基于跟踪共振峰的语音增强算法

杨　凌　　杨海波　　高新春
(兰州大学信息科学与工程学院　兰州　730000)

摘　要：该文通过实验方法研究和分析了汉字语音共振峰的特点，发现可跟踪并找到各个共振峰，结合汉字发音所具有的一般规律，提出了一种基于跟踪共振峰的语音增强算法。该算法能够有效地识别出带噪语音中的语音帧和非语音帧，简单且有效地去除非语音帧的全部噪声，明显抑制语音帧内的噪声。算法计算复杂度低并具有噪声环境可移植性。

关键词：语音信号处理；共振峰；共振峰区；语音帧
中图分类号：TN912.3　　**文献标识码**：A　　**文章编号**：1009-5896(2009)10-2536-05

Speech Enhancement Algorithm Based on Tracing the Formant

Yang Ling　　Yang Hai-bo　　Gao Xin-chun
(*School of Information Science and Engineering, Lanzhou University, Lanzhou* 730000, *China*)

Abstract: It is obtained that every formant can be traced and found out by studying and analyzing the Chinese speech in the way of experiment in this paper, combining with the generally articulation feature of the Chinese speech, a speech enhancement algorithm based on tracing the formant is proposed. The speech frames can be determined from the non-speech frames effectively using the algorithm, and then, all the noise in the non-speech frames can be wiped off simply and validly, and noise in the speech frame can be degraded largely. Also, the algorithm has some other good characteristics, such as simple computation and noise-environment transplantability and so on.

Key words: Speech signal processing; Formant; Formant area; Speech frame

1　引言

在诸多的语音通信领域，如：手机通信、网络电话、助听器、人机对话机器、语音识别、电话会议等，大量的需要噪声减少算法。1979 年，Boll 将谱减法应用到了数字信号领域[1]，从那时起，大量基于谱减法以及改进型谱减法的语音增强算法相继被提出[2−6]，谱减法虽然简单，但却引入了恼人的音乐噪声[7]。信号除躁的另一个重要方法是基于信号子空间分解的方法，它首先是由 Ephraim 和 Van Trees 于 1995 年提出的[8]，之后，Jabloun 和 Benoit 提出了语音增强的信号子空间技术[9]，近来，基于信号子空间原理和麦克风阵列的语音增强算法取得了一定的效果[10]，但此类算法的复杂度很大。谱减法和信号子空间分解法都是非参数的。1987 年，Paliwal 和 Basu 开启了参数类语音增强算法研究的先河[11]，该类算法近年来取得了相当的研究成果[12−14]，但它们都不能同时既改善语音的质量又提高语音的可懂度。基于小波和小波包分析方法的语音增强算法近年来发展很快[15−20]，并取得了良好的效果，然而由于该类算法中阈值的选取都是基于 Gauss 噪声模型，所以不具有噪声环境可移植性。

一般来说，上述算法要么简单但去噪效果差，要么运算复杂度大，要么不可能同时既改善语音的清晰度又提高语音的可懂度，要么不具有噪声环境可移植性。因此需要研究更为有效的语音去噪方法，以较小的运算量达到既能很好地去除噪声，又使语音几乎无扭曲的效果，并且还应具有噪声环境可移植性，本文介绍的算法在这方面取得了较好的效果。

2　共振峰的特点及跟踪的思想

2.1　共振峰的特点

表 1 列出了 6 个单元音(成年男性)的频谱特点和前 4 个共振峰的典型值。对于女性而言，共振峰值大约较男性高 25%，而小孩大约高 35% [21]。

由表 1 看出，各个共振峰分别占有一定的带宽；有的单元音有 4 个共振峰，考虑到第 4 个共振峰对语音的影响远不如第 1，第 2 和第 3 共振峰，所以

2008-10-17 收到，2009-04-09 改回
甘肃省自然科学基金(0803RJZA026)资助课题

表 1 6 个单元音的发音及频谱特点(前 4 个共振峰值)

韵母	典型字的韵母	收紧点	开口度	$F1$	$F2$	$F3$	$F4$	频谱特点
[a]	巴，大	后	大	850	1300	2600	3700	整体强度高，$F1$ 特别高
[o]	迫，魔	中	中	570	840	2400	–	开口由小到大，$F1$，$F2$ 皆由低到高
[e]	特，哥	中	中	520	1200	2400	–	开口由小到大，所以 $F1$ 由低到高
[i]	一，希	前	小	300	2300	3000	3500	$F2$ 弱，$F3$，$F4$ 近，故形成一个强区，频谱质心高
[u]	乌，路	后	小	350	650	2500	3300	$F1$，$F2$ 形成强区，$F3$，$F4$ 很弱，频谱质心低
[Ü]	玉，居	前	小	300	2000	2500	3500	圆唇音，$F2$，$F3$ 形成一个强区

表 2 元音[a]，[o]各谐波的峰值坐标及相邻两谐波之间的频差

第 1 共振峰谐波	谐波峰值坐标		相邻谐波	频差(Hz)		第 2 共振峰谐波	谐波峰值坐标		相邻谐波	频差(Hz)	
	[a]	[o]		[a]	[o]		[a]	[o]		[a]	[o]
$F11$	(74,23)	(74,29)	$F12$–$F11$	91.5	90.89	$F21$	(942,3)	(414,5)	$F22$–$F21$	117.12	125.05
$F12$	(224,8)	(223,10)	$F13$–$F12$	90.28	90.89	$F22$	(1134,9)	(619,11)	$F23$–$F22$	115.9	126.88
$F13$	(372,5)	(372,6)	$F14$–$F13$	90.28	90.28	$F23$	(1324,8)	(827,11)	$F24$–$F23$	115.9	127.49
$F14$	(520,2)	(520,2)	$F15$–$F14$	90.89	90.89	$F24$	(1514,7)	(1036,5)	$F25$–$F24$	111.02	124.44
$F15$	(669,4)	(669,5)	$F16$–$F15$	90.89		$F25$	(1696,9)	(1240,9)	$F26$–$F25$	115.9	
$F16$	(818,3)		$F1_{av}$	90.77	90.74	$F26$	(1886,11)		$F27$–$F26$	115.9	
						$F27$	(2076,5)		$F2_{av}$	115.29	125.97

本文只考虑元音的前 3 个共振峰。

我们通过大量实验进一步研究了 6 个单元音(成人男性)的频谱图。图 1 给出了元音[a]和[o]前两个共振峰的频谱图，文中语音抽样频率为 20 kHz。

表 2 列出了[a]和[o]各谐波的峰值坐标和相邻两谐波之间的频差值，其中，$F1_{av}$，$F2_{av}$ 分别为第 1，第 2 共振峰的平均基频。

由图 1，表 2 以及对其他单元音频谱图的分析，发现单元音具有如下特点：

(1)在第 1 共振峰区的频带内，基音频率基本保持不变。如表 2 所示，[a]和[o]的第 1 共振峰区的基频分别集中在 90.77 Hz 和 90.74 Hz 左右。

(2)在第 2 共振峰区的频带内，基音频率基本保持不变。如表 2 所示，[a]和[o]的第 2 共振峰区的基频分别集中在 115.29 Hz 和 125.97 Hz 左右。

(3)第 1 共振峰区的基音频率小于第 2 共振峰区的基音频率。

(4)有些元音各共振峰区之间有部分重叠。如图 1(b)所示元音[o]的频谱图清晰地体现了这一特点。

(5)不同元音的同一共振峰区在频谱上的位置和带宽是不一样的。对比图 1(a)，1(b)不难看出。

基频是人发声时声带的振动频率。由特点(1)和(2)知，第 1，第 2 共振峰各自的基频保持不变，说明在第 1 共振峰区，声带振动于一种模式(频率 $F1$)，在第 2 共振峰区，声带振动于另一种模式(频率 $F2$)。声带在发声前是静止的，正常发声时，声带先振动在较低的频率，后振动在较高的频率，所以由特点(3)可推知，首先出现第 1 共振峰区，接着迅速过渡到第 2 共振峰区，如此类推…由此得出：单元音各共振峰区的出现在时域上具有先后顺序性。

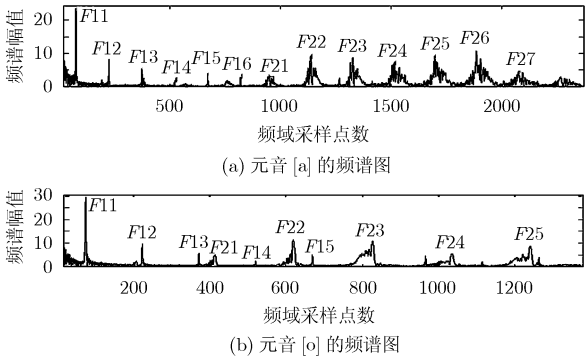


图 1 单元音[a]和[o]的频谱图

2.2 跟踪共振峰的思想

利用上述特性，可以在时域上对共振峰进行跟踪，基本思想是：在时域上将元音分成若干段，每一段对应一个共振峰区，这样，在每个语音段内，相应的共振峰区内的能量是最大的。

3 基于小波变换的语音段内噪声的去除

3.1 小波变换及语音频谱的分段

小波变换良好的时、频域局部特性，与语音信号的“短时平稳”特点正好吻合，适合做语音信号分析。根据对 Daubechies 小波时域能量集中性比较并适当考虑频率特性，本文选择 Db6 小波进行语音频谱分析。采用 Db6 小波 4 级变换后的语音频域分布如图 2 所示。

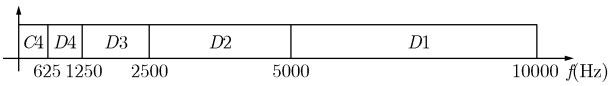


图 2 4 级小波变换后的语音频域分布图

由图 2 并结合以上关于共振峰的研究结果可以得出如下结论：

(1)与第 1 共振峰区有关的区域是：$C4$，$D4$。元音[a]，[o]，[e]的第 1 共振峰区涵盖了 $C4$ 和 $D4$，而[i]，[u]，[Ü] 3 个元音的第 1 共振峰区只包含 $C4$。

(2)与第 2 共振峰区有关的区域是：$C4$，$D4$，$D3$，$D2$，如表 3 所示。

表 3 6 个单元音第 2 共振峰所占的主要频谱区域

元音	[a]	[o]	[e]	[i]	[u]	[Ü]
频谱区域分布	$D4$, $D3$	$C4$, $D4$, $D3$	$D4$, $D3$	$D4$, $D3$, $D2$	$C4$, $D4$	$D4$, $D3$

(3)与第 3 共振峰区有关的区域是：$D3$，$D2$；

(4)与清音区有关的区域是：$D1$。

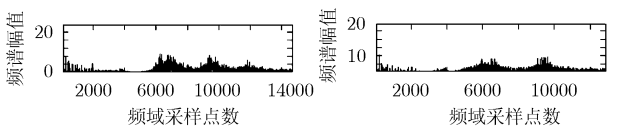
这样，在频域上，就把语音划分为了 4 个区域：第 1，第 2，第 3 共振峰区和清音区。

3.2 带噪语音信号语音段内的噪声去除

当跟踪到某一共振峰时，说明该共振峰区与当前正在处理的带噪语音信号序列相对应，于是保留跟踪到的共振峰区，并清除其他频带内的频谱成分。

4 带噪信号语音段与非语音段的区分原理

汉语普通话有一个很好的规律：绝大部分汉字的发音是由声母和韵母构成的。我们通过实验研究分析了声母(成人男性)的频谱特征，图 3 示出了声母[c]和[sh]在单独发声时的频谱图。



(a) 送气塞擦音 [c] 的频谱图　　(b) 清擦音 [sh] 的频谱图

图 3 声母[c]和[sh]的频谱图

由图 3(a)看出，声母[c]的频谱主要占据两大区域：0～1220 Hz (0～2000 采样区间，主要包含 $C4$) 和 3660～8750 Hz (6000～14000 采样区间，在 $D1$ 中)。由图 3(b)看出，声母[sh]占据的频谱区域与声母[c]的大致相同。对其他声母频谱的分析也可得到类似的特征。由此得出，声母在单独发声时，占据两个频区。由于声母在单独发声时，声带先是不振动的，而后才振动(产生第 1 共振峰)，所以，声母的两个语音区也满足在时域上的先后顺序性，而且在时间上相对滞后的一个语音区为第 1 共振峰区。

我们通过实验仔细分析了其他韵母(二合元音、三合元音及鼻韵母)发音的频谱特征，发现它们都具有和单元音相似的特征，即频域上具有类似共振峰的特点，所以把它们都当做一个准“元音”来处理。

当声母和韵母一起发声时，声母的第 1 共振峰区和韵母的第 1 共振峰区在时域上是相邻的。因此，在跟踪共振峰时，把它们当作一个共振峰来处理。从跟踪共振峰的角度来看，每个汉字的发音在时域上可分为若干段：声母发声的前半部分语音区+第 1 共振峰区+第 2 共振峰区+[第 3 共振峰区]。其中，[•] 表示若存在第 3 共振峰区，就加上；否则，不加。因此，判断一段带噪语音信号中是否含有纯净语音的依据是其是否具有上述的共振峰特点。

5 算法描述

基于跟踪共振峰的语音增强的主算法如下：

(1)建立一个大小为2×8 kb的缓冲区1(用于存放两段语音信号序列)和一个大小为 8 kb 的缓冲区 2；

(2)读入 128 个新的语音信号样本点，并放入缓冲区 2 中，和原先的数据一起组成新的序列；

(3)对缓冲区 2 中的数据进行小波变换，并找出能量变化最大的区域，该区域即为共振峰区；

(4)若该区域和加 128 个语音信号样本前的信号序列的能量变化区域相同，转(2)，否则，转(5)；

(5)除去缓冲区 2 中新加入的 128 个语音样本点；

(6)对缓冲区 2 中的数据进行小波变换，并去除相应的共振峰区外的噪声；

(7)进行小波逆变换；

(8)若当前语音信号序列对应于第 3 共振峰，且前一段语音信号序列为语音，则认为当前序列为语音，输出(7)中逆变换的结果，转(2)，否则，转(9)；

(9)若当前语音信号序列与第 2 共振峰相对应，且前一段语音信号序列为语音，则认为当前序列为语音，输出(7)中逆变换的结果，转(2)，否则，转(10)；

(10)若当前语音信号序列与第 2 共振峰相对应，且前一段语音信号序列与第 1 共振峰相对应，则认

为当前语音信号序列为语音，输出当前语音序列和前两段语音序列，转(2)，否则，转(11)；

(11)将缓冲区 1 中的前一段语音清零，并输出；

(12)缓冲区 1 中的数据向前平移，并将当前语音信号序列加入到缓冲区 1 中，转(2)。

6　实验仿真结果及评价

应用上述算法，对成人男性一段含噪语音(内容是“语音效果”)进行处理，实验结果如图 4 所示。其中，图 4(a)为纯净语音的时域图，图 4(b)，4(c)，4(d)，4(e)分别为不同信噪比下带噪语音增强前和增强后的时域图，其中噪声均为白 Gauss 噪声。

由图 4 可以看出，本文方法具有较好的去噪效果。从主观听觉上来讲，除了噪声被大幅度抑制外，几乎没有引入人为的噪声，且语音几乎无扭曲。

为了比较客观地评价本文算法的性能，我们还进行了以下两组实验，结果分别如表 4、表 5 所示。其中，信噪比(SNR)和分段信噪比(segSNR)的计算采用文献［22］的方法。

表 4 给出了在 Gauss 噪声环境下本文算法和谱减法的性能比较，可以看出，本文算法在取得良好去噪效果的同时，其运算速度堪与谱减法相媲美。

表 5 给出了本文算法在 Pink noise 和 F16 noise (来源于 Noise-92 噪声库)环境下的实验数据，可以看出，本文算法具有较好的噪声环境可移植性。

7　结论

本文在实验研究分析的基础上，发现了汉语普通话发音的一些特征，提出了一种基于跟踪共振峰的语音增强算法。实验结果表明，本文算法运算复杂度低，运行速度快，在有效提高信噪比的同时，可保证语音几乎无扭曲，且具有较好的噪声环境可移植性，从而有望在实时语音通信领域获得应用。

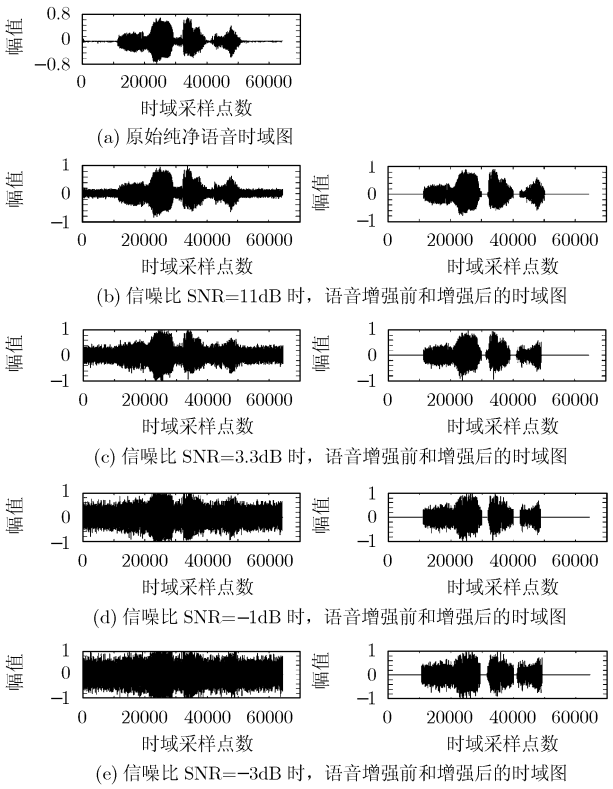


图 4　基于跟踪共振峰的语音增强算法的仿真实验结果

表 4 Gauss 噪声环境下本文方法和谱减法的性能比较

SNR			segSNR			CPU 耗时(s)	
SNR_{in}(dB)	SNR_{out}(dB)		$segSNR_{in}$(dB)	$segSNR_{out}$(dB)			
	谱减法	本文算法		谱减法	本文算法	谱减法	本文算法
11	17	19	2.5	9	10.5	0.167	0.180
3.3	10	15	−5	2.2	6	0.168	0.173
−1	7.2	12	−9.8	−1.8	3	0.166	0.170
−3	4.2	10	−11.8	−4	1.3	0.166	0.171

表 5 不同噪声环境下本文方法的性能比较

Pink					*F*16				
SNR		segSNR		CPU 耗时(s)	SNR		segSNR		CPU 耗时(s)
SNR_{in}(dB)	SNR_{out}(dB)	$segSNR_{in}$(dB)	$segSNR_{out}$(dB)		SNR_{in}(dB)	SNR_{out}(dB)	$segSNR_{in}$(dB)	$segSNR_{out}$(dB)	
13.9	20.4	1.4	4.9	0.182	11.8	15.8	−1.8	3.8	0.243
8.2	13.1	−4.6	2.8	0.165	9.8	13.4	−3.8	3.2	0.213
2.3	7.3	−10	0.1	0.143	3.9	11.5	−9.8	0.4	0.202
0.5	6.3	−12	−2.3	0.197	0.5	5.07	−13.3	−5.8	0.214

参 考 文 献

[1] Boll S F. Suppression of acoustic noise in speech using spectral subtraction[J]. *IEEE Transactions Acoustics Speech, Signal Processing*, 1979, 27(4): 113–120.

[2] Berouti M, Schwartz R, and Makhoul J. Enhancement of speech corrupted by acoustic noise[C]. Proc. of IEEE ICASSP, Washington, DC, USA, Apr 1979: 208–211.

[3] Lim J S and Oppenheim A V. Enhancement and bandwidth compression of noisy speech[J]. *Proc. of the IEEE*, 1979, 67(12): 1586–1604.

[4] McAulay R J and Malpass M L. Speech enhancement using a soft-decision noise suppression filter[J]. *IEEE Transactions Acoustics, Speech, Signal Processing*, 1980, 28(4): 137–145.

[5] Ephraim Y and Malah D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator[M]. *IEEE Transactions on Acoustics. Speech Signal Processing*, 1984, 32(12): 1109–1121.

[6] Wang Guang-yan, Wang Xia, and Zhao Xiao-qun. Speech enhancement based on a combined spectral subtraction with spectral estimation in various noise environment[C]. IEEE International Conference on Audio, Language and Image Processing, ICALIP-2008, Shanghai, China, July 2008: 1424–1429.

[7] Diethorn E J. Subband Noise Reduction Methods for Speech Enhancement[M]. Boston:Kluwer Academic Publishers, MA, 2004: 91–115.

[8] Ephraim Y and Van Trees H L. A signal subspace approach for speech enhancement[J]. *IEEE Transactions on Speech Audio Processing*, 1995, 3(7): 251–266.

[9] Benesty J, Makina S, and Chenk J. Speech Enhancement[M]. Berlin: Springer-Verlag, 2005: 1–8, 135–159.

[10] 张丽艳, 殷福亮. 一种改进的奇异值分解语音增强方法[J]. 电子与信息学报, 2008, 30(2): 357–361.
Zhang Li-yan and Yin Fu-liang. An improved speech enhancement method based on SVD. *Journal of Electronics & Information Technology*, 2008, 30(2): 357–361.

[11] Paliwal K K and Basu A. A speech enhancement method based on Kalman filtering[C]. Proc. IEEE ICASSP, 1987: 17–180.

[12] Li Hui, Wang Xin, Dai Bei-qian, and Lu Wei. A kalman smoothing algorithm for speech enhancement based on the properties of vocal tract varying slowly[C]. IEEE Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, networking, and Parallel/ Distributed Computing, SNPD-2007, Qingdao, China, July 2007, Vol.3: 832–836.

[13] You Chang-huai, Rahardja S, and Koh Soo Ngee. Autoregressive parameter estimation for Kalman filtering speech enhancement[C]. IEEE International conference on Acoustics,Speech and Signal Processing, ICASSP-2007, Honolulu, Hl, America, April 2007, Vol.4: IV-913–IV–916.

[14] Ruiz I, GarcÍa B, Méndez A, and Villanueva V. Oesophageal speech enhancement using Kalman filters[C]. IEEE International symposium on Signal Processing and Information Technology, Giza, Egypt, Dec. 2007: 1176–1179.

[15] Donoho D L. De-noising by soft thresholding[J]. *IEEE Transactions on Information Theory*, 1995, 41(3): 613–627.

[16] Shao Yu and Chang Chip-hong. A versatile speech enhancement system based on perceptual wavelet denoising[C]. IEEE International Symposium on Circuits and Systems, ISCAS-2005, Kobe, Japan, May 2005, Vol.2: 864–867.

[17] Xu Yao-hua, Wang Gang, Gu Ying, and Liu Hai-yang. A novel wavelet packet speech enhancement algorithm based on time-frequency threshold[C]. IEEE Second International Conference on Innovative Computing, Information and Control, ICICIC-2007, Kumamoto, Japan, September 2007: 492.

[18] Zhang L H and Rong G F. A kind of modified speech enhancement algorithm based on wavelet package transformation[C]. IEEE International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR-2008, Hong Kong, China, August 2008, Vol.1: 421–425.

[19] Hsung Tai-chui and Lun Pak-Kong D. Speech enhancement based on adaptive wavelet denoising on multitaper spectrum[C]. In IEEE International Symposium on Circuits and Systems, ISCAS-2008, Seattle, America, May 2008: 1700–1703.

[20] 徐耀华, 王刚, 郭英. 基于时频阈值的小波包语音增强算法[J]. 电子与信息学报, 2008, 30(6): 1363–1366.
Xu Yao-hua, Wang Gang, and Guo Ying. Wavelet package based speech enhancement algorithm using time-frequency threshold. *Journal of Electronics & Information Technology*, 2008, 30(6): 1363–1366.

[21] 杨行峻, 迟惠生. 语音信号数字处理[M]. 北京: 清华大学出版社, 1995: 13–20.

[22] Hangsen J and Pellom B. An effective quality evaluation protocol for speech enhancement algorithms[C]. Int. Conf. Spoken Language Processing, Sydney, Australia, Dec, 1998: 2819–2922.

杨 凌： 女，1966 年生，讲师，研究方向为信号与信息处理.

杨海波： 男，1983 年生，硕士生，研究方向为语音信号处理.

高新春： 女，1982 年生，硕士生，研究方向为通信数字信号处理.