

复倒谱域语音信号去混响研究

张德会, 陈光冶

(上海交通大学水声研究室, 上海 200030)

摘要: 语音信号去混响技术能明显提高语音通信系统和识别系统的性能。简要介绍了语音复倒谱法去混响原理, 对复倒谱域去混响滤波特性进行了仿真研究。根据多种去混响评价指标, 确定复倒谱域“低通滤波器”的最高截止点、过渡带宽和过渡带的曲线特性等参数, 发现在通常混响时间范围内, “低通滤波器”最高截止点与混响时间无关, 复倒谱域滤波前加高斯窗可以改善去混响效果。

关键词: 语音信号; 去混响; 复倒谱域滤波

中图分类号: TB556

文献标识码: A

文章编号: 1000-3630(2009)-01-0039-06

Speech signal dereverberation with cepstral processing

ZHANG De-hui, CHEN Guang-ye

(Underwater Acoustics Research Group, Shanghai Jiaotong University, Shanghai 200030, China)

Abstract: Speech signal dereverberation may greatly improve the performance of speech communication system or recognition system. The principle of dereverberation by cepstral processing is briefly introduced, and numerical simulation research is conducted to investigate filter properties in cepstral domain. By several dereverberation evaluation criterions, the cut-off points of the lowpass filter in complex cepstral domain are determined, and the transition bandwidth and transition section's curve characteristic are also fixed. It is shown that, the cut-off point of the lowpass filter is independent of the reverberation time in common reverberation time. It is also found that the dereverberation effect will improve if a Gauss window is used before complex cepstral filtering in time domain.

Key words: speech signal; dereverberation; complex cepstral filter

1 引言

在许多场合, 混响往往带来干扰, 导致声学接收系统性能变差。例如, 混响会导致语音识别系统性能显著下降, 在远程会议、振动声学、声纳通信和海洋声学中, 混响也主要带来负面影响^[1]。当混响严重时, 这些系统甚至无法正常发挥功能。因此, 如何减少混响对声音接收系统的影响是一个非常重要的课题。

针对在复倒谱域语音和房间冲击响应(RIR)容易分离的情况, 或者 RIR 的复倒谱为很突出峰值、便于检测的情况, 可以通过复倒谱滤波的方式或者去除尖峰值的办法减弱混响影响。在海洋声学中主动式声纳或鱼雷声制导系统可以将发射的波形设置为在复倒谱域中容易与海洋混响分离的形式, 便于海洋环境下去混响处理^[2]。现阶段国内外对于复倒谱去混响方法都有一定的研究, 主要涉及到两方

面: 一是对语音信号的预处理, 如对语音段加指数窗使信号具有最小相位特性^[3], 另一方面就是复倒谱域混响信号的低通滤波, 目前主要通过简单的低时窗进行信号截取。

本文简要介绍了声音复倒谱法去混响原理, 对复倒谱法去混响技术进行仿真研究。根据多种去混响指标, 确定了复倒谱域“低通滤波器”的最高截止点、过渡带宽和过渡带的曲线特性等参数, 并得出了一定混响时间范围内“低通滤波器”最高截止点与混响时间无关的结论。且实验显示, 复倒谱域滤波前加高斯窗能够改善去混响效果。

2 复倒谱域去混响原理及评价指标

在房间中不经过任何反射或散射直接到达接收点的声音叫直达声。而经过壁面一次或多次反射后到达接收点的声音叫做混响声, 是房间冲击响应与声源发出声音(激励)的卷积^[1,4]。去混响技术要从接收端减轻或消除混响声对原声音信号的影响。

在复倒谱域纯净语音信号的复倒谱通常分布在靠近原点的附近, 而房间冲击响应的复倒谱主要分布在远离原点的地方^[1,2]。目前一般采用低通滤波

收稿日期: 2008-02-21; 修回日期: 2008-05-31

作者简介: 张德会(1983-), 男, 江苏人, 硕士研究生, 研究方向为水声信号处理。

通讯作者: 张德会, E-mail: zdh013021@yahoo.cn

器的低时窗滤除相应的混响部分, 然后再通过逆向操作, 获得混响受到消减的声源信号。

本文针对语音和房间冲击响应(RIR)容易分离的情况进行讨论研究, 进一步探讨复倒谱域滤波的语音去混响方法。处理方法如图 1 所示, 其中 $x(n)$ 为混响信号, $s(n)$ 为纯净语音信号, $y(n)$ 为混响处理后的语音信号, 即 $s(n)$ 的估计。

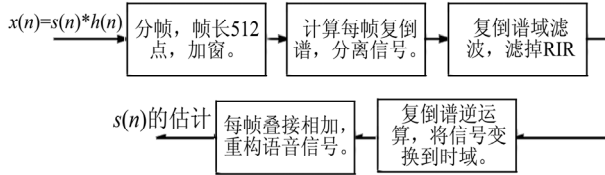


图 1 去混响系统流程框图

Fig.1 Flow frame of dereverberation system

目前评估去混响性能方法分为客观评测方法和主观评测方法, 本文引用时域信混噪比改善指标 $\Delta SRNR$, 线性预测倒谱系数距离改善指标 $\Delta LPCC$ 、MEL 频率倒谱系数距离改善指标 $\Delta MFCC$ 、语音的清晰度和自然度指标^[2,5,6]。其中 $\Delta LPCC$ 、 $\Delta MFCC$ 为两种频域客观评测系数, 是谱失真测度, 它们只与谱形状有关, 与谱能量无关, 本文均使用 12 阶倒谱系数, 原点的第一阶倒谱系数是反映频谱能量的, 没有计算在内。

2.1 信混噪比改善 $\Delta SRNR$ ^[2]

用 $SRNR_{in}$ 表示输入信混噪比、 $SRNR_{out}$ 表示输出信混噪比, 则它们以及 $\Delta SRNR$ 的表达式为(单位均为 dB):

$$SRNR_{in} = \frac{1}{N} 10 \lg \left[\frac{\sum_{n=1}^N s(n)^2}{\sum_{n=1}^N [x(n) - s(n)]^2} \right] \quad (1)$$

$$SRNR_{out} = \frac{1}{N} 10 \lg \left[\frac{\sum_{n=1}^N s(n)^2}{\sum_{n=1}^N [y(n) - s(n)]^2} \right] \quad (2)$$

$$\Delta SRNR = SRNR_{out} - SRNR_{in}$$

其中 $s[n]$ 、 $x[n]$ 、 $y[n]$ 分别是纯净语音信号、混响语音信号和输出的去混响信号, N 为每帧的样点个数。

2.2 线性预测倒谱系数距离改善 $\Delta LPCC$ ^[2]

用 $LPCC_{in}$ 、 $LPCC_{out}$ 表示输入、输出线性预测倒谱系数距离, 则它们及线性预测倒谱系数距离改善 $\Delta LPCC$ 的表达式为:

$$LPCC_{in} = \left\{ \sum_{n=1}^N [(c_n - c_n^x)^2] \right\}^{1/2} \quad (3)$$

$$LPCC_{out} = \left\{ \sum_{n=1}^N [(c_n - c_n^y)^2] \right\}^{1/2} \quad (4)$$

$$\Delta LPCC = LPCC_{in} - LPCC_{out} \quad (5)$$

其中 c_n 、 c_n^x 和 c_n^y 分别代表纯净语音信号 $s[n]$ 、

混响语音信号 $x[n]$ 和输出的去混响信号 $y[n]$ 的线性预测倒谱系数。

2.3 MEL 频率倒谱系数距离改善 $\Delta MFCC$ ^[2]

Mel 频率倒谱系数 (Mel Frequency Cepstral Coefficient, 简称 MFCC) 是一种基于短时傅立叶变换的谱包络参数表示。MFCC 分析不同于 LPCC 分析, 它不依赖于全极点语音产生模型的假定, 因而在混响和噪声环境下表现出更强的鲁棒性。本文求 MFCC 时使用 24 个带宽 256Mel 刻度, 间隔 128Mel 刻度的三角形滤波器。分别用式(6)~(8)计算输入、输出 Mel 频率倒谱系数距离和 Mel 频率倒谱系数距离改善 $\Delta MFCC$:

$$MFCC_{in} = \left\{ \sum_{n=1}^N [(m_n - m_n^x)^2] \right\}^{1/2} \quad (6)$$

$$MFCC_{out} = \left\{ \sum_{n=1}^N [(m_n - m_n^y)^2] \right\}^{1/2} \quad (7)$$

$$\Delta MFCC = MFCC_{in} - MFCC_{out} \quad (8)$$

其中 m_n 、 m_n^x 、 m_n^y 分别代表纯净语音 $s[n]$ 、混响语音 $x[n]$ 和去混响后的信号 $y[n]$ 的 Mel 频率倒谱系数。

2.4 清晰度和自然度, 主观评测方法

清晰度即去混响度, 即去掉混响语音信号中嗡嗡声的多少。自然度即语音的可懂度、失真度。本文采取用人耳听取仿真实验结果的方法来采集清晰度和自然度的情况, 其中清晰度和自然度各自分为四个等级: 优、良、中、差。

3 仿真研究

从语料库英文发音 1~10 中选定无混响单通道语音信号“six”作为研究对象, 时长 0.75s, 采样频率 10kHz, 采取 16 位 PCM 编码, 码速 160kbps。混响由音频软件产生, 混响时间 0.626s。

3.1 前处理的影响

3.1.1 确定分帧时的帧移

一般用短时能量的概念来描述语音信号的幅段, 将语音信号分成每 30ms 左右一段, 相邻两帧起始点之间间隔为 10ms, 即两帧之间有 10ms~20ms 的交叠。本文语音信号采样频率为 10kHz, 选定帧长为 512 点, 帧移为 256 和 128 情况下去混响各项指标异同如表 1 所示。

仿真时, 分帧后加汉明窗, 由表 1 可知, 当帧

表 1 帧移与去混响效果关系表

Table 1 The relationship between frame-shift and dereverberation effects

测评对象	客观评测			主观评测	
	$\Delta SRNR/dB$ (时域)	$\Delta LPCC/dB$ (频域)	$\Delta MFCC/dB$ (频域)	清晰度	自然度
帧移 256	0.5806	-0.6670	-8.4366	中	差
帧移 128	1.7941	0.1884	-0.5157	良	良

移为 128 点时,各项去混响指标均优于帧移 256 点,故帧移确定为 128 点。以下所做研究均在帧移 128 点的情况下进行。

3.1.2 确定为每帧加窗的窗函数^[7]

信号的截断会产生信息泄漏和波动等,加窗可对这些影响进行抑制。对语音信号进行分帧后,每帧均加窗函数,现比较加不同窗函数去混响效果,比较结果见表 2。

表 2 加窗与去混响效果关系表

Table 2 The relationship between windows and dereverberation effects

窗类型	客观评测			主观评测	
	$\Delta SRNR/dB$ (时域)	$\Delta LPCC/dB$ (频域)	$\Delta MFCC/dB$ (频域)	清晰度	自然度
汉明窗	1.7941	0.1884	-0.5157	中	良
汉宁窗	2.0871	0.2531	-0.5920	中	中
矩形窗	0.5166	0.0145	-5.5111	差	差
三角窗	1.9098	-0.0228	-1.5810	差	差
布莱克曼窗	2.1013	0.4275	1.0031	中	中
凯泽窗	0.6046	0.0522	-6.7394	差	差
高斯窗	1.7136	0.1840	1.5010	中	良

由表 2 可知,在同一去混响环境下,汉明窗、汉宁窗、布莱克曼窗和高斯窗的去混响效果较好,而其中高斯窗的各项指标最为均衡和优异,故针对高斯窗再予以深入研究。变换高斯窗的系数 a , 得其去混响各项指标,见表 3。以表 3 数据作出如图 2 所示曲线。

由表 3 和图 2 可以看出,在高斯窗系数为 2.8 左右的时候,去混响效果最好。因此选定分帧后的窗函数为高斯窗,系数为 2.8。

3.2 复倒谱域滤波器设计

该滤波器是一种广义上的低通滤波器,它是借用频域中低通滤波器的定义,在复倒谱域进行设计的一种特殊的低通滤波器。在复倒谱域,相应形成的向量空间称为“倒频域”,其横坐标从本质上是经过数学变换的时间轴^[2]。由于进行分帧处理,连续的帧与帧之间互有重叠,语音信号变换到复倒谱之后头尾部分均含有语音信号(见图 3),房间冲击响应混杂在远离头尾的部分,故需设定合理的滤波

器进行首尾的低通滤波,此复倒谱域的低通滤波器由三部分组成,分别为通带、过渡带和阻带,其基本构造如图 4 所示。

其中 L 为滤波器最高截止点, h 为过渡带宽,暂定为 $h=64$ 。

表 3 高斯窗系数与去混响效果关系表

Table 3 The relationship between Gauss window's coefficient and dereverberation effects

a	客观评测			主观评测	
	$\Delta SRNR/dB$ (时域)	$\Delta LPCC/dB$ (频域)	$\Delta MFCC/dB$ (频域)	清晰度	自然度
2.1	1.6796	0.2241	-0.3734	差	良
2.2	1.6981	0.0834	-1.3216	差	中
2.3	1.7607	0.1421	0.2514	中	中
2.4	1.7881	0.1835	1.5270	中	中
2.5	1.7136	0.1840	1.5010	中	中
2.6	2.0637	0.1935	-0.6237	差	中
2.7	2.0634	0.1829	-0.2151	中	中
2.8	2.1045	0.3913	2.2149	良	良
2.9	2.0768	0.3227	1.5198	良	良
3.0	1.9847	0.3125	1.8188	良	良

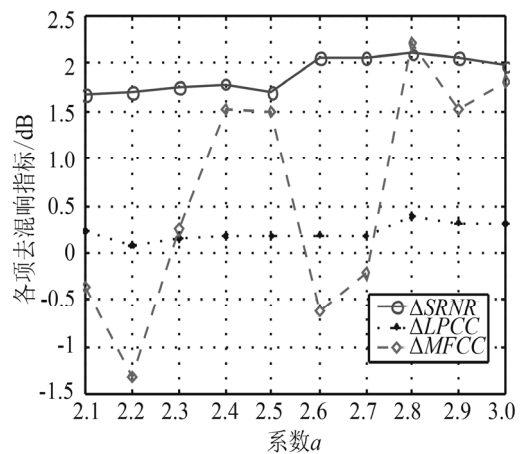


图 2 高斯窗不同系数与去混响效果关系图

Fig.2 The relationship between gauss window's coefficient and dereverberation effects

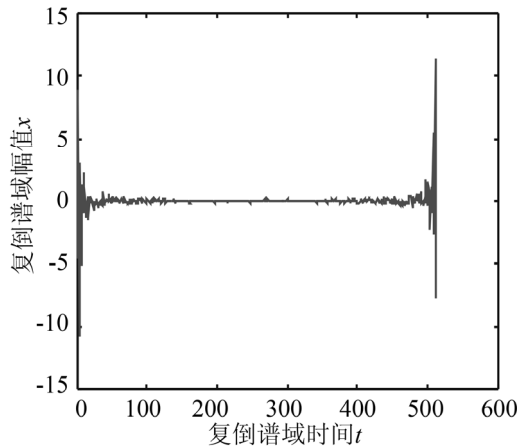


图 3 复倒谱滤波第二帧复倒谱域波形

Fig.3 Waveform of the second frame in complex cepstral fields

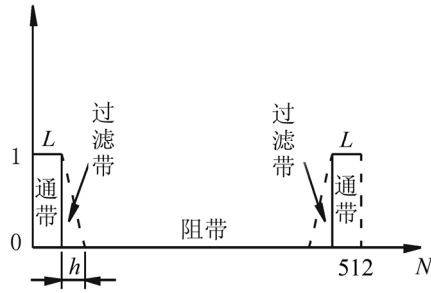


图 4 复倒谱域低通滤波器原理图
Fig.4 Principle of the lowpass in complex cepstral fields

3.2.1 确定最高截止点 L

表 4 为最高截止点与去混响效果关系，由表 4 作出曲线如图 5 所示。

表 4 最高截止点与去混响效果关系表
Table 4 The relationship between the cut-off point and dereverberation effects

L/N	客观评测方法			主观评测方法	
	$\Delta SRNR/dB$ (时域)	$\Delta LPCC/dB$ (频域)	$\Delta MFCC/dB$ (频域)	清晰度	自然度
1/512	1.8235	-0.1518	8.3738	差	差
2/512	1.8517	0.0297	8.2146	中	中
4/512	1.9753	0.0887	6.0412	良	良
8/512	1.9819	0.1283	1.3484	中	中
16/512	1.8312	0.2125	-3.4925	差	差
32/512	1.6735	-0.1418	-9.0476	差	差
64/512	1.7695	-0.0311	-12.3853	差	差
128/512	1.2849	0.1881	-0.9294	差	差

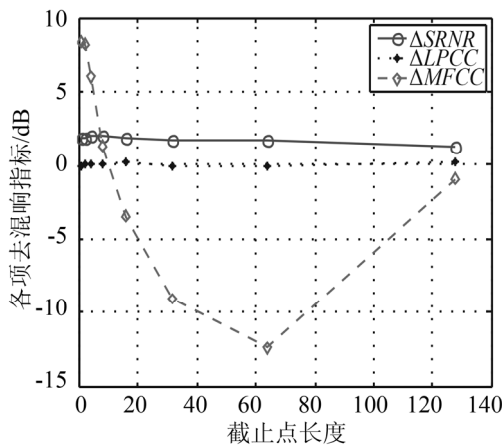


图 5 滤波器不同截止点去混响效果图
Fig.5 The relationship between the cut-off point and dereverberation effects

由表 4 和图 5 可知，最高截止点越低，去混响效果越好，但也不能太低。 $L=4$ 时，去混响效果最好，即最高截止点 L 与帧长 N 的比值 L/N 为 $1/128$ 时，去混响效果最好。

3.2.2 过渡带的优化

选定截止点 $L=4$ ，进行过渡带的优化。优化方法为从线性过渡依次到高阶曲线过渡。具体结果见表

表 5 过渡带阶数与去混响效果关系表
Table 5 The relationship between the transition curve's steps and dereverberation effects

阶数 n	客观评测			主观评测	
	$\Delta SRNR/dB$ (时域)	$\Delta LPCC/dB$ (频域)	$\Delta MFCC/dB$ (频域)	清晰度	自然度
1	1.9895	0.2763	4.8072	良	良
1.5	1.9753	0.0887	6.0412	良	良
2	1.9451	-0.1830	7.2400	中	中
3	1.8862	-0.5809	5.3248	中	中
4	1.9147	-0.8457	2.5515	差	差
5	1.9880	-1.0873	1.3788	差	差
6	2.0387	-1.2957	0.7734	差	差
7	2.0463	-1.3579	1.3926	差	差
8	2.0234	-1.4164	-0.0346	差	差
9	1.9899	-1.4782	-1.5398	差	差
10	1.9595	-1.5595	-2.1340	差	差

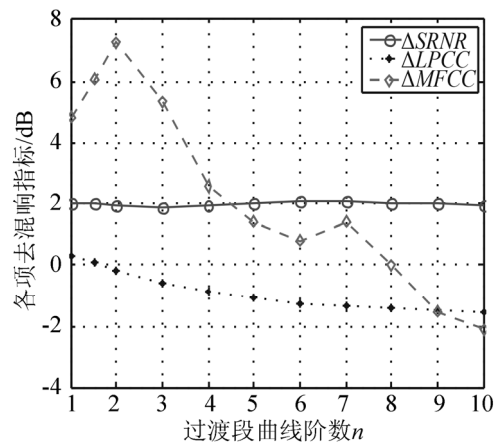


图 6 过渡段曲线阶数去混响效果图
Fig.6 The relationship between the transition curve's steps and dereverberation effects

5，由表 5 作出曲线图 6 所示。

通过表 5 和图 6 可以看出，在过渡段阶数小于 2 时，无论是从客观评测方法还是主观听觉来看，去混响效果都要优于高阶过渡曲线，因此过渡段阶数的选取应尽量选在线性阶数附近，为了达到更好效果，现进行进一步的研究，具体结果如表 6 所示，由表 6 数据作出图 7 所示曲线。

由表 6 和图 7 可以看出，当 $n>1$ 时，虽然 $\Delta MFCC$ 仍呈上升趋势，但 $\Delta SRNR$ 、 $\Delta LPCC$ 均呈下降趋势，综合考虑，当过渡段阶数处于 0.7~1.5 时，去混响各项指标比较好。

3.2.3 过渡带宽的优化

过渡带宽的大小对整个低通滤波器的去混响性能有着重要的影响，现在最高截止点 L 为 4，过渡段为线性曲线的前提下研究过渡段带宽对低通滤波器性能的影响。分别设定滤波器过渡带宽为 0、2、4、8、16、32、64、128，过渡带宽与去混响效

表 6 过渡带阶数与去混响效果关系表(进一步研究)

Table 6 The relationship between the transition curve's steps and deverbation effects(Further research)

阶数 n	客观评测方法			主观评测	
	$\Delta SRNR$ /dB (时域)	$\Delta LPCC$ /dB (频域)	$\Delta MFCC$ /dB (频域)	清晰度	自然度
0.3	1.8349	0.1452	-4.0449	差	差
0.4	1.8890	0.2348	-2.8865	差	差
0.5	1.9307	0.2928	-1.2293	差	差
0.6	1.9592	0.3325	0.4500	中	中
0.7	1.9765	0.3454	1.9703	中	中
0.8	1.9855	0.3330	3.3074	良	良
0.9	1.9891	0.3075	4.2142	良	良
1.0	1.9895	0.2763	4.8072	良	良
1.1	1.9881	0.2423	5.1829	良	良
1.2	1.9858	0.2098	5.4356	良	良
1.3	1.9828	0.1704	5.6636	良	良
1.4	1.9794	0.1316	5.8715	良	良
1.5	1.9753	0.0887	6.0412	良	良
1.6	1.9706	0.0404	6.2064	中	中
1.7	1.9651	-0.0130	6.4210	差	差
1.8	1.9590	-0.0697	6.6799	差	差
1.9	1.9523	-0.1272	6.9643	差	差

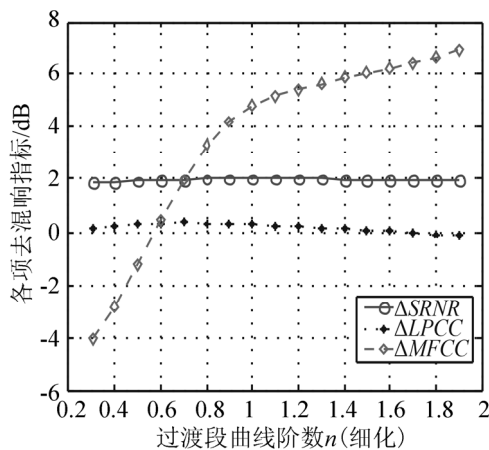


图 7 过渡段曲线与去混响指标的关系(进一步研究)

Fig.7 The relationship between the transition curve's steps and deverbation effects(Further research)

果关系见表 7, 由表 7 作出曲线图 8。由表 7 和图 8 可见, 当过渡带宽为 64, 即过渡带宽 h 与帧长 N 的比值 h/N 为 1/8 时, 去混响效果最好。

3.2.4 混响时间与滤波器最高截止点的关系

基于以上研究, 探讨混响时间与滤波器最高截止点的关系。分别对纯净语音信号“six”施加混响, 设定混响时间为 900ms、750ms、512ms、400ms、256ms, 利用已经得出的滤波器的最佳参数验证混响时间与滤波器最高截止点的关系, 得到以下表 8 中数据。

由表 8 中数据可得, 在一定混响时间范围内, 最佳最高截止点的确定与混响时间无关。

表 7 过渡带宽与去混响效果关系表

Table 7 The relationship between the transition bandwidth and deverbation effects

过渡带宽 h	客观评测			主观评测	
	$\Delta SRNR$ /dB (时域)	$\Delta LPCC$ /dB (频域)	$\Delta MFCC$ /dB (频域)	清晰度	自然度
0	-31.5196	-8.3428	-91.6149	差	差
2	-12.7548	-5.8257	-69.0932	差	差
4	1.0715	-2.8522	-36.3369	差	差
8	1.9218	-1.9674	-8.2340	差	差
16	2.0355	-1.5466	-0.7638	差	差
32	1.9192	-0.3949	4.6847	中	中
64	1.9895	0.2763	4.8072	良	良
128	2.0823	0.2436	6.1389	差	优

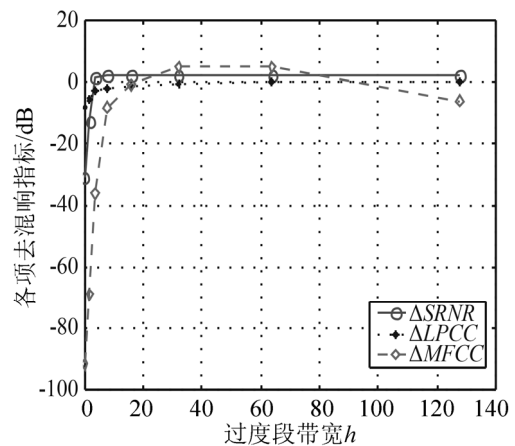


图 8 过渡带宽与各项去混响指标的关系

Fig.8 The relationship between the transition bandwidth and deverbation effects

表 8 混响时间与去最佳截止点关系表

Table 8 The relationship between reverberation time and the cut-off points

混响时间 T /ms	900	750	512	400	256
最佳截止点	4	4	4	4	4

针对以上去混响系统的预处理手段和复倒谱域低通滤波器的最佳参数, 对语音库中纯净语音“one~ten”, 进行重复研究, 发现研究结果在此范围内具有通用性。现给出语音信号为“six”时的去混响前后波形图, 如图 9 所示。

4 总结

本文对复倒谱去混响方法进行了进一步的研究, 确定了复倒谱域低通滤波器的基本参数, 得到以下结论:

- (1) 经仿真比较, 当帧移为 128 点, 即帧移为 13ms 左右, 帧长的 1/4 时, 去混响效果变优。在同一去混响环境下, 分帧后加汉明窗、汉宁窗、布莱克曼窗和高斯窗的去混响效果较好, 而其中高斯窗

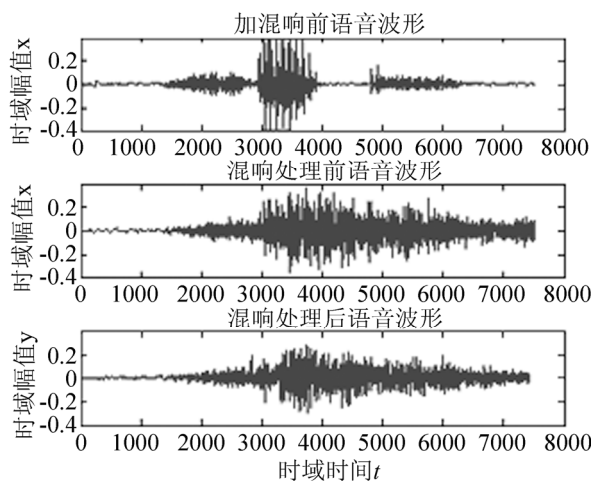


图9 复倒谱法去混响波形比较图

Fig.9 Wave forms compare in complex cepstral fields

的各项指标最为均衡和优异,且高斯窗的系数为2.8时,效果最好。

(2) 在复倒谱域低通滤波器的设计中,提出了这种滤波器的基本原理图,并对滤波器的各项参数进行了优化设计,得出滤波器的最佳最高截止点 L 为4,即与帧长 N 的比值为 $1/128$,过渡带的带宽 h 为64,与帧长 N 的比值为 $1/8$,过渡带优化阶数为0.7到1.5之间。

(3) 经过仿真验证,得出复倒谱域低通滤波器最高截止点与混响时间无关。利用其它语音信号对去混响系统进行验证,证明在研究范围内具有通用性。

此去混响系统有待改进的地方在于单纯的复

倒谱法去混响的效果不是太明显,对复倒谱法继续优化和与其他方法结合值得进一步研究,以改善去混响效果。

参 考 文 献

- [1] 吴佳栋, 陈光冶. 语音信号去混响原理与技术[J]. 电声技术, 2006, (05): 63-67.
WU Jiadong, CHEN Guangye. Principle and technologies of speech signal dereverberation[J]. Audio Engineering, 2006, (05): 63-67.
- [2] 苏先礼, 何培宇. 语音去混响研究[D]. 成都四川大学, 2006, 14-28.
SU Xianli, HE Peiyu. The research on speech dereverberation[D]. Chengdu: Sichuan University, 2006, 14-28.
- [3] Bees D, Blosein M, Kabal P. Reverberant speech enhancement using cepstral processing[A]. IEEE ICASSP-91[C]. Toronto, Canada: 1991, 977-980.
- [4] 杜功焕, 朱哲民, 龚秀芬. 声学基础[M]. 南京: 南京大学出版社, 2001, 417-432.
DU Gonghuan, ZHU Zhemin, GONG Xiufeng. Acoustics elements[M]. Nanjing: Nanjing University Press, 2001, 417-432.
- [5] 易克初, 田斌, 付强. 语音信号处理[M]. 北京: 国防工业出版社, 2000, 83-118.
YI Kechu, TIAN Bin, FU Qiang. Speech signal processing[M]. Beijing: National Defence Industry Press, 2000, 83-118.
- [6] L. R. Rabiner(拉宾纳), R. W. schaffer(谢弗), 朱雪龙等译. 语音信号数字处理[M]. 北京: 科学出版社, 1983, 272-347.
Rabiner L R, Schaffer R W, ZHU Xuelong. Digital processing of speech signal[M]. Beijing: Science Press, 1983, 272-347.
- [7] 陈怀琛. 数字信号处理教程——MATLAB 释义与实现[M]. 北京: 电子工业出版社, 2004, 220-225.
CHEN Huicheng. Digital signal processing tutorial-MATLAB paraphrase and realization[M]. Beijing: Publishing House of Electronics Industry, 2004, 220-225.