

小波分析在语音信号分析中的应用

PB17000123 张湛

2019 年 7 月 18 日

摘要

语言信息主要包含在语音信号的参数之中，因此准确而迅速地提取语言信号的参数是进行语音信号处理的关键。常用的语音信号参数有：共振峰幅度、频率与带宽、音调和噪音、噪音的判别等。后来又提出了线性预测系数、声道反射系数和倒谱参数等参数。

20 世纪 80 年代之后，研究语音信号非平稳参数分析方法迅速发展，人们提出了一整套快速的算法，还有利用优化规律实现以合成信号统计分析参数的新算法，取得了很好的效果。其中，有些便是基于小波和小波包变换而得出的算法。如：语音信号在其频域中的分析技术，小波变换和 EMD 分解在端点检测中的应用，以及小波-自相关函数法在带噪音的语音中进行的基音检测。

本文通过在 matlab 中进行实验，探寻小波分析在语音信号分析与合成中的应用，以此探究进一步理解小波变换在实际中的应用，更深入的感悟小波分析在应用中带来的方便。具体学习和实践了利用窗口傅里叶变换绘制语音信号的语谱图 (Spectrogram.m)，通过小波变换提取简单的基因频率 (Peaks.m)，用小波包对语音在 Bark 域上分解并压缩后重构 (Wavlet Barkms Decomposition.m)，还有有关端点检测的小波分解平均幅值法 (WT_VAD.m) 与小波包的 BARK 子带分离频带方差法 (WPB_VAD.m)，这一系列的编程实践与理论的学习都让我体会到了小波分析在信号处理中独特的地位。

目录

1	语音的基本概念	4
1.1	语音信号的数字模型	4
1.1.1	激励模型	4
1.1.2	声道模型	5
1.1.3	辐射模型	6
1.2	语音信号的时域、频域特性	6
1.2.1	短时傅里叶变换	7
1.2.2	语谱图	7
2	语音信号在小波变换域中的分析技术和特性	8
2.1	利用小波变换提取基音信息	8
2.2	用小波包分解设计 Bark 滤波器	8
2.2.1	Bark 域介绍	8
2.2.2	用小波包分解设计 Bark 滤波器	9
3	小波变换在端点检测中的应用	10
3.1	语音端点检测的常用方法	10
3.2	小波分解系数的平均幅值法进行端点检测	11
3.3	小波包 BARK 子带方差的端点检测	11
	参考文献	13

1 语音的基本概念

语音是人类发出的声音，它由一连串的音信号组成，而各个音的排列是符合一些经验的规律的。语音中最小的基本单位是音素，它分为元音和辅音两类，是人类区别一个单词和另一个单词的基础。在音信号中，还有一个重要的特点是其具有特定的声调，它是一种音节在念法上的高低升降变化。

1.1 语音信号的数字模型

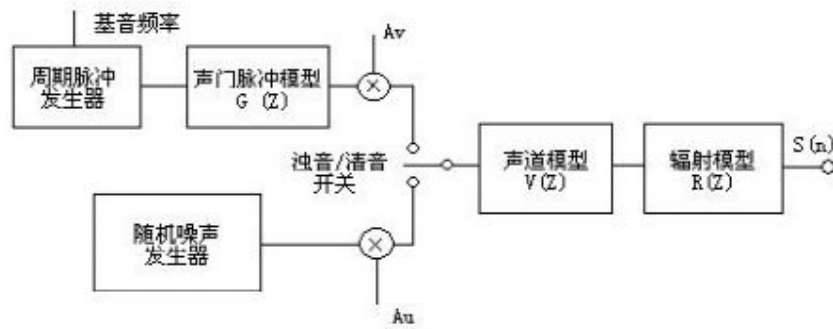


图 1: 语音信号产生的时域模型

1.1.1 激励模型

发浊音时，由于声带不断的开启和关闭，将产生间歇的脉冲波，这个脉冲波类似于斜三角形的脉冲。因此，此时的激励信号是一个以基音周期为周期的斜三角脉冲串。单个的斜三角波形的数学表达式为

$$g(n) = \begin{cases} \frac{1}{2}[1 - \cos(\pi n/N_1)], & 0 \leq n \leq N_1 \\ \cos[\pi(n - N_1)/2N_2], & N_1 \leq n \leq N_1 + N_2 \\ 0, & \text{else} \end{cases}$$

式中， N_1 为斜三角波上升部分的时间， N_2 是三角波下降部分的时间。单个三角波的频谱 $G(e^{j\omega})$ 如图 2 所示。由图可见，这是个低通滤波器，通常将其表示为 Z 变换的全极点模型的形式

$$G(z) = \frac{1}{(1 - e^{-CT}z^{-1})^2}$$

式中， C 是一常数。

上式表明斜三角波可以描述为一个二极点的单位脉冲串激励上述单个斜三角波模型的结果。而该单位脉冲串及幅值因子可以表示成下面的 Z 变换形式

$$E(z) = \frac{A_v}{1 - z^{-1}}$$

所以，整个激励模型可以表示为

$$U(z) = G(z)E(z) = \frac{A_v}{(1 - z^{-1})(1 - e^{-CT}z^{-1})^2}$$

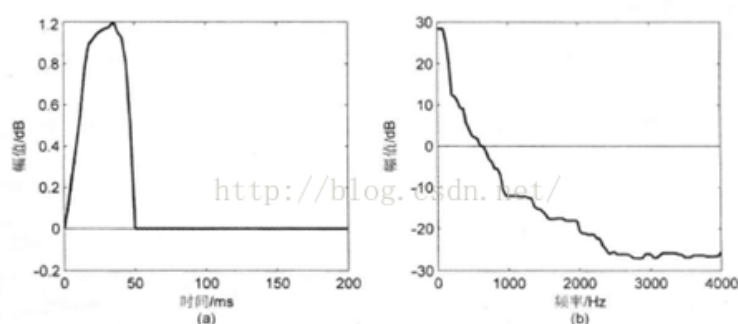


图 2: 单个斜三角形波及其频谱

另一种是发清音的情况。这时声道被阻碍形成湍流，所以可以模拟成随机白噪声。实际上，可使用均值为 0、方差为 1，并在时间或幅度上均匀分布的序列。

1.1.2 声道模型

声道部分的数学模型目前有两种观点，一种是将声道视为多个不同截面积的管子串联而形成的系统，即“声管模型”，另一种是将声道视为一个谐振腔，即“共振峰模型”。

$$V_k(z) = \frac{a_K}{1 - b_k z^{-1} - c_k z^{-2}}$$

式中 a_k 是常数，有

$$\begin{cases} c_k = -\exp(-2B_k \pi T) \\ b_k = 2\exp(-B_k \pi T) \cos(2\pi F_k T) \\ a_k = 1 - b_k - c_k \\ a_1 a_2 \dots a_M = G \end{cases}$$

M 是小于 $(N+1)/2$ 的整数， N 是极点个数， G 是幅值因子

(1) 级联型

$$V(z) = \prod_{k=1}^M V_k(z)$$

(2) 并联型

$$V(z) = \sum_1^M V_k(z)$$

(3) 混合型

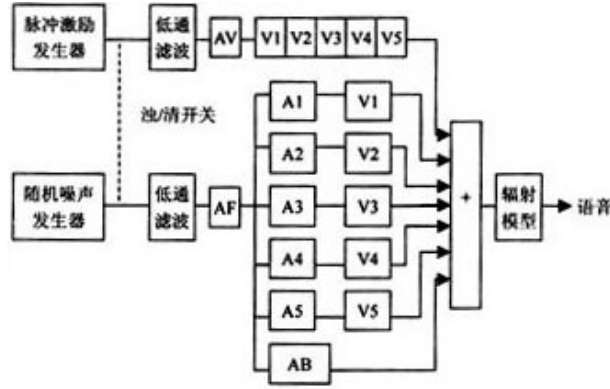


图 3: 混合型共振峰模型

1.1.3 辐射模型

声道的终端为口和唇。从声道输出的是速度波，而语音信号是声压波，二者之比的倒数称为辐射阻抗 Z_L ，它表明了口和唇的辐射效应。

研究表明，口唇端辐射在高频端较为显著，在低频端时影响较小，所以辐射模型 $R(z)$ 应是一阶类高通滤波器的形式。口唇的辐射效应可以表示为

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

它是一阶后向差分。

综上，完整的语音信号模型可用激励模型、声道模型和辐射模型三个子模型的串联来表示。该模型的 Z 变换函数为

$$X(z) = U(z)V(Z)R(Z)$$

应当指出，上式所示的模型并不和语音产生的物理过程相一致，但这种模型和真实模型在输出处是等效的。

1.2 语音信号的时域、频域特性

由语音信号的数字模型我们知道，语音信号是一个非稳态的、时变的信号。但在“短时间”范围内可以认为语音信号是准稳态的信号，短时间一般指 10-30ms，这便构成了语音信号的“短时分析技术”。

1.2.1 短时傅里叶变换

由于语音信号是短时平稳的，因此可以对语音进行分帧处理，计算某一帧的傅里叶变换，这样就得到了一个短时傅里叶变换，本质上这就是一个窗口傅里叶变换。其定义为

$$X_n(e^{i\omega}) = \sum_{m=-\infty}^{\infty} x(m)\omega(n-m)e^{i\omega m}$$

1.2.2 语谱图

将信号的强弱由记录在纸上的灰度来表示：若某个滤波器输出的信号强，相应的记录为浓黑；反之，则浅淡一些。语谱图在语音分析中经常用到，在后文中的提取基频后常与语谱图叠加在一起，以观察参数的正确性。

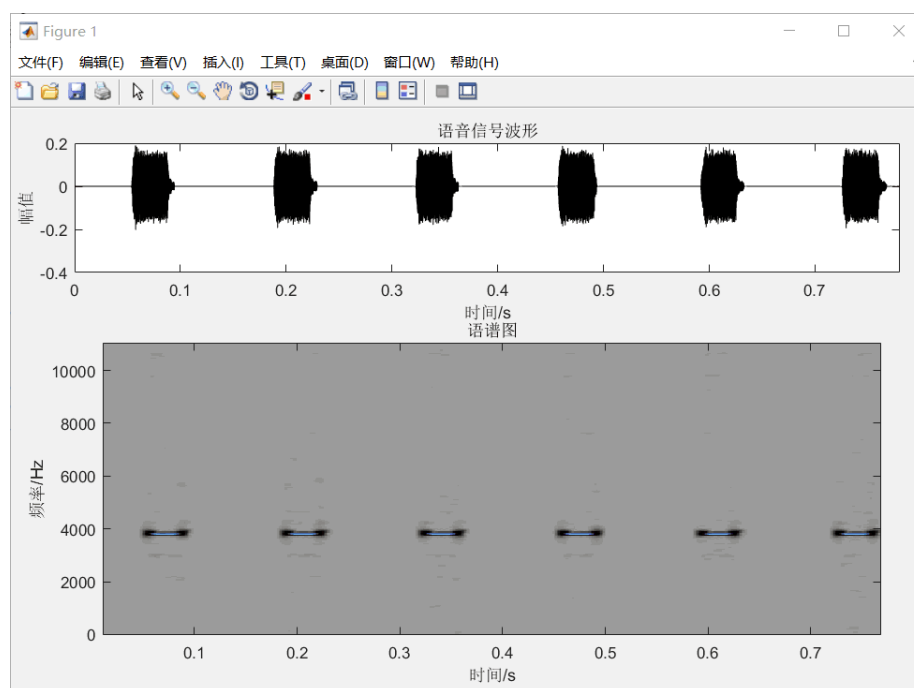


图 4: “QQ” 提示音的语谱图

2 语音信号在小波变换域中的分析技术和特性

2.1 利用小波变换提取基音信息

在语音信号处理中，可以用小波变换提取基音信息。在小波变换后，可利用近似系数的峰值获取该帧的基音频率 [2]，下面给出的就是对元音/e/的峰值图，我们得到基音频率为 254Hz。

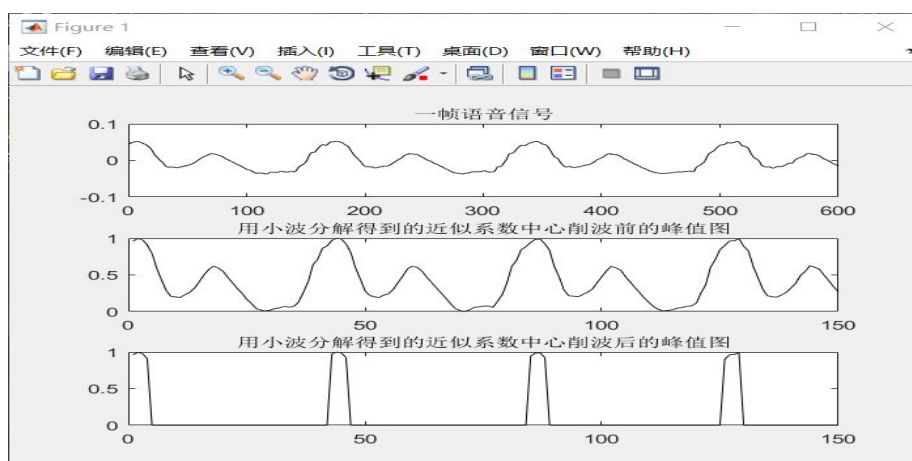


图 5: 洛天依音源/e/的语音信号处理

2.2 用小波包分解设计 Bark 滤波器

2.2.1 Bark 域介绍

耳蜗是内耳中听觉的传感器官，临界频带指的是由于耳蜗构造产生的听觉滤波器的频率带宽。概况地说，临界频带是声音频率带，在临界频带中第一个单音感知性会被第二单音的听觉掩蔽所干扰。声学研究中，人们使用听觉滤波器来模拟不同的临界频带。后来研究者发现人耳结构大致会对 24 个频率点产生共振，这就是 Bark 域。

根据相关的心理声学的研究，由于人耳的特殊结构，在同一个临界频带里头信号容易发生掩蔽效应，即：主要信号容易被能量大并且频率接近的掩蔽信号所掩蔽。因此我们可以认为 Bark 域约近的信号越容易产生掩蔽效应。Zwicker 和 Terhardt 于 1980 提出了近似表示 Bark 域的一种方式

$$Bark = 13 \tan^{-1}(0.76f/1000) + 3.5 \tan^{-1}(f/7500)^2$$

Bark 域对低频具有放大作用，高频具有压缩作用。这点能够从上面的转换公式明显看出来。回顾声音的掩蔽效应，从 Bark 域来看，能够更清晰地分析出哪些信号容易产

生掩蔽，哪些噪声比较明显。同时对于某些声音特征例如基音，音调等在 Bark 域分析也有独特的效果。

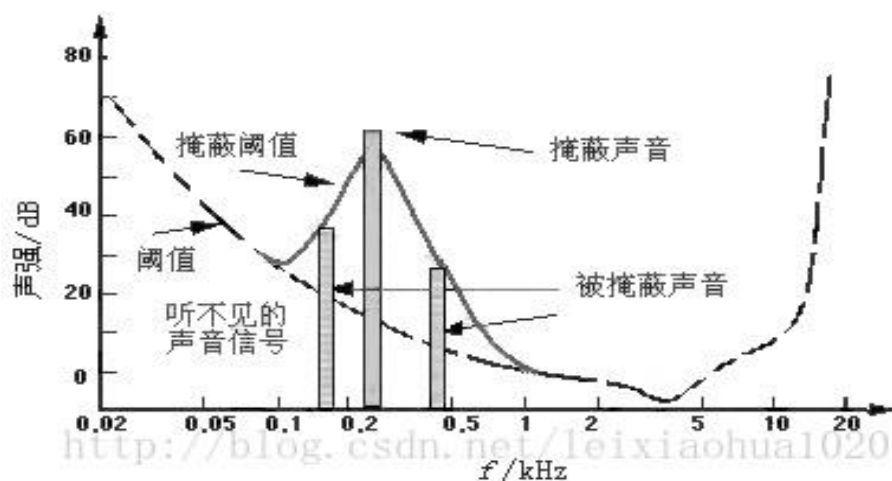


图 6: 掩蔽效应

2.2.2 用小波包分解设计 Bark 滤波器

掩蔽效应用小波包分解，我们可以设计出 Bark 滤波器，在采样率为 8000Hz 时，小波包分解对应于 21 个 Bark 尺度的滤波器 [3] db2 小波母函数对元音/a/进行分解，然

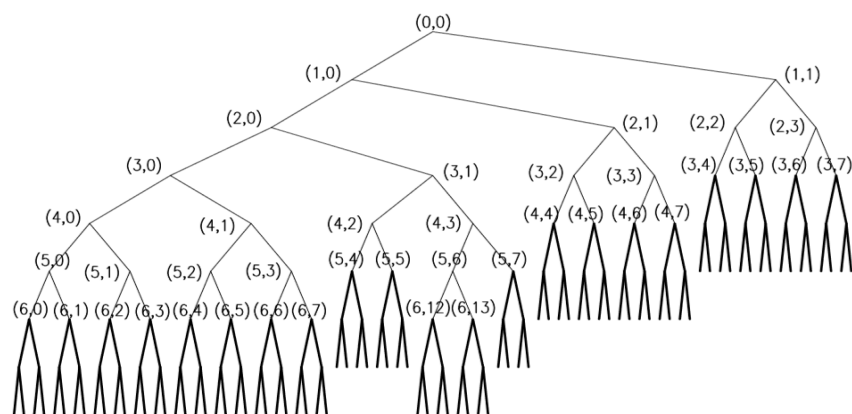


图 7: Bark 滤波器由小波包分解的结构如意图

后取前 8 个 Bark 滤波器的输出进行重构，得到的波形为/a*/。

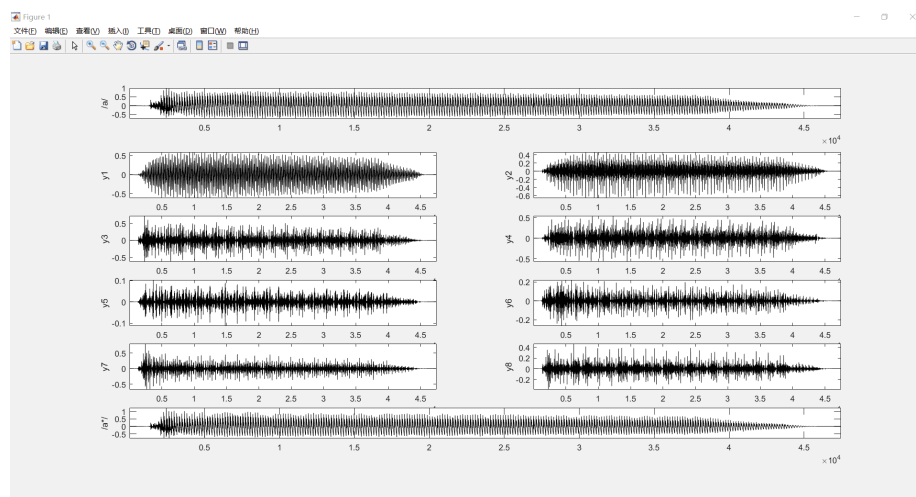


图 8: 元音/a/的波形和前 8 个 Bark 滤波器的输出以及/a*/的波形

3 小波变换在端点检测中的应用

3.1 语音端点检测的常用方法

(1) 双门限法

这最初是基于短时平均能量和短时平均过零率而提出的,其原理是汉语中有元音,能量较大,所以可以从短时平均能量中找到韵母,而声母是辅音,他们的频率较高,相应的短时平均过零率较大,所以可以用这两个特点找出声母和韵母,等于找出了完整的汉语音节。它是通过如下图 9 的二级判决来实现的。

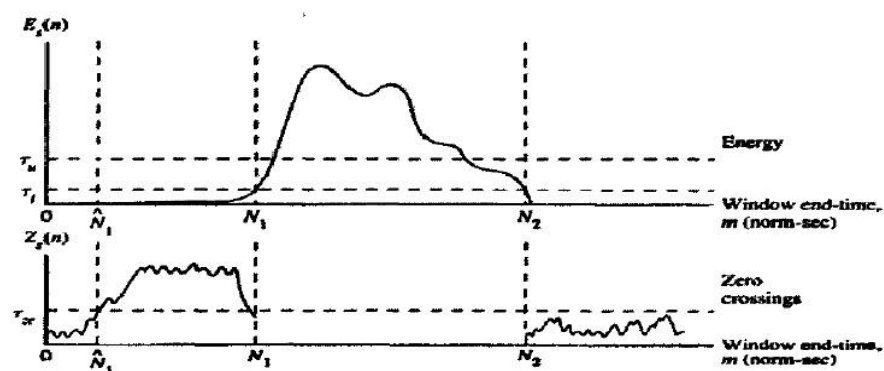


图 9: 双门限法端点检测示意图

(2) 相关法

语音信号 $x(n)$ 分帧后有 $x_i(m)$ ，下标 i 表示第 i 帧 ($i = 1, 2, \dots, M$), M 为总帧数。每帧数据的短时自相关函数定义为

$$R_i(k) = \sum_{m=1}^{L-k} x_i(m)x_i(m+k)$$

式中, L 为语音分帧后每帧的长度, k 为延时量。

为避免语音端点检测过程中绝对能量带来的影响, 把自相关函数进行归一化处理。

$$R_i^n(k) = R_i(k)/R_i(0)$$

对自相关函数进行低通滤波后, 发现与噪声存在差别, 于是利用这点我们便可对语音进行端点检测。

3.2 小波分解系数的平均幅值法进行端点检测

小波变换分解后, 可以被看作由低通滤波器和带通滤波器组成的滤波器组, 低通滤波器输出的是近似系数, 带通滤波器输出的是细节系数。小波系数较大者携带信号能量较多, 系数较小者则携带信号能量较小。对不同子带内的小波系数进行统计分析, 我们可得到噪声和语音的不同分布特征, 在这我们选 db4 小波进行分解。

语音信号 $x(n)$ 分帧后有 $x_i(m)$ ，下标 i 表示第 i 帧 ($i = 1, 2, \dots, M$), M 为总帧数, N 为帧长。对每帧信号做 db4 小波的 10 层分解, 细节系数为 $S_j^k (k = 1, 2, \dots, 10)$, 表示第 k 层第 j 个小波系数。对每帧分解的 10 层, 把它的细节系数分为两部分, 1-5 层为第一部分, 6-10 层为第二部分。分别求出每层的平均幅值, 第 k 层的小波系数平均幅值为

$$E_i^k = \frac{1}{L(k)} \sum_{j \in L(k)} |S_j^k|$$

式中, 下标 i 表示第 i 帧; $L(k)$ 表示第 k 层小波细节系数的长度。再令

$$\begin{cases} M_1(i) = \max E_i^1, E_i^2, E_i^3, E_i^4, E_i^5 \\ M_2(i) = \max E_i^6, E_i^7, E_i^8, E_i^9, E_i^{10} \end{cases}$$

在 1-5 和 6-10 两部分中找出各自平均幅值的最大值, 再进一步计算 $M_1(i)$ 和 $M_2(i)$ 的乘积:

$$MD_i = M_1(i) * M_2(i)$$

这是利用噪声和语音在小波分解的各子带中有不同的特性, 再利用单参数的双门先检测法来区分。[5]

3.3 小波包 BARK 子带方差的端点检测

我们在时域中进行方差的计算, 设含噪语音信号的时域波形为 $x(n)$, 加窗分帧处理后得到的第 i 帧语音信号为 $x_i(m)$, 则 $x_i(m)$ 满足

$$x_i(m) = \omega(m) * x(iT + m), \quad 1 \leq m \leq N$$

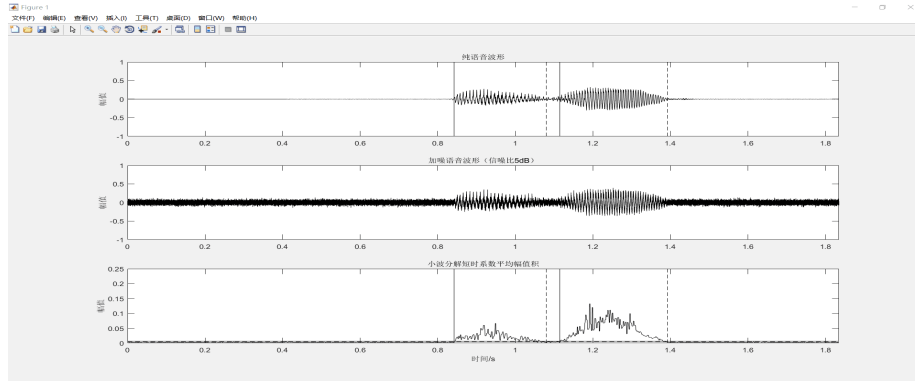


图 10: 小波分解系数的平均幅值法进行端点检测的结果

把 $x_i(m)$ 分解在 J 个 BARK 子带中, 每个子带中的信号 $x_i^j(m)$ 为

$$x_i^j(m) = h^j * x_i(m), \quad j = 1, 2, \dots, J$$

式中, j 指第 j 个 BARK 子带, h^j 是第 j 个 BARK 子带滤波器的冲激响应, $*$ 是卷积。在这我们通过小波包变换获得, 即用小波包分解设计 Bark 滤波器。

对第 i 帧第 j 个 BARK 子带中的信号 $x_i^j(m)$ 的方差我们可以求得为

$$D_i^j = E[\{x_i^j(m) - E[x_i^j(m)]\}^2]$$

所以第 i 帧各个 BARK 子带的平均方差值为

$$\overline{D}_i = E[D_i^j]$$

利用 \overline{D}_i 的值, 再利用单参数的双门先检测法我们可对语音信号进行进行端点检测。

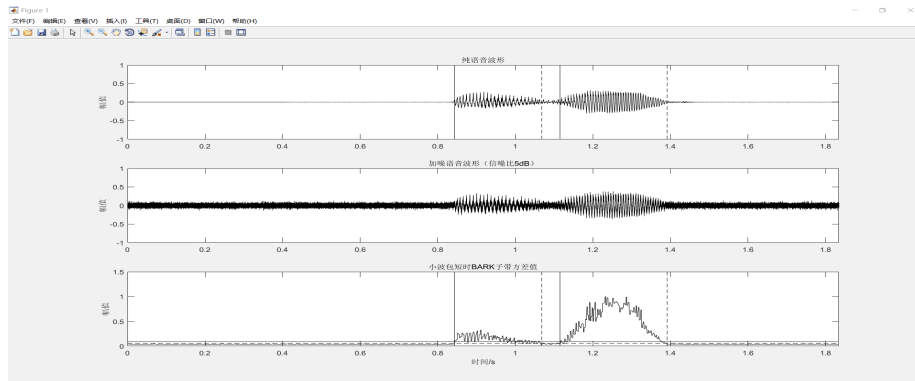


图 11: 用 BARK 子带分离频带方差法进行端点检测的结果

参考文献

- [1] MATLAB 语音信号分析与合成. 北京航空航天大学出版社,2017.
- [2] 王长富, 林志刚, 戴蓓倩, 等. 基于小波变换的语音基音周期检测 [J]. 中国科学技术大学学报,1995,25(1):47-52
- [3] 冯流保. 基于听觉掩蔽效应的小波包语音增强 [J]. 通信技术,2010,43(3):139-141
- [4] Cohc I. Enhancement of Speech Using Bark-Scaled Wavelet Packet Decomposition[J]. Interspeech,2001:1933-1936.
- [5] 郭永亮, 施玉霞. 基于小波系数能量及其方差的语音端点检测 [J]. 微型电脑应用,2009,25(11):36-37.