

基于复倒谱的声门下噪音源研究

赵守国¹, 孙义和¹, 王素品²

(1. 清华大学微电子学研究所设计室, 北京 100084; 2. 西安交通大学生物医学工程系, 陕西 西安 710049)

摘要: 根据噪音 - 呼吸系统生理和生物物理原理, 运用研制的噪音测试系统, 获取颈前 V 型槽处的传导语音信号, 基于复倒谱技术提取声门下噪音源特性。研究男女胸声区不同发声条件下声带振动和声门下噪音源特性。结果表明, 声门下噪音源能够反映声带振动特征, 是研究声带振动特性和喉部疾病诊断的一种有效方法。
关键词: 复倒谱; 声门下噪音源; 传导语音
中图分类号: Q811.3 **文献标识码:** A **文章编号:** 1000-6737(2002)04-0434-06

声带高速、复杂、精微的三维振动限制了各种先进的医学测试与成像技术在发声研究中的应用。过去有关发声特性的研究主要从喉部解剖方面出发, 利用喉动态镜观察声带振动, 进行噪音声学分析与评价等。近年来, 国外学者利用反滤波声门图技术研究了声门气流体积、速度、波形特征随性别、年龄的变化关系^[1-4], Tong 等^[5]从发声空气动力学与发声效率的角度研究了儿童、妇女和成年男性噪音源特性。目前对声带振动和噪音源特性的研究还远不够系统深入^[6-14], 实验研究方法尚未得到一致的认可, 实验数据比较分散, 难以进行比较。

根据噪音 - 呼吸系统生理和生物物理原理, 各部分在发声过程中所起的作用不同但又有内在的联系。我们以声门为中心从概念上将这一声学与空气动力学耦合系统分割为上声道、声门上噪音源、声带、声门下噪音源、下声道、肺胸系统和呼吸噪音等几部分。运用研制的噪音测试系统, 获取颈前 V 型槽处的传导语音信号, 研究了男女胸声区不同发声条件下声带振动和声门下噪音源特性。对发声基

础研究、语言声学、艺术噪音学和肺胸系统声传递特征等领域有重要意义, 并得到了一些有意义的实验结果。

1 声门下噪音源产生的物理模型

在语音产生过程中, 声带振动产生的声源通过上声道传输产生语音, 同时也经过下声道传输至体表形成传导语音, 它的物理模型如图 1 所示^[12], 声门下噪音源与声门上噪音源是不同的。传导语音在向下传播过程中, 在气管、支气管和肺部组织产生噪声, 这些噪声本身不是由肺胸系统传播声门下噪音源产生的, 而是在气管、支气管和肺部组织产生湍流而形成的, 根据传导语音声传播物理特征, 把噪音 - 呼吸系统等效为如图 1 所示的物理模型。 $p_{sb}(n)$ 为发稳态元音时的声门下噪音源, 是一准周期脉冲序列, $r_{sb}(n)$ 为发清音时的声门下噪音源, 是一随机脉冲序列。

下声道被看成是一个缓变的时变滤波器, 并用

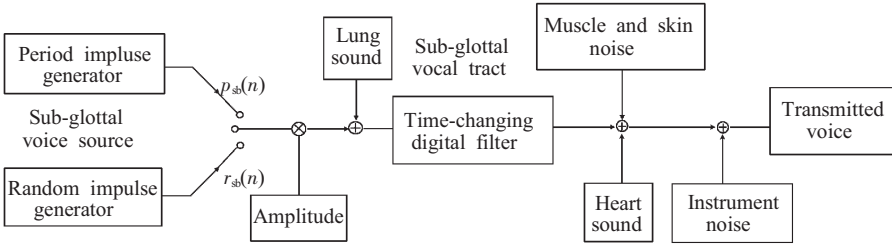


Fig.1 Physical model of transmitted sound production

它的一些自然频率来刻画, 这些自然频率就是下声道传输中的谐振点, 其传递函数要比上声道简单^[15,16]。在保持发音器官的位置和形状不变的一段短时间内, 也把下声道看作是一个线性非时变系统, 这时, 下声道的传递函数也由声道分量和附加分量组成, 声道分量是:

$$V_{sb}(z) = \frac{|A|}{\prod_{k=1}^p (1-d_k z^{-1})(1-d_k^* z^{-1})} \quad |d_k| < 1 \quad (1)$$

式中 d_k 表示声道的自然频率, A 为常数。附加分量是:

$$G_{sb}(z) = |B| \prod_{k=1}^{m_1} (1-b_k z^{-1}) \prod_{k=1}^{m_0} (1-s_k z^{-1}) \quad |b_k| < 1, |s_k| < 1 \quad (2)$$

B 为常数。下声道的传递函数是:

$$H_{sb}(z) = V_{sb}(z) \cdot G_{sb}(z) \quad (3)$$

下声道系统的脉冲响应为 $h_{sb}(n)$ 。发浊音时经下声道传至体表的传导语音信号 $y_{tr}(n)$ 为:

$$\begin{aligned} y_{tr}(n) &= p_{sb}(n) \cdot h_{sb}(n) + N_m(n) \cdot h_{Nsb}(n) + H(n) + M(n) + S(n) \\ &= p_{sb}(n) \cdot v_{sb}(n) \cdot g_{sb}(n) + N_m(n) \cdot h_{Nsb}(n) \\ &\quad + H(n) + M(n) + S(n) \end{aligned} \quad (4)$$

从发声生理及生物物理原理得知, 在发稳态元音时, 声门下噪音源为一准周期脉冲信号^[3]。从(4)式可以看出, 输出 $y_{tr}(n)$ 主要由两部分组成, 第一部分 $p_{sb}(n) \cdot h_{sb}(n)$ 是由声门下噪音源通过下声道和颈部肌肉组织传至体表的准周期信号, 第二部分为随机噪声信号, 包括在气管、支气管和肺部组织形成湍流而产生的噪声 $N_m(n)$ 经过系统 $h_{Nsb}(n)$ 而产生的噪声输出、心音信号 $H(n)$ 、肌肉和皮肤产生的噪声 $M(n)$ 、仪器噪声 $S(n)$ 等^[12]。

2 复倒谱技术提取声门下噪音源

我们主要研究浊音发音时的声门下噪音源。这是因为, 在发清音和爆破音时, 输出信号的特征主要是由声道决定的, 对其研究并不能很好地反映声带振动特性, 而浊音则不同, 研究它不仅可以获得下声道的特征, 同时也可以更好地获得声带振动特性。采用如图2所示的数据采集系统纪录颈前V型槽处的传导语音和口腔输出语音。

根据上述提出的语音和传导语音传播的物理模型, 传导语音输出包括准周期的谐波分量和噪声信号。如果能滤除噪声分量而保留谐波分量, 并利用

谐波分量的输出提取声门下噪音源, 结果能够更好地反映声带振动特性。我们采用文献[17]论述的小波去噪方法, 首先滤除传导语音中的噪声分量, 再利用复倒谱技术提取声门下噪音源^[15,16]。

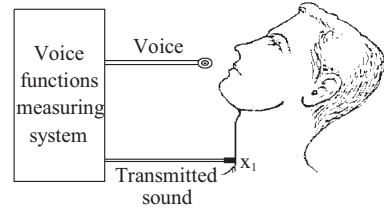


Fig.2 Acquiring transmitted sound

声门下噪音源到胸前颈下V型槽 x_1 处的系统脉冲响应为 $h_{sb}(n)$, x_1 处输出的传导语音 $y_{tr}(n)$ 经小波去噪后的传导语音为 $y_{tr}'(n)$, $p_{sb}(n)$ 为声门下噪音源, 则有:

$$y_{tr}'(n) = p_{sb}(n) \cdot h_{sb}(n) = p_{sb}(n) \cdot v_{sb}(n) \cdot g_{sb}(n) \quad (5)$$

从上式中我们可以看出, $y_{tr}'(n)$ 是由激励源和下声道脉冲响应卷积而生成的。当发浊音时, 声门下噪音源 $p_{sb}(n)$ 为一等间隔的脉冲串, 设周期为 N_p , 则有:

$$p_{sb}(n) = \sum_{r=0}^M a_r \delta(n - rN_p) \quad (6)$$

M 为采样期间的脉冲数, a_r 为脉冲幅度。利用复倒谱技术估计声门下噪音源 $p_{sb}(n)$ 。

根据(3)式, 由复倒谱定义得到:

$$\hat{y}_{tr}'(n) = \hat{p}_{sb}(n) + \hat{h}_{sb}(n) \quad (7)$$

其中 $p_{sb}(n)$ 的复倒谱为:

$$\hat{p}_{sb}(n) = \log a_0 \delta(n) - \sum_{k=1}^{\infty} \left[\frac{1}{k} \sum_{r=1}^M (a_r')^k \delta(n - kN_p) \right] \quad (8)$$

a_r' 为 $p_{sb}(n)$ 的Z变换多项式的根, 从上式可以看出 $p_{sb}(n)$ 的复倒谱是周期为 N_p 的一系列脉冲, 当采样点数有限时, 即 M 为有限长时, 它的复倒谱是无限长的, 幅度随时间变化而变化, 且随时间的增加而衰减。

下声道及相应的复倒谱为:

$$\hat{h}_{sb}(n) = \begin{cases} \log |AB| & n=0 \\ -\sum_{k=1}^{m_1} \frac{b_k^n}{n} + \sum_{k=1}^{m_1} \frac{d_k^n}{n} + \sum_{k=1}^p \frac{(d_k^*)^n}{n} & n>0 \\ \sum_{k=1}^{m_0} \frac{s_k^{-n}}{n} & n<0 \end{cases} \quad (9)$$

从(9)式可以看出:

(1) $\hat{h}_{sb}(n)$ 是一个双边序列, 即在 $-\infty < n < \infty$ 范围内,

$\hat{h}_{sb}(n)$ 皆有值。

(2) 由于 $|b_k| < 1$, $|s_k| < 1$, $|d_k| < 1$, 所以倒谱 $\hat{h}_{sb}(n)$ 是一个衰减序列, 即随着 $|n|$ 的增大, $|\hat{h}_{sb}(n)|$ 减小。

(3) 倒谱 $\hat{h}_{sb}(n)$ 随 $|n|$ 增大而衰减的速度至少比 $\frac{1}{|n|}$ 要

快, 具体为: $\left| \hat{h}_{sb}(n) \right| < C \left| \frac{r^n}{n} \right|$, $(-\infty < n < \infty)$

这里, r 是 $|b_k|$ 、 $|s_k|$ 、 $|d_k|$ 中的最大者, C 为一常数因子, 因而, 倒谱 $\hat{h}_{sb}(n)$ 比 $h_{sb}(n)$ 更集中于原点附近, 或者说 $\hat{h}_{sb}(n)$ 更具有短时性。

(4) 当 $h_{sb}(n)$ 为最小相位序列时, 即 $h_{sb}(n)$ 的 Z 变换的极点和零点都在单位圆内部, 则 $\hat{h}_{sb}(n)$ 只在 $n \geq 0$ 范围内有值, 即为因果序列。

从 $p_{sb}(n)$ 和 $h_{sb}(n)$ 的复倒谱知道, $\hat{h}_{sb}(n)$ 主要集中在低时部分, 这样当 $p_{sb}(n)$ 为周期脉冲串时, $\hat{p}_{sb}(n)$

与 $\hat{h}_{sb}(n)$ 分布在不同的时段, 利用时域滤波器就可以分别求出 $p_{sb}(n)$ 和 $h_{sb}(n)$ 。当采样频率为 10 kHz 时, 复倒谱时域滤波器采用如(10)式的计算公式, M 的取值与性别和发声方式有关, 一般选择 $20 < M < 50$, 当 $M=40$ 时, 相当于 4 ms。

$$l(n) = \begin{cases} 1, & |n| \leq M \\ 0, & |n| > M \end{cases} \quad (10)$$

3 胸声区发声方式的声门下噪音源

选择能正常发声的男性和女性各 20 名, 按图 2 所示的方法同时记录口腔输出语音和颈前 V 型槽处的传导语音。采样频率为 10 kHz。图 3-6 给出了一位健康成年男性采用胸声区发声方式, 分别发浊音 [a:]、[i:]、[u:] 和 [æ] 时所得到的处理结果。在图中, (A) 是颈前 V 型槽处的传导语音; (B) 是利用复倒谱技术由传导语音得到的声门下噪音源; (C)

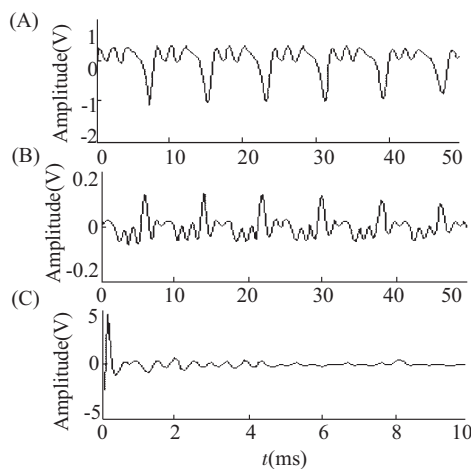


Fig.3 [a:] transmitted sound and sub-glottal voice source of healthy man. (A) Transmitted sound in V-slot location; (B) Acquired sub-glottal voice source; (C) Complex cepstrum of transmitted sound

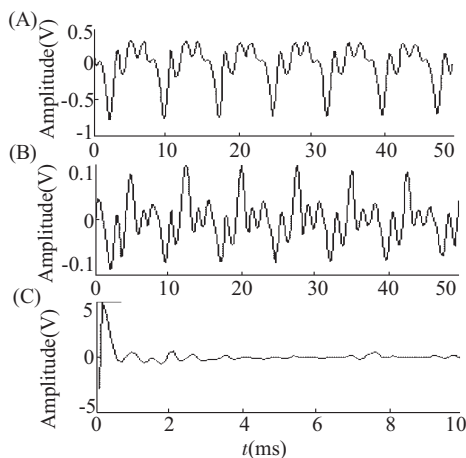


Fig.4 [i:] transmitted sound and sub-glottal voice source of healthy man. (A) Transmitted sound in V-slot location; (B) Acquired sub-glottal voice source; (C) Complex cepstrum of transmitted sound

是传导语音的复倒谱。

从所记录的传导语音和口腔输出语音中可以看出, 如消除因声道长度不同引起的时延, 传导语音的极大值对应语音的极小值。不同的浊音, 由口腔共鸣腔的作用, 使其语音输出变化很大, 但传导语音的输出十分相似, 变化较小。这主要是因为下声道在发不同的浊音时, 基本是不变的, 这与我们前面的理论分析是一致的, 由得到的传导语音复倒谱可以看出, 其在 8 ms 处有一个明显的峰值, 这是由声门下噪音源的周期性引起的。从语音和复倒谱图中都可以得到该试验者的基频为 125 Hz 左右。

从声门下噪音源结果可以看出, 它与以前得到的噪音源结果有着很大的区别。不同的发音, 其声门下噪音源变化不大, 与以前得到的声门下压力波

形也有着很大的差别，但我们可以看出它比由口腔输出语音所得到的噪音源能更好地反映声带振动特性。

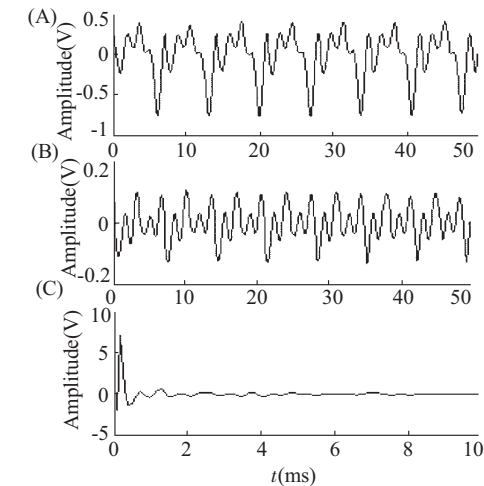


Fig.5 [u:] transmitted sound and sub-glottal voice source of healthy man. (A) Transmitted sound in V-slot location; (B) Acquired sub-glottal voice source; (C) Complex cepstrum of transmitted sound

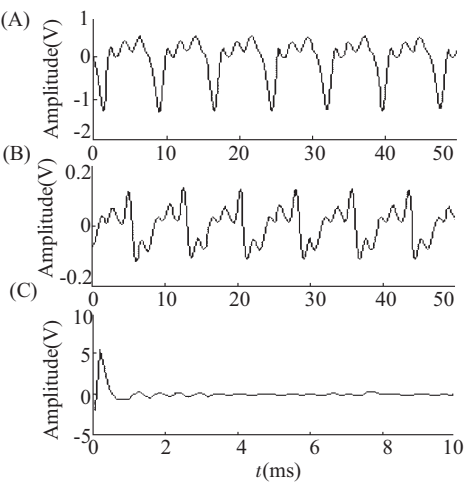


Fig.6 [æ] transmitted sound and sub-glottal voice source of healthy man. (A) Transmitted sound in V-slot location; (B) Acquired sub-glottal voice source; (C) Complex cepstrum of transmitted sound

图 7-10 给出了一位健康成年女性采用胸声区发声发式，分别发浊音[a:]、[i:]、[u:]和[æ]时所得到的处理结果。该受试者的语音和传导语音的基频为 200 Hz 左右，复倒谱在 5 ms 处出现一个明显的峰值。从结果可以看出，女性的声门下噪音源比男性变化的幅度要大，如在发[a:]时的基频在 200 Hz，而发[i:]时的基频则上升为 300 Hz 左右。在复倒谱运算中，时域滤波器的 M 值取 20 左右。

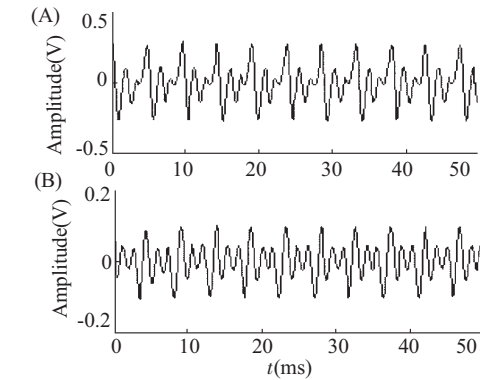


Fig.7 [a:] transmitted sound and sub-glottal voice source of healthy woman. (A) Transmitted sound in V-slot location; (B) Acquired sub-glottal voice source

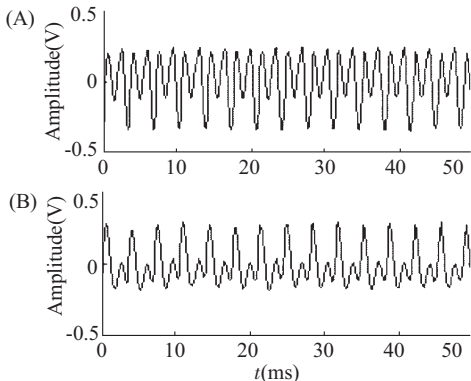


Fig.8 [i:] transmitted sound and sub-glottal voice source of healthy woman. (A) Transmitted sound in V-slot location; (B) Acquired sub-glottal voice source

从结果可以看出，不同的发音，女性声门下噪音源的变化要比男性声门下噪音源变化大，但与口腔语音输出的结果相比变化幅度还是比较小的。实验过程中可以看出，女性胸声区自然发声的声门下噪音源受个体差异影响比较大，与声带的振动特征的相关性比男性差一些，为了解决这个问题，可以让其发比自然发声更低沉的声音，尽可能降低声带振动频率，使其声带振动方式类似于男性胸声区的声带振动方式。这样可以很好地消除自然发声条件

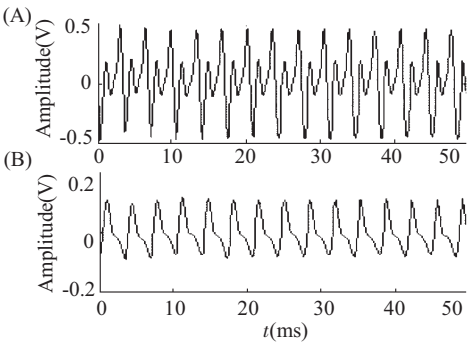


Fig.9 [u:] transmitted sound and sub-glottal voice source of healthy woman. (A) Transmitted sound in V-slot location; (B) Acquired sub-glottal voice source

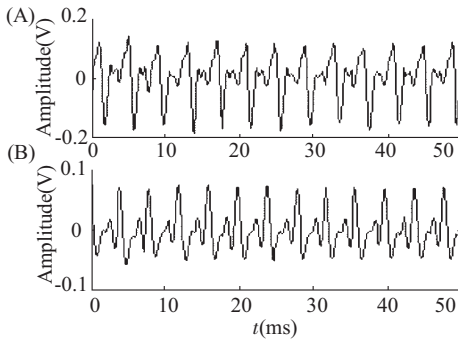


Fig.10 [æ] transmitted sound and sub-glottal voice source of healthy woman. (A) Transmitted sound in V-slot location; (B) Acquired sub-glottal voice source

下女性声带振动的个体差异，更好地利用声门下嗓音源研究声带振动的特性。

4 结论与讨论

本文据嗓音 - 呼吸系统生理和生物物理原理，运用研制的嗓音测试系统，获取颈前 V 型槽处的传导语音信号。利用复倒谱技术提取声门下嗓音源，研究男女胸声区不同发声条件下声带振动和声门下嗓音源特性。从人体实验结果可以看出，声门下嗓音源与声带振动具有较好的相关性，与理论分析的结果是一致的。该方法是研究声带振动的一种有效方法。对发声基础研究、语言声学、艺术嗓音学和肺胸系统声传递特征等领域有重要意义。

参考文献:

[1] Titze IR. Principles of voice production [M]. Englewood Cliffs WJ: Prentice Hall, 1994.

[2] Titze IR. Physiologic and acoustic differences between male and female voices[J]. *J Acoust Soc Am*, 1989,85:1699-1707.

[3] Fant G. 嗓音声源研究[J]. *声学学报*, 1988,13:135-146.

[4] Arend MS, Hero PW. Glottal volume velocity waveform characteristics in subjects with and without vocal training, related

to gender, sound intensity, fundamental frequency, and age[J]. *J Acoust Soc Am*, 1996,100:3360-3373.

[5] Tong JN. Vocal efficiency as a function of vocal intensity: A study of children, women and men[J]. *J Acoust Soc Am*, 1995,97:1885-1892.

[6] Holmberg EB, Hillman RE, Perkell JS. Glottal airflow and transglottal air pressure measurements for male and female in soft, normal and loud voice[J]. *J Acoust Soc Am*, 1988, 84:511-529.

[7] Berke GS, Kreiman J, Gerratt BR, et al. The effect of air flow and medical adductory compression on vocal efficiency and glottal vibration[J]. *Head and Neck Surgery*, 1990,102: 212-218.

[8] Van JW Berg D . Direct and indirect determination of the mean subglottal pressure[J]. *Folia Phoniat*, 1956,52:1-24.

[9] Isshiki N. Vocal efficiency index[A]. In: *Vocal Fold Physiology*[M]. University of Tokyo Press, 1981. 193-203.

[10] Kakita Y. Wave-flow index: A measure of vocal efficiency [C]. *The Proceedings of IEEE ICASSP 86*, 1986. 685-688.

[11] 王素品, 万明习, 赵守国, 等. 男女嗓音源特性研究[J]. *声学学报*, 1999,24:128-135.

[12] 赵守国. 声门 - 呼吸 - 肺胸耦合系统分割与特征表达[D]. 博士学位论文, 西安交通大学, 2000.

[13] 赵守国, 万明习, 王素品, 等. 基于声门下嗓音源的肺胸系统声传递特征研究[J]. *生物物理学报*, 2001,17:142-148.

[14] 万明习, 程敬之, 王素品, 等. 声门发声效率的研究[J]. *科学通报*, 1992,37:1507-1509.

[15] Oppenheim AV, Schafer RW. Homomorphism analysis of speech[J]. *IEEE Transactions on Audio and Electroacoustics*, 1968,16:221-226.

[16] Noll AM. Cepstrum pitch determination[J]. *J Acoust Soc Am*, 1967,41:293-309.

[17] Mallat S, Zhang S. Characterization of signals from multi-scale edges [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992,38:617-643.

A STUDY ON SUB-GLOTTAL VOICE SOURCE BASED ON COMPLEX CEPSTRUM

ZHAO Shou-guo¹, SUN Yi-he¹, WANG Su-pin²

(1. Institute of Microelectronics, Tsinghua University, Beijing 100084, China;

2. Department of Biomedical Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Based on the physiology and biophysics of voice-respiration system, the voice production and aerodynamic coupling system can be divided into supra-glottal vocal tract, supra-glottal voice source, vocal cord, sub-glottal voice source, sub-glottal vocal tract, lung, respiratory sound source and lung-thorax sections. Using the voice functions measuring system developed, the transmitted sound signals were acquired. Based on complex cepstrum, the vocal tract vibrations and sub-glottal voice sources of both male and female in chest region were examined. Results showed vocal cord vibration modes and characteristics. It would be one of good methods for measurement of vocal cord vibration characteristics and diagnosis of thorax diseases.

Key Words: Complex cepstrum; Sub-glottal voice source; Transmitted sound