**基于谱减法的变频电子喉语音增强方法对比研究**

李阳 牛海军 王立 冯亦军

**【摘要】** **目的** 电子喉因具有发声简单、使用方便等特点而成为喉头切除患者最常用的言语发声辅具，但是目前的电子喉语音仍存在辐射噪声大、发声机械等诸多缺陷，通过消除辐射噪声来改善语音质量成为诸多研究者的目标之一。谱减法是消除基频恒定电子喉辐射噪声最常用的方法，为探究谱减法对变频电子喉语音的去噪效果，本文对基于谱减法的变频电子喉语音增强方法进行了研究。**方法** 分别采用经典谱减法、多带谱减法、基于感知加权技术的改进谱减法以及基于加权函数的改进谱减法对变频电子喉语音进行去噪处理，并对去噪后语音进行主客观评价，比较不同谱减方法的去噪效果。**结果** 经典谱减法在去除部分辐射噪声的同时引入了音乐噪声，降低了增强后语音质量；多带谱减法、基于感知加权技术的改进谱减法以及基于加权函数的改进谱减法都有效减少了辐射噪声，抑制了音乐噪声，提高了变频电子喉语音的可接受度，但是对于语音的可懂度影响较小。**结论** 基于谱减法的语音增强方法可以有效减少变频电子喉辐射噪声，提高语音的可接受度，对改善变频电子喉语音的听觉质量有较大帮助。

**【关键词】：**电子喉；变频；谱减法；语音增强[[1]](#footnote-1)

**Contrastive Study on the Enhancement of Pitch Adjustable Electrolarynx Speech Based on** **Spectral Subtraction Technology**

Li Yang1 NIU Hai-Jun1\* Wang Li1 Feng Yi-Jun1

*School of Biological Science & Medical Engineering, Beihang University, Beijing 100191,*

*Key Laboratory of the Ministry of Education for Biomechanics and Mechanobiology*

*(\*Correspondence author:**hjniu@buaa.edu.cn)*

【Abstract】**Objective** Electrolarynx (EL) is the most common device to provide a voice for laryngectomees since it is easy to use and phonate. However, the EL speech has several shortcomings, such as the unnaturalness and the radiated noise. To improve the speech quality by eliminating the radiated noise becomes one of the goals for researchers. Spectral subtraction method is the most common way to eliminate the radiated noise of EL whose pitch is constant. In order to explore the denoising effects of the spectral subtraction methods for pitch adjustable EL speech, in this paper, we conduct a study of enhancement of pitch adjustable EL speech based on spectral subtraction methods. **Methods** The classical spectral subtraction, the multi-band spectral subtraction, the improved spectral subtraction based on perceptual weighting technology and the improved spectral subtraction based on weighting function were used to remove the radiated noise of EL speech, and subjective and objective estimation were used to evaluate and compare the enhanced speeches. **Results** The classical spectral subtraction method removed part of the radiated noise while it introduced the music noise, which reduced the quality of enhanced speech; The other three methods including multi-band spectral subtraction method, the improved spectral subtraction method based on perceptual weighting technology and the improved spectral subtraction method based on weighting function could reduce the noise obviously, suppress the music noise and improved the speech acceptability effectively, but had little influence on speech Intelligibility. **Conclusions** The enhancement methods based on spectral subtraction can effectively reduce the radiated noise of EL speech and improve the acceptability, which is greatly helpful for improving speech quality.

**【Key Words】**Electrolarynx; Pitch adjustable;Spectral subtraction; Speech enhancement

# 0 引言

食管发声法、气管食管发声法与电子喉发声法是喉头切除患者常用的三种语音恢复方法，其中电子喉因具有发声简单、使用方便等特点而成为喉头切除患者最常用的言语发声辅具[[[2]](#endnote-1),[[3]](#endnote-2),[[4]](#endnote-3)]。但是目前的电子喉语音仍存在辐射噪声大、发声机械等诸多缺陷，如何有效消除辐射噪声，以改善电子喉语音质量成为诸多研究者追寻的目标之一。

Barney等人的测试实验结果显示，当口闭合时，电子喉的辐射噪声强度约为20-25dB[[[5]](#endnote-4)] ，且能量集中于400-1kHz和2k-4kHz这两个频率段。对于同一种电子喉，Weiss的报告中指出当口闭合时，根据使用者的颈部特性不同，噪声强度大约在4-15dB间变化，同时主观测试显示，当信噪比低于4dB时，语音可懂度明显降低[[[6]](#endnote-5)]。为了有效消除辐射噪声，研究者使用了物理屏蔽和信号处理等多种手段[[[7]](#endnote-6),[[8]](#endnote-7)]，其中信号处理方法由于可以在声传输环节消除辐射噪声而引起多人关注。基于信号处理的电子喉消噪方法主要有两种，即：自适应方法和谱减法。自适应方法因额外需要一个参考输入，算法较复杂[[[9]](#endnote-8)]；谱减法通过减去噪声谱来获取更纯净的语音，具有运算量小、易于实现、增强效果较好等特点，应用更广泛[[[10]](#endnote-9),[[11]](#endnote-10),[[12]](#endnote-11)]。但是经典谱减法在语音增强的同时会引入“音乐噪声”，影响处理后语音的质量。为了抑制音乐噪声，Li等人根据电子喉辐射噪声的频谱非均匀分布的特点，在去噪时采用了不同频率段用不同谱减系数的多频带非线性谱减方法，去噪效果有所提高[[[13]](#endnote-12)]；Liu等人基于人耳掩蔽效应，将低速率编解码器中的感知加权技术应用于谱减法，也较好的抑制了残余音乐噪声[[[14]](#endnote-13)]；此外，Verteletskaya引入了基于LPC谱的加权函数，并利用该函数对处理后的语音进行重塑，衰减人感知域内的噪声谱，在有效抑制残余音乐噪声的同时，还避免了语音失真[[[15]](#endnote-14),[[16]](#endnote-15)]。

综上所述，研究者们在处理基频恒定的电子喉语音时，采用上述多种谱减方法在一定程度上改善了电子喉语音的质量，但是对于近年来发展的基频可调电子喉语音（如Trutone 电子喉）的处理效果如何，并没有涉及。由于基频可调，辐射噪声的频率与强度必然不稳定，因此，本文将结合语音主观与客观评价方法，比较不同谱减技术对于变频电子喉语音的去噪效果。

# 1 方法

## 1.1 谱减法

谱减法的基本原理是从带噪语音谱中减去噪声估计谱，从而得到较为纯净的语音。假设语音和加性噪声是短时平稳的[[[17]](#endnote-16)]，令s(n)为纯净语音，d(n)为加性噪声，s(n)与d(n)统计不相关，则带噪语音y(n)可表示为：

(1)

用，和分别表示y(n)，s(n)和d(n)的傅里叶变换。假设s(n)与d(n)相互独立，则

(2)

即为估计的纯净语音谱，再利用原始带噪语音的相位，通过逆傅里叶变换即可得到估计的纯净语音信号。

上述方法即为经典谱减法，由于采用固定的谱减系数，即在整个频率段减去固定的噪声估计谱，当带噪语音帧某频谱噪声分量大于估计的噪声分量时，相减后会有部分噪声残留，在频谱上表现为随机出现的尖峰，听起来像有节奏的“音乐噪声”，降低了增强后的语音质量。为了抑制音乐噪声，Li等人根据电子喉辐射噪声的频谱在整个频率段是非均匀分布的特点，将整个频谱分为5个非叠加频带，即0~300Hz，300~1KHz，1K~2KHz，2K~3KHz，3K~5KHz，在每个频带根据信噪比动态调节谱减系数，以有效去除辐射噪声，减少噪声的残留。因此，第i个频带的增强语音谱的计算如下[12]：

(3)

其中，是第i个频带的过减系数，可以有效抑制音乐噪声，但是过大会引起语音失真，是控制过减系数大小的系数，为频谱最小值系数，使帧与帧之间的频谱更加平滑，从而抑制音乐噪声，和是第i个频带的起止频率和截止频率。过减因子是第i个频带的带噪语音信号信噪比的函数，而的取值依次为1.2，1.5，1.8，1.6，1.2，为0.001。

针对经典谱减法中谱减系数固定不变的问题，Liu等人在电子喉语音的去噪研究中将低速率编码器中的感知滤波器应用于谱减法中，即根据人耳掩蔽效应，利用感知滤波器的频响调节谱减系数，以有效去除辐射噪声。感知滤波器的计算如下[13]：

(4)

其中，A(z)为线性预测多项式，是短时线性预测系数，是控制共振峰区域的误差能量的系数，p是线性预测阶数，因此，感知加权滤波器及其频响均可由上式计算得到，本文中，p为15，，可通过带噪语音计算得到。

感知加权滤波器P(z)的频响为，谱减系数可基于自调节，以适应不同的语音环境。

此外，Verteletskaya提出引入一个加权函数，对谱减增强后语音频谱进行重塑，以实现噪声去除与语音不失真的最佳平衡。加权函数W(ω)可通过线性预测分析得到的频谱包络L(ω)推导得到[15]。如下所示：

(5)

通过多次实验发现，以带噪语音频谱最大值的5%作为阈值，取得了最佳去噪效果。指数参数γ=2，较大的γ值会使衰减后声音刺耳。

## 1.2 实验

在安静的环境下，让一名经过训练能够熟练使用电子喉的受试者（男）使用变频电子喉（基频的变化范围为60-150Hz）朗读给定的20句日常汉语普通话语句，在每次朗读前先开启电子喉1~2秒，作为录制语音的静音段以估计噪声，将麦克风放置于受试者嘴前10cm处，以48kHz的采样率录制电子喉语音原始数据。

分别采用经典谱减法、多带谱减法、基于感知加权技术的改进谱减法和基于加权函数的改进谱减法对电子喉语音数据进行去噪处理，窗函数采用汉明窗，窗长2048，帧移512，使用Praat软件对增强前后的语音进行波形图和语谱图分析以及基频的计算，并通过式(6)计算增强后语音和原始语音的能量比。能量比(Energy Ratio, ER)计算公式为：

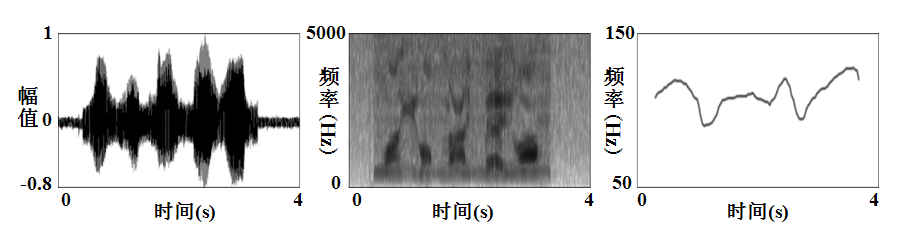
(6)

其中，x(n)为增强后电子喉语音，y(n)为原始电子喉语音

主观分析中，采用MOS（Mean Opinion Score）法对增强前后语音的质量进行评估。评价内容包括可懂度（听懂：1分，没有听懂：0分）和可接受度（优：5分，良：4分，中：3分，差：2分，劣：1分）两个方面，受试者为8位双耳具有正常听力且不熟悉电子喉语音的标准普通话说话人，测试语料为电子喉原始语音及不同谱减方法增强后的电子喉语音共100句，在安静环境中乱序播放，并要求受试者写下听懂的语句。

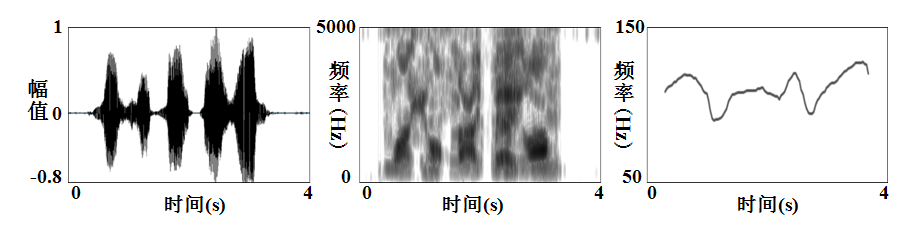
# 2 结果

## 2.1 语音信号分析



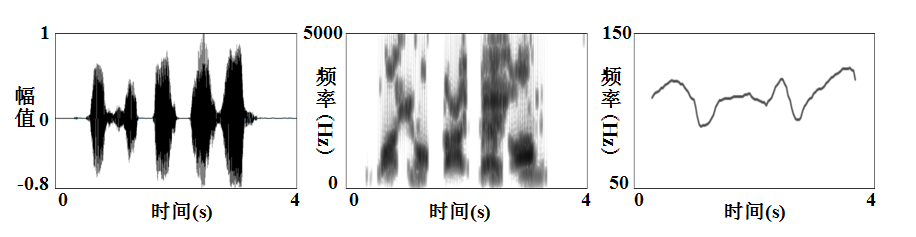
(a) 原始电子喉(ORIGINAL)语音

(a) Original EL speech



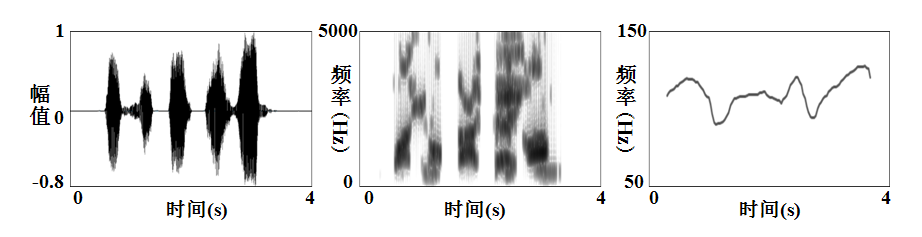
(b) 经典谱减法(Classical Spectral Subtraction, CSS)增强后电子喉语音

(b) Enhanced EL speech by CSS



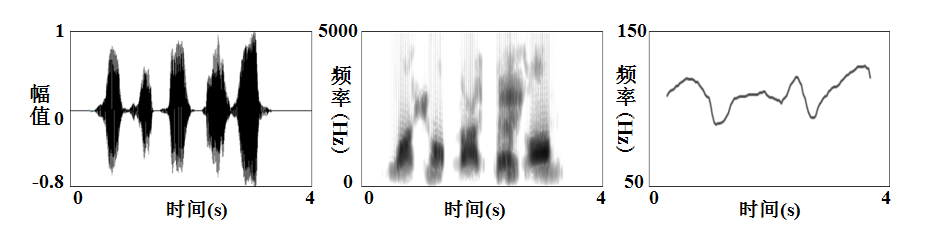
(c) 多带谱减法(Multi-Band Spectral Subtraction, MBSS)增强后电子喉语音

(c) Enhanced EL speech by MBSS



(d) 基于感知加权技术的改进谱减法(Perceptual Weighting Spectral Subtraction, PWSS)增强后电子喉语音

(d) Enhanced EL speech by PWSS



(e) 基于加权函数的改进谱减法(Weighting Function Spectral Subtraction, WFSS)增强后电子喉语音

(e) Enhanced EL speech by WFSS

图1 电子喉语句“欢迎来北航”的波形图、语谱图和基频图

Figure 1 Waveform, spectrograms and pitch of EL speech “huan ying lai beihang”

图1(a)中波形图和语谱图显示，原始电子喉语音存在明显的辐射噪声，尤其在语音间隔段，并且辐射噪声的能量存在于整个频率段。从图1(b)中波形图及图2的ER值可以发现，与原始电子喉语音相比，经CSS法增强后的语音，虽然有相当于原始电子喉语音能量7%的辐射噪声被去除，但是在静音段仍然有噪声残留，同样从语谱图也可以看到，在语音段除了正常语音成分外，仍有噪声残留。图1(c)、(d)、(e)的波形图表明，经WBSS法、PWSS法和WFSS法增强后的语音，辐射噪声都被明显去除，尤其在语音间隔段。不过，相比PWSS法和WFSS法，WBSS法增强后的语音在静音段仍有少量噪声残留。而从语谱图可以看出，MBSS法和PWSS法增强后的语音，在语音段仍然存在少量的噪声，而用WFSS法增强后的语音在语音段和间隔段，辐射噪声都被去除，不过在高频段有少量语音的损失，也使得WFSS法的ER值最低。基频图显示，不同谱减方法增强后的电子喉语音的基频较原始电子喉语音没有变化。

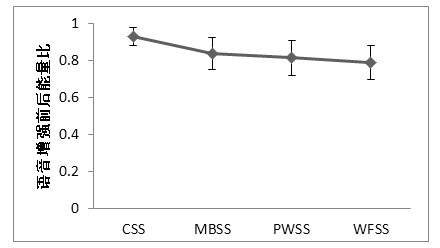


图2 增强后电子喉语音与原始电子喉语音的ER值

Figure 2 ER values of the enhanced and original EL speech

如图2所示，增强后的电子喉语音能量相比原始电子喉语音都有所降低。其中，CSS法增强后语音的ER值为0.93，能量减少的最少，而其他三种方法增强后的语音ER值比较接近，依次为0.84，0.82和0.79，WFSS法增强后的语音能量减少最多。

## 2.2 语音信号主观评价

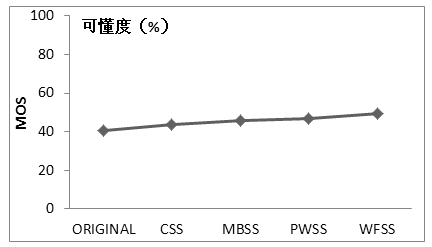
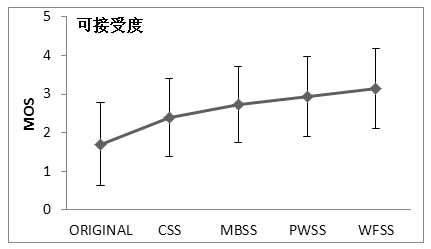


图3 原始电子喉语音及增强后电子喉语音的主观测试结果

Figure 3 Perceptual results of the original and enhanced EL speech

如图3所示，原始电子喉语音的可接受度和可懂度分别为1.701.08和40.63%，而CSS法、MBSS法、PWSS法及WFSS法增强后语音的可接受度和可懂度依次为2.391.01和43.75%，2.730.99和45.63%，2.941.03和46.87%，3.141.04和49.38%。

# 3 讨论

为探究不同谱减方法对变频电子喉语音的去噪效果，本文分别采用经典谱减法(CSS)、多带谱减法(MBSS)、基于感知加权技术的改进谱减法(PWSS)以及基于加权函数的改进谱减法(WFSS)对变频电子喉语音进行了去噪处理。波形图和语谱图的结果表明，MBSS法、PWSS法和WFSS法的去噪效果均优于CSS法，同样，在Li和Liu的实验中，针对非变频电子喉语音，CSS法的去噪程度也分别低于MBSS法和PWSS法，但增强后语音在语音间隔段的噪声要少于本文中CSS法去噪后的结果，这说明频率变化的辐射噪声相比非变频辐射噪声更难以去除。此外，基频图的结果表明，基于谱减法的语音增强方法不会影响电子喉语音的基频。.

增强后电子喉语音的ER值虽然不能代表语音的质量，但从客观上反映了噪声去除的程度。本文中，CSS法的ER值最高，而从语谱图可以看出CSS法去除的噪声最少，其他三种方法的ER值比较接近，其中，WFSS法的ER值最低，这与语谱图分析结果相一致。同样，在Li的实验中，在低信噪比和高信噪比两种噪声条件下，MBSS法的SNR值相比CSS法分别提高了6dB和12dB。在Verteletskay的实验中，在低信噪比和高信噪比的噪声条件下，WFSS法的SNR值相比CSS法分别提高了3dB和0.2dB。以上结果均反映了MBSS法和WFSS法的去噪效果优于CSS法。

主观评价结果中，各谱减方法增强后语音的可接受度有所提高，这是因为增强后的语音噪声减少，主观上让人听起来更加舒适。其中，CSS法得分最低，而MBSS法、PWSS法增强后的语音因存在少量的噪声，得分略低于WFSS法，与客观分析结果相一致。而各谱减方法增强后语音的可懂度比较接近，这可能是因为各谱减方法虽然降低了噪声，但是对语意影响不大。Li的实验中，原始电子喉语音由无喉患者使用基频固定电子喉录制，可接受度为2.1，而CSS法和MBSS法增强后的语音可接受度分别为2.5和3.25，均高于本文的实验结果，原因可能有两点，(1)变频电子喉辐射噪声变化不定，降低了CSS法和MBSS法的去噪效果。(2)无喉患者发出的电子喉语音质量优于正常人发出的电子喉语音质量，这是因为无喉患者的颈部组织特性优于正常人的，因此发出的电子喉语音中辐射噪声少于正常人的。Liu的实验中，由正常人使用基频固定电子喉发出的语音可接受度得分为1.25，而CSS法和PWSS法增强后的语音得分分别为1.6和2.05，均低于本文的实验结果，可能的原因为本文采用的变频电子喉基频可调，发出的电子喉语音更加自然，尤其是短词，提高了语音的可懂度，一定程度上影响了受试者对语音的主观评价。而可接受度有较大标准差，这主要是因为不同受试者对电子喉语音的熟练程度不同，再加上不同个体打分习惯差异较大，从而造成了同一语料同一内容的打分结果有的为1分，有的为4分。

综上所述，基于谱减法的语音增强方法可以有效减少变频电子喉语音的辐射噪声，提高语音的可接受度，对改善变频电子喉语音的听觉质量有较大帮助。

# 参考文献

1. 基金项目：国家科技支撑计划（2012BAI33B03）

   作者单位：北京航空航天大学生物与医学工程学院（北京 100191）

   作者简介：李阳（1989-），男， 硕士研究生

   通信作者：牛海军，教授 [↑](#footnote-ref-1)
2. [] Hanjun Liu, Mingxi Wan, Supin Wang, et al. Aerodynamic characteristics of laryngectomees breathing quietly and speaking with the electrolarynx[J]. Journal of Voice, 2004,18(4):567-577. [↑](#endnote-ref-1)
3. [] Ng M.L., Gilbert H.R., Lerman J.W.. Fundamental frequency, intensity and vowel duration characteristics related to perception of Cantonese a laryngeal speech[J]. Proquest Nursing. 2001, 53 (1): 201-204. [↑](#endnote-ref-2)
4. [] Isshiki N., Tanabe M. Acoustic and aerodynamic study of a superior electrolarynx speaker[J]. Folia Phoniatrica. 1972, 24(1): 65-76 [↑](#endnote-ref-3)
5. [] H. L. Barney, F. E. Haworth, H. K. Dunn. An experimental transistorized artificial larynx[J].Bell Syst. Tech, 1959, 38:1337-1356. [↑](#endnote-ref-4)
6. [] Weiss MS, Yeni-Komshian GH, Heinz JM. Acoustic and perceptual characteristics of speech produced with an electronic artificial larynx[J].Journal of the Acoustical Society of America, 1979, 65(5):1298-1308. [↑](#endnote-ref-5)
7. [] R. L. Norton, R. S. Bernstein. Improved Laboratory Prototype electrolarynx (LAPEL): using inverse filtering of frequency response function of the human throat[J]. Ann. Biomed. Eng.,1993, 21:163-174. [↑](#endnote-ref-6)
8. [] C. Y. Espy-Wilson, V. R. Chari, J. MacAuslan, et al. Enhancement of electrolaryngeal speech by adaptivefiltering[J]. Speech Lang. Hear. Res.,1998, 41: 1253-1264. [↑](#endnote-ref-7)
9. [] H. J. Niu, M. X. Wan, S. P. Wang, et al. Enhancement of electrolarynx speech using adaptive noise cancelling based on independent component analysis[J]. Medical & Biological Engineering & Computing,2003, 41:670-678. [↑](#endnote-ref-8)
10. [] Liu H., Zhao Q., Wan M., et al. Application of spectral subtraction method on enhancement of electrolarynx speech[J].Journal of the Acoustical Society of America,2006,120:398-406. [↑](#endnote-ref-9)
11. [] Pandey P. C., Bhandarkar, S. M., et al. Enhancement of alaryngeal speech using spectral subtraction[J]. Digital Signal Processing, 2002, 2:591-594. [↑](#endnote-ref-10)
12. []D. Cole, S. Sridharan, M. Moody, et al.Application of noise reduction techniques for alaryngeal speech enhancement[C].in Proc. IEEE TENCON, 1997, 2: 491-494. [↑](#endnote-ref-11)
13. [] S. Li, M. Wan, S. Wang. Multi-Band spectral subtraction method for electrolarynx speech enhancement[J]. Algorithms, 2009,2:550-564. [↑](#endnote-ref-12)
14. [] Liu H., Zhao Q., Wan, M., et al. Enhancement of electrolarynx speech based on auditory masking[J]. IEEE Transactions on Biomedical Engineering,2006, 53:865-874. [↑](#endnote-ref-13)
15. [] E.Verteletskaya, B. Simak. Noise reduction based on modified spectral subtraction method[J]. International Journal of Computer Science,2011,38:1-7. [↑](#endnote-ref-14)
16. [] E.Verteletskaya, B. Simak, Speech distortion minimized noise reduction algorithm[C]. Proceedings of the world congress on engineering and computer science, 2010, 1:581-584. [↑](#endnote-ref-15)
17. [] S. Boll. Suppression of acoustic noise in speech using spectral subtraction[J].Signal Processing,1979, 27:113-120. [↑](#endnote-ref-16)