to 3000 Hz, and then flattening out until 4000 Hz. Based on the good qualitative agreement, it is concluded that our procedure of measuring the NFRF is feasible and, once again, confirmed that the neck acts essentially like a low-pass filter with a peak ∼300 Hz.

The biggest advantage of this method is the use of the reflectionless tube to directly estimate the volume velocity that excites the vocal tract. The reflectionless tube is used to minimize the influence of vocal tract characteristics. The performance of the reflectionless tube determines the estimation accuracy of the volume velocity, as well as the NFRF. However, the tube performance depends on the specific vowels produced. The neutral vowel is the best condition to obtain the glottal volume velocity with minimal distortion.[15] Thus, as the results show in Fig. 3, it is concluded that the measured NFRF of vowel "e" can be considered as close to the actual value due to the accurate estimation of volume velocity.

Furthermore, the results in Fig. 5 show that the average standard deviation in NFRF across different vowels for each subject is small enough to ensure accurate estimations for all the vowels. Nevertheless, the NFRF of vowel "e" is still the closest to the average NFRF, while the vowel "i" has the largest deviation from the mean value (see Fig. 6). This outcome indicates that the NFRF measurement is influenced more in the case of vowel "i" than other vowels, which might be related to the larger variability of the vocal tract and the smaller cross-sectional areas of the oral cavity. In addition, it is possible that the difference of the NFRF is partly due to the small changes of neck tissue resulting from the different vocal tract postures. For this reason, all the NFRFs in this paper were characterized using the average value of five vowels, rather than the value of vowel "e." The small intervowel variability confirms that the average NFRF is still reliable.

Although the NFRFs found here are similar to those reported by Meltzner *et al.*,[14] there are still some differences. In the range of 100–310 Hz, the rising slope of the NFRF measured here is a little greater than that of Meltzner *et al.* (approximate to zero). On the other hand, above 2000 Hz, the NFRFs measured here remain constant, while those of Meltzner *et al.* roll off until 3000 or 4000 Hz. These small differences might be explained by the difference of the two methods for volume velocity estimation. First, the radiation noise has an obvious negative effect on Meltzner's NFRFs, especially in the low-frequency zone.[14] In contrast, the signal acquisition in the reflectionless tube is not sensitive to noise. Second, for the inverse filtering used by Meltzner *et al.*, the estimation accuracy of vocal tract transfer function (especially the formant magnitude) is very dependent on *a priori* knowledge of the NFRF. Accordingly, the NFRF measured by our method might be more accurate based on the good performance of the reflectionless tube.

In addition, Meltzner *et al.* found that the closed glottis may introduce spectral zeros into the vocal tract transfer function.[14] In our experiment, subjects were asked to keep their glottis open as much as they could, but it was difficult to guarantee this condition for each recording. Thus, we set up an experiment to study the influence of the glottal condition on the NFRF measurement. A male subject was asked to repeat the experiments with the glottal condition monitored by electroglottography. Figure 10 shows the average NFRFs measured with closed and open glottis. The differences between the two NFRFs are not significant, with a RMSE of 2.2 dB and a maximum difference of 4.5 dB. Despite the high-frequency difference, this result indicates that the glottal condition will not influence the NFRF measurement, and confirms the validity of our experiments.

Therefore, the proposed method in this paper is a simple and effective way to directly measure the NFRF without sacrificing accuracy. The results obtained here can be used to characterize the sound transmission features of normal neck tissue.

B. NFRF of normal subjects

In this work, the results are based on data from ten normal subjects (five male and five female). The low-pass characteristics of the neck tissue found here are not only consistent with previous works,[9,14] but also similar to those of other tissues, such as the subglottic respiratory system.[19] Thus, the outcomes are valuable and meaningful to reflect the average NFRF of normal subjects, and the NFRF variations with different shaker locations and different subjects.

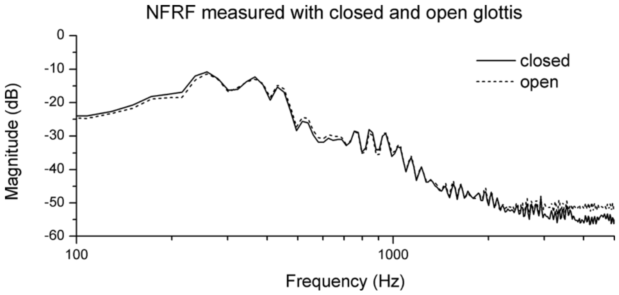

FIG. 10. The average NFRFs measured with closed glottis (solid line) and open glottis (dotted line).

For different locations, the shapes of the NFRFs are not significantly different for all the subjects. This result indicates that the differences of normal neck tissue will not have a remarkable effect on the spectral shape of sound transmission. In some individual cases, the gain of the NFRF was highly dependent on the shaker locations (see the bottom panel in Fig. 7). This outcome demonstrates that the attenuation of sound energy may be related with the tissue differences of normal neck (e.g., stiffness and thickness of tissues), and these differences are subject dependent. For the average results of all subjects, the NFRFs of positions 1 and 2 have no significant difference between each other, but with lower gains than that of position 3. This might be explained by the fact that both positions 1 and 2 are located over the thyroid cartilage, while position 3 is located over the thyrohyoid membrane. On one hand, the thyrohyoid membrane is less rigid than thyroid cartilage, and thus provides better coupling with the shaker. On the other hand, the smaller thickness at position 3 may also reduce transmission loss. Consequently, it is possible to deliver more sound energy through the thyrohyoid membrane than the thyroid cartilage. And it may be the reason why the position 3 is the common location for EL placement among laryngectomy patients.[18]

For different subjects, the individual variations of the NFRFs are not significant, but the gender differences of the NFRF are more remarkable. For each subject group, the limited variations might be due to the small tissue difference among individuals. In particular, the standard deviation of the female group is much less than that of the male group. This result demonstrates larger individual tissue property variations in men than in women, which might be due to the bigger laryngeal skeleton in normal male subjects, especially the thyroid cartilage. The physiological structure differences in the larynx between the male and female also led to a notable difference in their NFRFs, mainly in the low frequency ranging from 100 Hz to the frequency of maximum peak (∼310 Hz). Although the peak frequencies of these two groups are similar, the magnitude of the male group is significantly higher than that of the female group, and this finding is similar to Meltzner's results.[14] This might be explained by the fact that the male laryngeal cartilages tended to be ossified to a greater extent than female cartilages.[20] Furthermore, the ossification progress of thyroid cartilage is correlated with the ages for both genders. So, the results in this work might only reflect the neck characteristics for the ages of 20–30 yr.

C. Applications

Since the method has been proved to be easy and effective to measure an accurate NFRF, we conclude this paper by suggesting some potential uses based on these results.

As mentioned by Meltzner *et al.*,[14] the NFRF can be accurately modeled as a low-pass filter for improving the EL driving signal. More than that, for normal speech production, the model of the NFRF can also be used in filtering the neck characteristics from the signal captured by contact microphone or accelerometer sensor for speech recognition or vocal function evaluation. In addition, individual differences of the NFRFs show some relationship with the physiological structure differences, so an intriguing possibility is that of using NFRF to identify physiological features of the neck tissue, such as tissue stiffness and cartilage ossification. In such an application, it would be necessary to figure out the relationship between NFRF and anatomical structures.

In this work, all the participants are normal subjects. The method can also be applied to measure the NFRF of laryngectomized subjects. For EL speech production, the NFRF can be used not only in the inverse filtering of the sound source, but also in the training of using EL. The NFRF variation with different shaker locations is a good guide for selection of EL placement to produce loud speech. Therefore, in the next step, we will use this method to measure the NFRF of the laryngectomized patients for improving the EL sound source.

V. CONCLUSION

Using the reflectionless tube is an easy and effective way to measure the NFRF without sacrificing accuracy. The NFRFs were measured in ten laryngeal subjects (five males and five females) producing different Mandarin vowels and at three different positions on the neck. The general NFRF shape confirmed that the neck resembled a low-pass filter, with a peak frequency near 310 Hz, rolling off at a slope of -8.4 dB/octave to 2000 Hz, and then remaining constant until 5000 Hz. For each subject, the neck tissue differences at different positions had no impact on the shape of the NFRF. The physiological structure differences between genders significantly influenced the NFRF. Therefore, these results may be useful for research in speech recognition and vocal function evaluation, and this method can be used to measure the NFRF of laryngectomees for EL speech improvement.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under Grants No. 11274250 and 61271087, and Research Fund for the Doctoral Program of Higher Education of China under Grant No. 20120201110049, and Project funded by China Postdoctoral Science Foundation under Grant No. 2014M552457.

[1] H. A. Cheyne, "Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck," Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2002.

[2] H. A. Cheyne, H. M. Hanson, R. P. Genereux, K. N. Stevens, and R. E. Hillman, "Development and testing of a portable vocal accumulator," J. Speech Lang. Hear. Res. **46**, 1457–1467 (2003).

[3] J. G. Švec, I. R. Titze, and P. S. Popolo, "Vocal dosimetry: Theoretical and practical issues," in *Proceeding Papers for the Conference Advances in Quantitative Laryngology, Voice and Speech Research* (2003), pp. 1–8.

[4] J. G. Švec, I. R. Titze, and P. S. Popolo, "Estimation of sound pressure levels of voiced speech from skin vibration of the neck," J. Acoust. Soc. Am. **117**, 1386–1394 (2005).

[5] R. E. Hillman, J. T. Heaton, A. Masaki, S. M. Zeitels, and H. A. Cheyne, "Ambulatory monitoring of disordered voices," Ann. Otol. Rhinol. Laryngol. **115**, 795–801 (2006).

[6] D. D. Mehta, M. Zañartu, S. Feng, H. I. Cheyne, and R. E. Hillman, "Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform," IEEE Trans. Biomed. Eng. **59**, 3090–3096 (2012).

[7] Y. Zheng, Z. Liu, Z. Zhang, M. Sinclair, J. Droppo, L. Deng, A. Acero, and X. Huang, "Air- and bone-conductive integrated microphones for robust speech detection and enhancement," in *IEEE Automatic Speech Recongnition and Understanding Workshop (ASRU)*, Virgin Islands, USA (2003), pp. 249–254.

[8] M. Garciarena, H. Franco, K. Sonmez, and H. Bratt, "Combining standard and throat microphones for robust speech recognition," IEEE Signal Process. Lett. **10**, 72–74 (2003).

[9] S. Dupont, R. Christophe, and D. Bachelart, "Combined use of close-talk and throat microphones for improved speech recognition under non-stationary background noise," in *COST278 and ISCA Tutorial and Research Workshop (ITRW) on Robustness Issues in Conversational Interaction* (2004).

[10] Z. Zhang, L. Zicheng, M. Sinclair, A. Acero, L. Deng, J. Droppo, X. Huang, and Y. Zheng, "Multi-sensory microphones for robust speech detection, enhancement and recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (2004), Vol. III, pp. 781–784.

[11] E. Erzin, "Improving throat microphone speech recognition by joint analysis of throat and acoustic microphone recordings," IEEE Trans. Audio Speech Lang. Process. **17**, 1316–1324 (2009).

[12] T. Dekens, W. Verhelst, F. Capman, and F. Beaugendre, "Improved speech recognition in noisy environments by using a throat microphone of accurate voicing detection," in *18th European Signal Processing Conference* (2010), pp. 23–27.

[13] R. L. Norton and R. S. Bernstein, "Improved Laboratory Prototype Electrolarynx (LAPEL): Using inverse filtering of the frequency response function of the human throat," Ann. Biomed. Eng. **21**, 163–174 (1993).

[14] G. S. Meltzner, J. B. Kobler, and R. E. Hillman, "Measuring the neck frequency response function of laryngectomy patients: Implications for the design of electrolarynx devices," J. Acoust. Soc. Am. **114**, 1035–1047 (2003).

[15] M. M. Sondhi, "Measurement of the glottal waveform," J. Acoust. Soc. Am. **57**, 228–232 (1975).

[16] M. Tanabe, N. Isshiki, and K. Kitajima, "Application of reflectionless acoustic tube for extraction of the glottal waveform," Stud. Phonol. **12**, 31–38 (1978).

[17] J. Walker and P. Murphy, "A review of glottal waveform analysis," in *Progress in Nonlinear Speech Processing* (Springer, Berlin, 2007), pp. 1–21.

[18] G. S. Meltzner, "Perceptual and acoustic impacts of aberrant properties of electolaryngeal speech," Doctoral dissertation, Harvard-MIT Division of Health Sciences and Technology, Harvard University, Cambridge, MA, 2003.

[19] G. R. Wodicka and D. C. Shannon, "Transfer function of sound transmission in subglottal human respiratory system at low frequencies," J. Appl. Physiol. **69**, 2126–2130 (1990).

[20] K. D. Dang-Tran, F. Dedouit, F. Joffre, D. Rougé, H. Rousseau, and N. Telmon, "Thyroid cartilage ossification and multislice computed tomography examination: A useful tool for age assessment?," J. Forensic. Sci. **55**, 677–683 (2010).