

基于语音频谱包络抽取的 MFCC 算法*

李 波^{1,2}, 王成友¹, 杨 聪¹, 蔡宣平¹, 张尔扬¹

(1. 国防科技大学电子科学与工程学院, 湖南 长沙 410073; 2. 空军工程大学电讯工程学院, 陕西 西安 710077)

摘 要 : MFCC 可以更加准确地表示语音低频处的频谱包络, 在语音转换中是一种很好的频谱包络表示方法, 提出一种新的有效的求解 MFCC 的算法, 先对频谱通过谐波抽取得到阶梯谱包络, 再对阶梯谱包络进行 Mel 尺度变换, 最后通过余弦变换求得 MFCC 参数。该方法实现简单, 运算结果准确。

关键词 : 语音处理; MFCC; 语音识别; 说话人识别; 语音转换

中图分类号 : TP391 **文献标识码** : A

An Algorithm for MFCC Based on Speech Spectrum
Harmonic Samples ExtractionLI Bo^{1,2}, WANG Cheng-you¹, YANG Cong¹, CAI Xuan-ping¹, ZHANG Er-yang¹

(1. College of Electronic Science and Engineering, National Univ. of Defense Technology, Changsha 410073, China;

2. The Telecommunication Engineering Institute, Air Force Engineering Univ., Xi'an 710077, China)

Abstract : MFCC can express speech spectrum more accurately at low frequency, so MFCC is a good method to express the spectral envelope in voice conversion; A new and efficient algorithm for MFCC is proposed. At first, the staircase envelope is obtained through harmonic extraction, then the staircase envelope is transformed by Mel frequency scale, at last the MFCC coefficients are extracted by cosine transformation. The algorithm can be realized easily, and the result is good.

Key words : speech processing; MFCC; speech recognition; speaker recognition; voice conversion

语音的频谱包络反映声道的滤波特性, 语音内容不同或者语音内容相同但说话人不同, 其声道的滤波特性就不同, 这种声道的不同滤波特性由不同的频谱包络反映出来。语音编码中的 LPC 系数及 LPC 系数的推演参数, 语音识别和说话人识别中的 LPC 倒谱、倒谱和 Mel 倒谱 (MFCC)^[4], 都与语音的频谱包络密切相关, 不同的参数表示不同的频谱包络, 语音转换^[2] (voice conversion) 就是要保持语音的语义内容不变, 而改变语音的说话人个性特征, 对于由声道特性决定的语音音色是语音转换的一项很重要的内容, 其中的频谱包络可以由 LPC 的推演参数 (如 LSF、PARCOR、LAR 等)、倒谱和 MFCC 来表示, 以便进行转换。语音识别和说话人识别中的 LPC 倒谱、倒谱和 MFCC 只要求其不同的频谱包络对应, 以便在识别的过程中能区分不同的频谱包络, 即区分不同的语音, 不必要求能由这些参数反过来准确地求出频谱包络, 而语音编码中和语音转换中则要求其 LPC 系数、LPC 系数的推演参数、倒谱、MFCC 能准确地求出频谱包络, 这样才能使合成的语音达到高的质量。

语音的频谱在低频处变化是最大的, LPC 系数和倒谱系数虽然在总体上可以较好地表示频谱包络, 但在低频处, 却往往对频谱包络表示得不够准确, 而人耳对语音频率的感知是非线性的, 对低频的感知要比对高频灵敏, Mel 尺度的频率描述了人耳对频率感知的非线性特征, 在语音识别中, 已证明 MFCC 优于其他倒谱系数。语音转换中要高质量地转换频谱包络, 精确地表示低频处的频谱包络是前提, 在 Stylianou^[2] 的语音转换系统就采用 MFCC 来作为频谱包络的表示参数。

MFCC 有不同的提取算法, 在语音识别中普遍采用对能量谱用 Mel 频率尺度三角滤波法来求解^[7], 这种方法如前面所述, 不能精确地表示频谱包络, 在语音转换中是不能采用这种方法来表示频谱包络的, Tokuda^[6] 给出了由 LPC 系数计算 MFCC 的方法, 这种方法得到的结果直接受 LPC 系数提取精度影

* 收稿日期 2004 - 02 - 25
作者简介: 李波 (1974—), 男, 博士生。

响,低频处 LPC 谱包络不精确,则在此基础上得到的 MFCC 也不会精确,而且 LPC 谱包络在清音时往往效果不理想。Stylianou^[2]采用解矩阵求解方程法来求解 MFCC,这种方法比较精确,为一些语音研究者所采用,但该方法要求解矩阵的逆,随着矩阵阶数的增加,计算量将很大,而且这种方法是基于谐波处的谱幅度值来求解的,当谐波特性不好时,性能就会受影响,对于清音的情况,该方法的效果就不理想。

在文献 [1][3] 中给出了求解倒谱的算法,受此启发,提出一种简单有效的求解 MFCC 算法,首先与 [1] 中方法一样求得阶梯谱包络,再对阶梯谱包络通过 Mel 尺度进行变换,再用求解倒谱相同的方法进行余弦变换,即可得到 MFCC 参数。该方法得到的 MFCC 参数能较 LPC 系数和倒谱系数更加准确地表示低频处的谱包络,克服了 Stylianou 的矩阵求解算法的缺点,不必求解矩阵的逆,几乎不受语音频谱谐波特性影响,在求解清音的 MFCC 的情况下依然是准确而有效的。用本文的算法求得的 MFCC 参数也应该可以应用于语音识别和说话人识别中,是基于能量谱用 Mel 频率尺度三角滤波法来求解方法^[7]的一种改进算法。

1 MFCC 求解算法

下面以一个例子的求解过程来说明 MFCC 的求解方法。取一帧浊音为例,采样频率为 11 025 Hz,一帧为 400 个采样点,语音表示为 $x(n), 0 \leq n < 400$ 。MFCC 求解的前提是已进行基音频率 F_0 提取,此处用我们研究提出的谐波自相关法^[5]。

(1) 用汉明窗进行加窗处理,减少 Gibbs 效应。

(2) 对加窗语音进行 1024 点的 DFT 变换得到 $S[k]$,计算求得该帧语音的幅度谱, $S[k] = 20 \log(|S[k]|), 0 \leq k < 512$ 。

(3) 逐个在基音频率周期内搜索局部极大值 $A_{f_n} (nF_0 - F_0/2 < f_n < nF_0 + F_0/2), m$ 为整数。

(4) 如果 f_n 与 f_{n+1} 的间隔 $d = f_{n+1} - f_n > 1.5F_0$,则在间隔 $[f_n + d/4, f_n + 3d/4]$ 内再进行极大值搜索。

(5) 对得到的极大值序列进行线性插值,得阶梯谱包络 $S_l[k]$,如图 1 所示。

(6) 对阶梯谱包络直接进行 Mel 尺度拉伸(说明:没有像 Madlov^[1]那样对阶梯谱包络进行 Blackman 窗平滑,是因为实验中发现平滑与否对最后求得的效果影响很小)。Mel 频率与 Hz 频率间的关系如下:

$$m = \frac{1000 \cdot \ln\left(1 + \frac{f}{700}\right)}{\ln\left(1 + \frac{1000}{700}\right)} \approx 1127 \ln\left(1 + \frac{f}{700}\right) \quad (1)$$

m 的单位为 Mel, f 的单位为 Hz。由上式可推得:

$$f = 700 \cdot (e^{\frac{m}{1127}} - 1) \quad (2)$$

当 $f = 11 025/2$ Hz 时,由式(1)得 $m = 2460.5$ Mel,此处的实验为每 5 Mel 进行一次由 Hz 到 Mel 的阶梯谱包络计算,得到 Mel 尺度的包络函数 $S_l[m], m = 0, 1, \dots, M-1$,结果如图 2(1)所示。

(7) 求 MFCC,计算公式如式(3):

$$c_n = \sum_{m=0}^{M-1} S_l[m] \cdot \cos\left(\frac{2\pi mn}{2M}\right) \quad n = 0, 1, \dots, N-1 \quad (3)$$

在实验中,选取的 MFCC 阶数为 20。这样就得到了 MFCC 系数。

上面即为求解 MFCC 的整个算法过程。此处是以浊音帧为例来描述算法的,当在清音帧时或语音频谱的浊音特性不够好情况下,显然阶梯谱包络一样是可以很容易求得,用该方法计算 MFCC 一样是有效的。

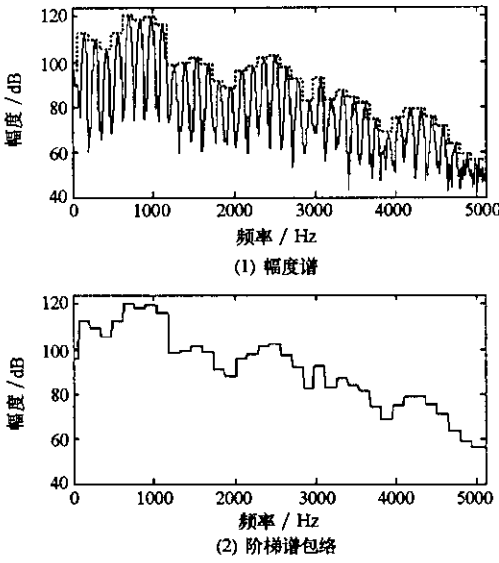


图 1 阶梯谱包络提取结果图

Fig.1 Staircase envelope extracted results

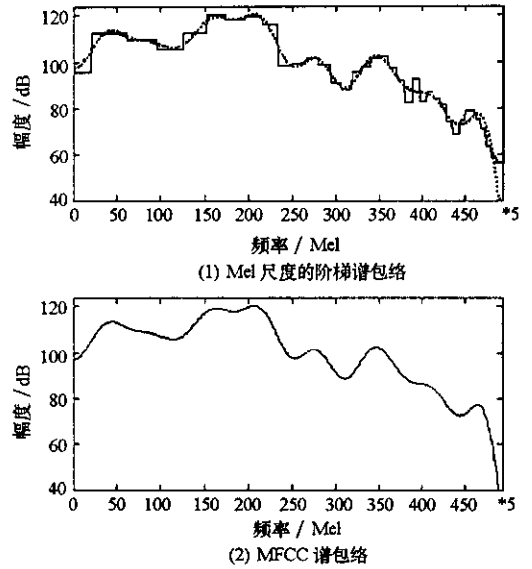


图 2 Mel 尺度阶梯谱包络和 MFCC 谱包络

Fig.2 Mel-scale staircase envelope and MFCC envelope

2 通过对比实验验证 MFCC 表示谱包络的有效性

我们由 MFCC 系数求解频谱包络,与倒谱频谱包络和 LPC 频谱包络进行对比,验证其有效性,阶数都为 20。

首先由 MFCC 系数求解频谱包络,根据式(4)由 MFCC 直接计算包络函数,此处称之为 MFCC 谱包络,结果如图 2 所示。

$$S[m] = c_0 + 2 \sum_{n=1}^{N-1} c_n \cos\left(\frac{2\pi mn}{2M}\right) \quad m = 0, 1, \dots, M-1 \quad (4)$$

将由式(4)计算的 MFCC 谱包络按公式(2)反变换回以 Hz 为单位的频谱包络,此处称之为 MFCC-线性频谱包络,结果如图 3 所示。

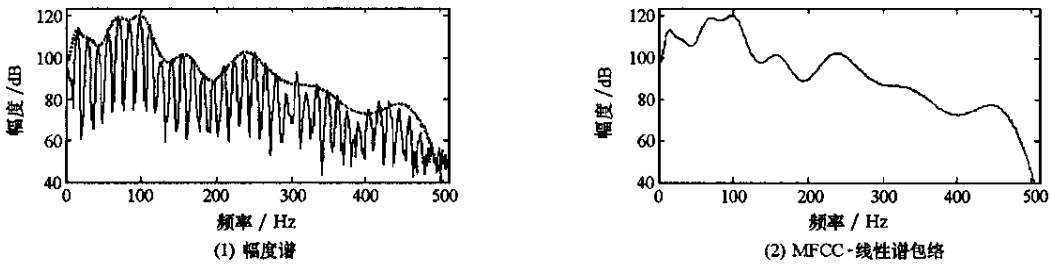


图 3 MFCC-线性谱包络

Fig.3 MFCC-linear envelope

再计算倒谱包络,对进行 Mel 尺度拉伸之前的阶梯谱包络 $S[k]$ 直接按式(3)进行计算得到的是倒谱,再按公式(4)求得倒谱包络^[1],如图 4(1)所示。

用 Durbin 递推算法求解 LPC 系数^[7],再求解 LPC 频谱包络,结果如图 4(2)所示。

对比图 3 图 4,很容易可以看出由 MFCC 系数可以更加精确地表示语音信号在低频处的频谱包络。

