一种新的基于倒谱的共振峰频率检测算法

赵 毅[1†] 尹雪飞[1] 陈克安[2]
（1 西北工业大学电子信息学院 西安 710129）
（2 西北工业大学航海学院 西安 710072）

摘要 共振峰频率是语音信号的一个重要参数。传统的基于线性预测的共振峰检测算法由于受到计算量的限制，很难实现实时处理。本文提出一种基于倒谱变换的共振峰频率检测算法，采用后置处理，比较声道冲击响应对数幅频特性的二次导数和相频特性一次导数检测出的结果，删除伪峰数值和甄别合并共振峰，提高检测精度。仿真结果证明，该算法计算效率高，低信噪比下仍能保持较好的检测性能。

关键词 共振峰，倒谱，对数幅频特性，相频特性

A new formant detection algorithm based on cepstrum

ZHAO Yi[1] YIN Xue-Fei[1] CHEN Ke-An[2]
(1 *School of Electronics and Information, Northwestern Polytechnical University, Xi'an* 710129)
(2 *School of Marine Technology, Northwestern Polytechnical University, Xi'an* 710072)

Abstract Formant is one of important parameters for speech signal. Due to huge computational load, the traditional speech formant detected by linear prediction coding can not be used for real-time implementation. In this paper, a new formant detection algorithm based on cepstrum transform is presented, which compares the second derivative of the logarithmic amplitude-frequency characteristics with the first ones of the phase-frequency characteristics, deleting the false formants, and discriminating the merging of the formants. The simulations show that the proposed algorithm is of high efficiency, and can still keep good performances under low signal-to-noise ratio conditions.

Key words Formant, Cepstrum, Logarithmic amplitude-frequency characteristics, Phase-frequency characteristics

1 引言

语音信号处理中，共振峰频率是一个特别重要的参数。它是表征发音时声道特性的关键参数，也是区别不同韵母的重要依据[1]。寻求一种可实时实现的共振峰频率检测算法，在语音合成、语音识别、说话人识别等领域中有重要的应用价值。共振峰信息包含在语音频谱包络中[1]。因此，共振峰频率检测的关键在于估计自然语音频谱包络，并认为谱包络最大值对应的频率就是共振峰频率。在此基础上，人们发展了两类较为常用

2009-05-17 收稿; 2009-09-25 定稿

作者简介：赵毅 (1985-)，男，济南市人，硕士研究生，研究方向：语音信号处理。
尹雪飞 (1968-)，女，副教授，博士，硕士生导师。 陈克安 (1965-)，男，教授，博士，博士生导师。
†通讯作者：赵毅，E-mail:q19851007@163.com

的共振峰频率提取算法：基于功率谱的峰值提取法和基于线性预测（linear prediction coding，简称为LPC）的求根法[2-4]。前者认为共振峰频率在频域以功率谱极大值点的形式出现，只需在功率谱中检测出极大值点所对应的频率及可确定共振峰频率。后者用线性预测对语音信号进行解卷积，得到声道响应的全极点模型，并通过牛顿-拉夫逊（Newton-Raphson）方法确定线性预测系数，进一步得出共振峰中心频率和3 dB带宽[1]。这两类算法各有不足之处：峰值提取法会受到合并共振峰和虚假共振峰（伪峰）的影响，检测结果不精确；求根法只适用于所有根都为共轭复根的情况[2]，而且由于线性预测算法收敛速度慢，运算量大，难以快速准确地找到根，从而不适用于实时实现的场合。目前，共振峰检测这一领域的主流算法是线性预测法[4]。

本文提出一种新的基于复倒谱变换的共振峰频率快速检测算法，从语音信号中分离出声道冲击响应函数，经一系列变换后采用峰值检测提取共振峰频率，避免了因线性预测带来的计算量大的问题，便于实时实现；并且增加后置处理，筛选检测到的峰值数据，降低了伪峰频率出现的可能性，并对共振峰合并有一定的分辨效果，提高了检测精度。仿真结果证明，该算法可以快速准确地检测出语音模型和实际语音的共振峰频率。

2　语音信号倒谱分析

已有研究表明，如果直接在语音频谱提取共振峰频率，误差会很大。倒谱分析技术可以较好地分离出语音信号频谱包络结构，从而为共振峰频率的检测开辟一条新途径。倒谱分析在检测语音信号基音频率中应用较为成熟[1,5]，在共振峰频率检测中应用不多。2004年Thomas F. Quatieri系统阐述了同态分析（倒谱分析是其中的一个典型方法）的具体原理[6]，王晓亚提出了倒谱分析在共振峰检测中的应用前景[5]，2005年赵力等人也阐述了倒谱分析应用于共振峰检测领域的基本原理和步骤[1]。但是，倒谱分析在共振峰检测中的应用存在若干缺陷，使其无法代替传统的LPC方法。本文首先建立语音模型，通过设置模型的声门激励函数和声道冲击响应函数，模拟不同的语音效果，然后说明本文算法采用倒谱分析的意义以及改进措施。

2.1 语音模型的建立

人类的发音过程总体可分为3个阶段，即声门激励、声道调制和口唇辐射。在此基础上，对语音进行建模，可以认为语音模型是由声门激励模型，声道模型以及口唇辐射模型3者级联所产生的[1,7]（如图1所示）。

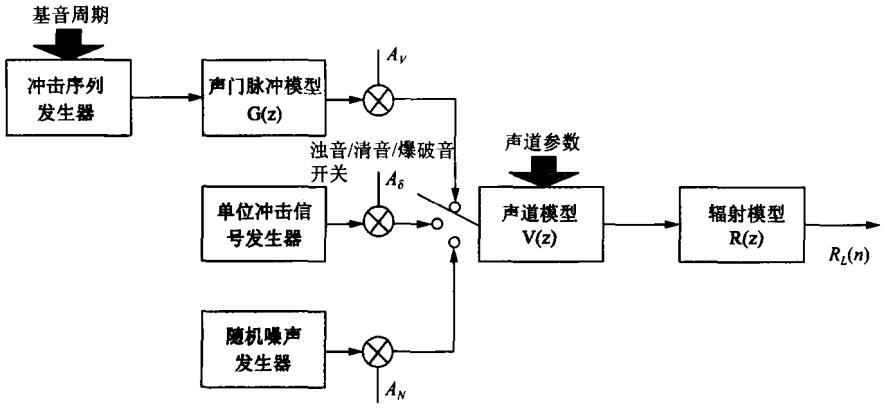


图1　语音产生模型

2.1.1 声门激励模型

不同的声门激励模型可产生不同的典型语音，如浊音、清音和爆破音[6,7]。采用脉冲序列调制声门波函数（声门波函数是一个类余弦函数，主要参数是声门打开和关闭时间长度比，基音周期和占空比）即可模拟浊音信号的声门激励，用随机噪声模拟清音的声门激励，用单位冲激函数模仿爆破音的声门激励。本文所用声门波函数的数学表达式为：

$$g(n)=\begin{cases}(1/2)[1-\cos(\pi n/N_1)] & 0\le n<N_1\\ \cos[\pi(n-N_1)/2N_2] & N_1\le n\le N_1+N_2\\ 0 & else\end{cases} \tag{1}$$

其中，N_1表示声门激励上升时间，N_2表示声门激励下降时间。

2.1.2 声道模型

声道模型可看作一个稳定的最小相位系统[7]，其数学表达式为：

$$V(z)=\frac{G}{1-\sum_{k=1}^{N}a_k z^{-1}} \tag{2}$$

式中，N是极点个数，G是幅度因子，a_k是常系数。基于线性预测的求根法就是通过求解一系列常系数a_k来确定声道模型，该模型对于一般的元音有很好的模拟效果。

如果把声道看成是一个共鸣器，当准周期脉冲激励进入声道时会引起共振特性，产生一组共振频率，这组共振频率就是共振峰频率或简称共振峰[1]。清音和爆破音是没有明显的共振峰特性的。在建立声道模型的时候，如考虑共振峰参数的影响就得到声道的共振峰模型。其数学表达式如下：

$$V(z)=\frac{G}{\prod(1-z_k z^{-k})} \tag{3}$$

式中：

$$z_k=r_k\exp(j\theta_k) \tag{4}$$

$$\theta_k=2\pi TF_k \tag{5}$$

$$r_k=\frac{1}{\exp(\pi TB_k)} \tag{6}$$

其中，r_k是极点的幅度信息，θ_k是极点的相位信息，T是采样周期，F_k是共振峰频率，B_k是共振峰3 dB带宽[8,9]。我们采用声道模型的一对共轭极点来表示一个共振峰，一般取10～12阶的声道模型即可满足要求。

2.1.3 口唇辐射模型

口唇辐射模型表征口唇的辐射效应，也包括圆形的头部的绕射效应等[1]。辐射效应可以用一个一阶差分方程来近似的描述，它的数学形式如下：

$$R(z)=R_0(1-z^{-1}) \tag{7}$$

其中R_0是一个常量系数。嘴唇辐射影响引起的输出信号高频提升作用大约为每倍频程6 dB[5]。

将以上3个模型级联，就可以模拟语音产生的全过程。

2.2 复倒谱分析的作用

复倒谱是一种同态处理技术，它通过对输入信号求Z变换，取对数，求逆Z变换3步运算将呈卷积关系的两个信号转化为叠加的关系[10]。前文介绍的语音模型是由声门激励模型和声道模型卷积而成的，声道模型在频域表征语音频谱的包络结构，共振峰结构就包含在声道模型中。经过复倒谱分析之后，

原本呈卷积关系的声门激励模型和声道模型在倒频域呈加性关系，声道冲击响应向零点靠近，声门激励则平铺在整个倒频域，只需设计合适的低频倒谱窗便可将两者分离，得到语音信号频谱包络信息，通过后值算法提取出共振峰频率。

3　共振峰提取的后置处理

3.1 待筛选数据的检测

郁伯康等人[9]的研究成果表明，声道冲击响应对数幅频特性的二次导数与相频特性的三次导数，相对于对数幅频特性本身，均具有较高的频率分辨率，可以提高共振峰频率检测的精度。本文通过大量蒙特卡洛仿真得出，相频特性三次导数的频率分辨率比对数幅频特性二次导数的频率分辨率略高，但是出现伪峰值的概率也高于后者。基于在峰值检测的高精度和错误率之间做出折中的考虑，本文采用对数幅频响应二次导数检测到的极小值点频率，作为共振峰频率点的待筛选数据。本文使用峰值检测算法，在共振峰频率集中的较低频段（一般集中在0～6000 Hz），按照峰值点能量的大小，从高到低依次提取10个极小值频率点作为待筛选数据，供后置处理使用。一般情况下，对数幅频响应二次导数出现的极小值点可能不会达到10个（尤其在信噪比较高的情况下），本文建立一个包含10个元素的向量来存储这些极小值频率点，称为初始向量。之所以选取较大的向量长度，是为了不漏掉任何一个可能是共振峰的频率点（当然也有可能是伪峰）。

3.2 后置处理

伪峰频率点和共振峰合并是共振峰检测中最为突出的两个问题[1,11]，目前还没有一种算法可以彻底解决这两个问题。本文针对上述问题，提出以下后置处理算法。

3.2.1 对伪峰数据的处理

伪峰的出现主要有两个原因，首先是混杂在语音信号中的噪声的影响，其次是声门激励函数对声道冲击响应函数的干扰。针对上述原因，我们通过以下3个方面的处理来筛选数据，删除伪峰。

首先，求取声道冲击响应函数相频特性的一次导数，按照极值点能量的高低，从大到小依次检测极小值点，并将能量较大的点对应的频率存入一向量，我们称之为“标尺”向量。这样做是因为，声道冲击响应函数相频特性的一次导数频率分辨率低于对数幅频响应的二次导数，出现伪峰值的概率较小，将该特性中检测到的极小值点作为“标尺”，并设定误差阈值，然后将初始向量中与“标尺”向量元素误差大于阈值的数据删除，即可初步去除伪峰数值。我们将“标尺”向量的元素个数控制在5～6个（即将相频响应一次导数中能量最大5～6个极小值点对应频率存入“标尺”向量），避免引入伪峰值。

其次，自适应控制低通倒谱窗函数的长度，减小声门激励函数的干扰并保持较高的频率分辨率，选择合适的窗函数点数是本文算法的一个关键。若该窗函数取的点过少，则无法提取出完整的声道冲击响应，破坏了语音信号共振峰特征，且后期处理频率分辨率降低；若取的点过多，则引入了过多声门激励的影响，经过一系列的变换后会产生伪峰值，同样影响检测的准确性。本文首先取较多点数的窗长（一般在20点以上），来保证较高的频率分辨率，然后通过检测声道冲击响应相频特性一次导数的极小值频率点个数来控制窗长，若极小值点个数过多，则减小窗长，然后继续检测，如此往复，直到声道冲击函数相频响应的一次导数中极小值点个数控制在合理范围内（一般取7～8个认为合理，本文取8个），然后按照峰值能量由大到

小的顺序依次确定“标尺”元素，认为此时“标尺”中已经没有伪峰数据。低通倒谱窗一般取矩形窗，本文选用一种类正弦函数窗（如式8所示）

$$window(n)=\begin{cases}|\sin(\pi n/n_0)| & |n|<n_0\\ 0 & |n|\geq n_0\end{cases} \quad (8)$$

作为低通倒谱窗，式中n表示窗函数样点，n_0表示窗长。仿真结果证明使用该窗函数的效果优于矩形窗。

最后，通过事先设定的误差阈值，对初始向量中的元素逐一排查，若“标尺”中没有元素与它“对应”，则认为该数值是伪峰，予以删除。这里的“对应”是指“标尺”中的任何一个元素与数据的欧氏距离均大于误差阈值。

3.2.2 对合并共振峰的处理

经过复倒谱分析，我们提取出信号频谱的包络信息，但是声道冲击响应变得平滑，共振峰带宽被展宽，这就使得共振峰合并问题更加突出（如图2所示）。本文提出算法对于该类问题的处理分以下两个方面：

首先是选取声道冲击响应对数幅频特性的二次导数来代替声道冲击响应幅频特性提取共振峰频率的待筛选数据（即初始向量），这样可以大大提高频率分辨精度，有效避免了共振峰合并现象。

其次，本文在后置处理中引入了频率分辨率较低的声道冲击响应相频特性一次导数这一特性。在该特性中可能存在共振峰合并现象，这样会出现以下状况（如图3所示），即待筛选数据中不存在共振峰合并，但是由于“标尺”中存在因共振峰合并而出现的坏点，使得正确的共振峰频率在排查时由于没有“标尺”元素与其“对应”而被当作伪峰值删除。本文的解决方案是，在待筛选数据与“标尺”数据进行比对时，采用不同的误差阈值。若初始向量中任意两个元素欧氏距离均大于某一设定值，则认为没有共振峰合并，在比对时设置较小的误差阈值；若初始向量中有元素欧氏距离太近，则在比对时选择较大的误差阈值。这样将误差阈值扩大，可以检测出相距比较近的共振峰频率。

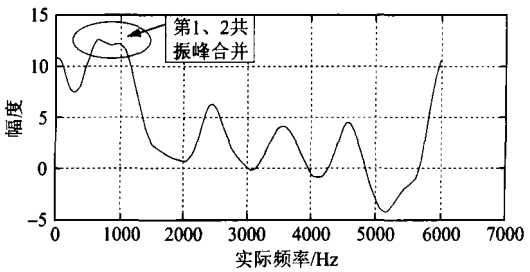


图2 共振峰合并现象

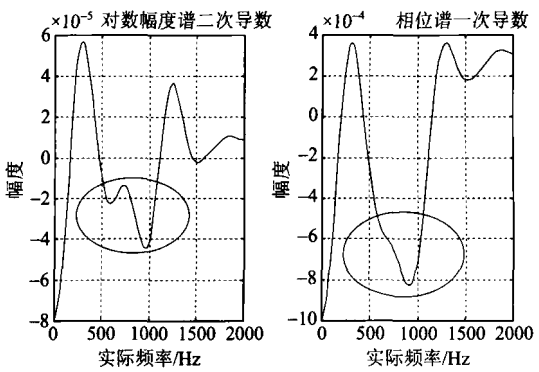


图3 共振峰合并在相频特性一次导数中的体现

3.3 算法流程

本文算法的流程如图4所示。原始语音信号输入以后，首先进行分帧和预加重处理，并求复倒谱，利用低通倒谱窗分离出语音信号的声道冲击响应函数。然后通过计算声道冲击响应对数幅频特性的二次导数和相频特性的一次导数并检测极小值点，得出初始向量和“标尺”向量，在每一帧的处理过程中，都将根据声道相频特性一次导数中极小值点的个数来调整低通倒谱窗的长度。最后，利用后置处理算法，去除伪峰，甄别合并峰，得到最终的共振峰频率。

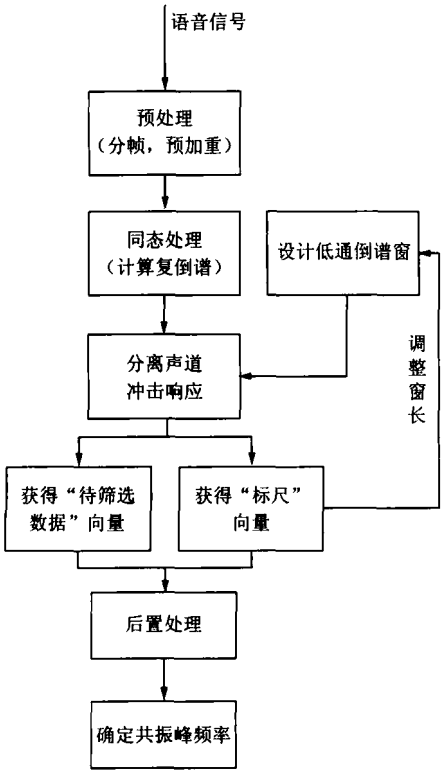


图4　算法流程

4　仿真及结果分析

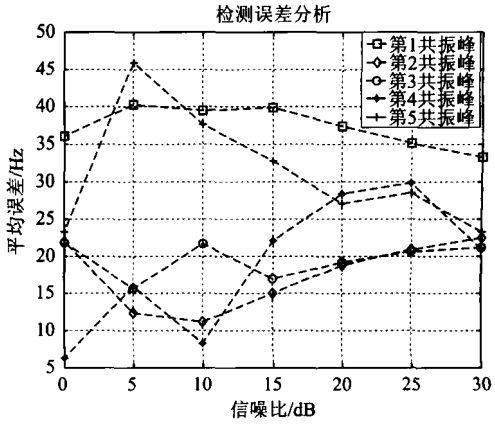
通过仿真对所提出算法的性能进行了验证。仿真包括两部分，首先采用合成的语音模型进行检测，然后采用实际的元音音素进行检测。

4.1 合成语音模型仿真结果

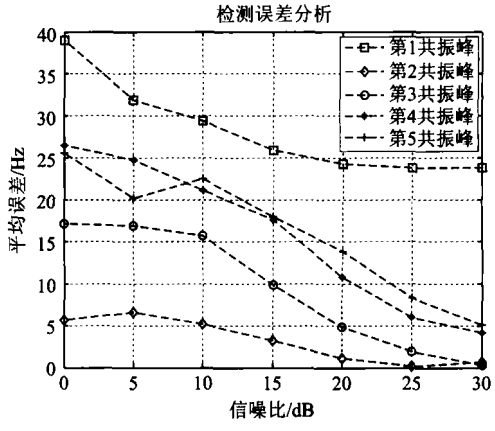
在不同信噪比下各做500次蒙特卡洛仿真，将检测到的共振峰频率平均值与设定的初始值相比较，并计算出每个共振峰频率检测出的次数在500次蒙特卡洛仿真中所占的百分比。两组仿真数据分别是人工合成的音素/a/和Portonff基于声道管壁振动、粘滞损耗和热损耗而建立的语音模型[7]。通过式（3）～（6）将已经设定好的共振峰频率和3 dB带宽合成语音声道模型，并与声门激励模型和口唇辐射模型（式（7））级联，形成仿真所用的纯净语音模型。本文仿真中，设定采样频率12000 Hz，频率范围0～6000 Hz，设定的共振峰频率如表1所示，信噪比从30 dB到0 dB，每隔5 dB检测一组数据，并统计检测数据的均值和错检率，两组仿真的误差和错检率分析如图5～6所示。

表1　设定共振峰频率

人工合成/a/音素					
共振峰次序	1	2	3	4	5
共振峰频率 (Hz)	650.3	1075.7	2463.1	3558.3	4631.3
Portonff语音模型					
共振峰次序	1	2	3	4	5
共振峰频率 (Hz)	502.5	1508.9	2511.2	3513.5	4518.0

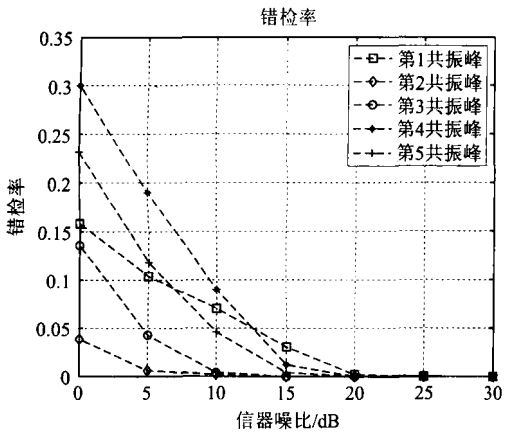


(a) Portonff语音模型

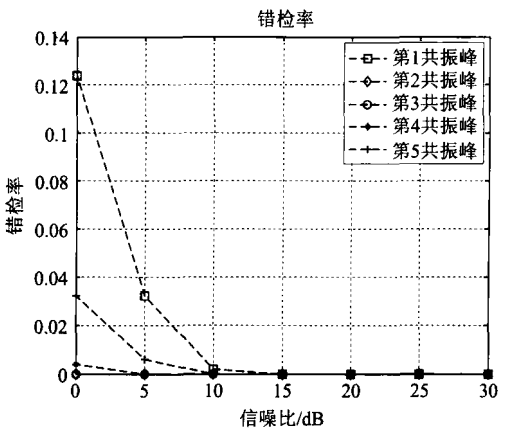


(b) 共振峰检测误差曲线

图5　合成音素/a/



(a) Portonff语音模型



(b) 共振峰检测错检率曲线

图6 合成音素/a/

从仿真结果可以看出，对于合成语音，该算法可以比较准确的提取出语音信号前五个共振峰。图5(a)和(b)中的5条曲线分别表示人工合成音素/a/和Portonff语音模型的5共振峰频率随信噪比不断降低，检测误差逐渐增大。图6(a)和(b)中的5条曲线分别表示人工合成音素/a/和Portonff语音模型的5共振峰频率随信噪比不断降低，错检率逐渐上升。对于合成音素/a/，该算法对前3个共振峰频率的错检率，在信噪比0 db时仍然可以控制在15.0%左右，第4共振峰的错检率最高，为30%。虽然第1、2共振峰靠的比较近，但本文算法仍然可以将两者分开。对于Portonff语音模型，0 db信噪比情况下，前3个共振峰的错检率控制在12.4%以内，在信噪比大于10 dB时，错检率几乎为零，基本可以满足实际需要。

4.2 实际语音仿真结果

我们采用实际的元音音素/ʌ/ /ai/ /æ/来验证本文算法的有效性。首先用语谱图方法读出以上3个音素的前5个共振峰频率，然后分别用传统的LPC法和本文算法进行检测，并将比较结果以表2～4的形式给出。另外，分析了两种算法的检测误差和检出率（见图7～8）。原始的语音文件采样率为44100 Hz，本文降采样为10000 Hz，最高频率为5000 Hz[12]，用256点汉明窗对语音分帧，每帧帧长256点，

表2 实际语音仿真结果/ʌ/（共7帧）

共振峰序号	1	2	3	4	5
语谱图(Hz)	591.2	1248.3	2357.2	3379.2	4633.4
LPC法(Hz)	597.7	1191.0	2358.1	*	*
检出帧数	7	7	7	0	0
本文方法(Hz)	578.7	1258.7	2381.0	3350.7	4677.3
检出帧数	7	7	7	7	3

*表示未检出，下同

表3 实际语音仿真结果/ai/（共11帧）

共振峰序号	1	2	3	4	5
语谱图(Hz)	412.0	1107.9	3241.0	4056.5	4671.1
LPC法(Hz)	394.0	1169.7	3269.2	4053.6	*
检出帧数	10	11	11	10	0
本文方法(Hz)	379.4	1147.9	3231.2	4037.5	4719.0
检出帧数	11	11	11	10	1

表4 实际语音仿真结果/æ/（共12帧）

共振峰序号	1	2	3	4	5
语谱图(Hz)	653.2	2032.9	2710.9	3714.9	4547.0
LPC法(Hz)	650.3	2025.6	2694.3	3755.3	*
检出帧数	12	12	12	12	0
本文方法(Hz)	670.2	1999.0	2743.9	3761.5	4577
检出帧数	12	12	12	12	3

帧间重叠128点，每段语料的处理帧数见各表右上角。LPC法采用Durbin递推算法，设递推阶数 $p=14$ 。

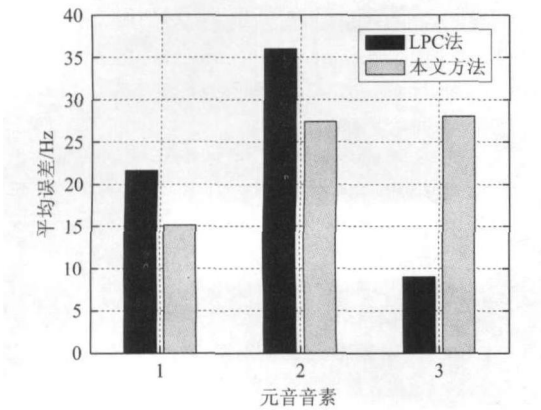


图7　平均误差分析

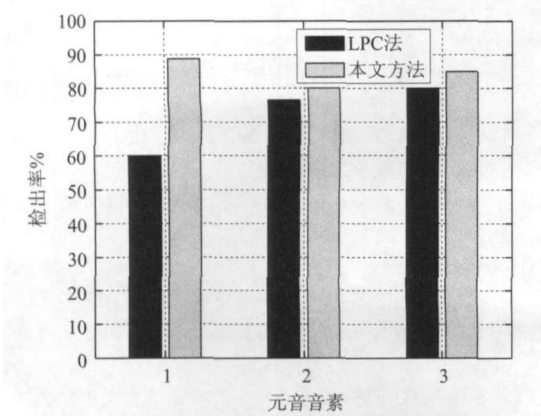


图8　检出率分析

从仿真数据（表2～4）可以看出，传统的LPC法对语音前3个共振峰的检测精度较高，但对第4、第5共振峰的检测结果不理想。我们将两种方法对实际语料前3个共振峰的检测结果的平均误差用图7表示，可以看出，对于前两段语料，本文所提出的算法检测误差小于LPC法，第3段预料本文算法的检测误差大于LPC法。综合3段语料的检测结果，两种算法的精度大体相当。然而，从检测出的共振峰数目来看，LPC法检测到11个共振峰频率（第一组数据，LPC法只能检测到前3个共振峰），而本文算法可以检测到全部15个共振峰，检出率明显高于LPC法（如图8所示）。总之，传统的LPC法受线性预测系数个数的影响，很难检测出语音全部5个共振峰。而本文算法直接从语音频谱分离包络结构，最大限度的保留了原始语音频谱的峰值信息，在计算量不变的情况下检测出语音信号全部5个共振峰，并保持和LPC法相同的检测精度，更加适合实时处理的领域。

4.3 运算量分析

以本文4.2元音音素/ʌ/为例，取其中第一帧做计算量分析。对于帧长为N的预料，LPC法在求信号自相关函数时进行 $p\times N$ 次实数乘法运算，N次实数加法运算，然后为确定自相关函数的递归初始值进行向量与其自身转置的乘法运算，需要 N^2 次实数乘法运算，最后从 $i=2$ 开始进行 $p-1$ 次递归，总共需要 $\frac{3p^2}{2}$ 次实数乘法运算，$p+\frac{p^2}{2}$ 次实数加法运算。最终的乘加次数为：$(p+N)N+\frac{3p^2}{2}$ 次实数乘法，$N+p+\frac{p^2}{2}$ 次实数加法。本文算法最大的计算量开销是两次FFT变换和一次IFFT变换，FFT变换需要 $\left(\frac{N}{2}\right)\log_2^N$ 次复数乘法，$N\log_2^N$ 次复数加法[13]。因为一次复数乘法相当于4次实数乘法，两次实数加法，一次复数加法相当于2次实数加法，所以总共需要 $\frac{9}{2}N\log_2^N$ 次实数乘法运算，$9N\log_2^N$ 次实数加法运算。计算机中做乘法运算是通过加法运算来完成的，因此乘法运算的系统开销远大于加法运算。传统LPC法的乘法次数取决于 $(p+N)N$ ，本文算法的乘法次数取决于 $N\log_2^N$ 。其中 $N=256$ ， $p=14$ ，因此可知本文算法计算量要小于传统的LPC法。4.2中3段语音信号使用传统LPC法和本文算法

进行共振峰频率检测的仿真时间对比如表5所示。该表中列出两种方法做共振峰检测时，利用MATLAB软件的计时函数计算处理每一帧语音信号的平均时间，单位是“毫秒(ms)”，由该表可以看出本文算法处理每帧数据比LPC法约节省0.002 ms，在计算时间上优于LPC法。

表5 计算时间对比

元音/ʌ/	LPC法	0.023728
	本文方法	0.020909
元音/ai/	LPC法	0.023212
	本文方法	0.020388
元音/æ/	LPC法	0.021066
	本文方法	0.019395

5 总结

本文针对传统的基于LPC的共振峰频率检测算法存在的各种不足，提出了一种基于复倒谱变换的共振峰频率检测算法，并引入长度可变的低通倒谱窗和后置处理方法，该方法在一定程度上克服了倒谱方法用于共振峰频率检测的固有缺陷。最终的仿真结果证明了该算法在实际语音的共振峰频率检测中的有效性。

参 考 文 献

[1] 赵力. 语音信号处理, 北京: 机械工业出版社, 2005: 76-80.

[2] 何峰, 陈晓清, 李国锁, 等. 一种新的语音信号共振峰提取算法, 信号处理, 2007, 23(4): 618-621.

[3] Lutz Welling, Hermann Ney. Formant estimation for speech recognition. IEEE Transactions on Speech and Audio Processing, 1998, 6(1): 36-48.

[4] Stephanie S. Mccandless. An algorithm for automatic formant extraction using linear prediction spectra. IEEE Transactions on Acoustics, Speech and Processing. 1974, 22(2): 135-141.

[5] 王晓亚. 倒谱在语音基音和共振峰提取中的应用, 无线电工程, 2004, 34(1): 57-61.

[6] Thomas F. Quatieri（著), 赵胜辉, 谢湘, 刘家康, 等(译). 离散时间语音信号处理, 北京: 电子工业出版社, 2004: 197-238.

[7] 易克初, 田斌, 付强. 语音信号处理, 北京: 国防工业出版社, 2001: 31-35.

[8] Roy C. Snell, Fausto Milinazzo. Formant location from LPC analysis data. IEEE Transactions on Speech and Audio Processing, 1993, 1(2): 129-134.

[9] 郁伯康, 郁梅. LPC方法提取语音信号共振峰的分析, 电声技术, 2003, 3(1): 3-8.

[10] http://qsliu.spaces.live.com/blog/cns!e620e6807a69b58b!152.entry.

[11] 刘建新, 曹荣, 赵鹤鸣. 一种 LPC 改进算法在提取耳语音共振峰中的应用, 西华大学学报, 2008, 27(3): 77-80.

[12] 胡广书. 数字信号处理——理论、算法与实现, 第 2 版, 北京: 清华大学出版社, 1998: 115-119.

[13] 程佩青. 数字信号处理教程, 第 2 版, 北京: 清华大学出版社, 2001: 144-150.