基于 LPCC 和 MFCC 的藏语语音端点检测算法

李洪波1，于洪志2
中国科学院 自动化研究所

摘要 端点检测是语音识别系统预处理阶段遇到的第一个关键技术。该算法根据藏语元音/辅音频谱特性差异，对语音信号分高/低频带后再分别处理的思想，符合藏语语音的清/浊对立信息分布特点，然后分别提取倒谱系数作为端点检测的特征，因为倒谱特征所含的信息比其他参数多，能较好地表征语音信号，语音质量好，识别正确率高；检测时采取自适应噪声参数估计，根据倒谱距离轨迹判决起止端点，仿真结果表明了它的优越性。

关键词 端点检测；LPCC；MFCC；安多藏语

Endpoint Detection Algorithm of Tibetan Pronunciation Based on LPCC and MFCC

Li Hongbo[1] , Yu Hongzhi[2]

Institute of Automation Chinese Academy of Sciences

Abstract Endpoint detection is the first essential technology which speech recognition system meets in pre-processing stage. This algorithm which based Tibetan vowel/consonant frequency spectrum characteristic, separately is processed again through the pronunciation signal minute high/low-frequency band, conforms to Tibetan pronunciation clear/muddy opposition information distribution characteristic, then separately withdraws but actually is scored the cepstral coefficient to take the endpoint detection characteristic, because the cepstral coefficient actually scores the information which the characteristic contains compared to other parameters many, can attribute the better attribute pronunciation signal, the pronunciation quality is good, the recognition accuracy is high; When examination adopt the auto-adapted noise parameter to estimate that, decided beginning/end vertex according to ceptrum distantce, the simulation result indicated its superiority.

Key words speech endpoint detection；Lpc cepstral coefficient；Mel-frequency cepstral coefficient；Ando Tibetan

---

[1]**作者简介**：李洪波(1971—)，女，副教授、研究生，主要研究方向：语音识别，语音编码，计算机应用

[2]于洪志，教授、博导，研究方向为多文种信息处理、中文信息

项目资助：本项目得到中国科学院自动化研究所模式识别国家重点实验室开放课题“安多藏语语音合成文本分析基础研究”资助

1 引言

藏语语音信号预处理技术是进行语音处理的重要环节，藏语语音识别的预处理也是藏语语音识别系统中的重要的一步。端点检测是语音识别系统预处理阶段遇到的第一个关键技术。藏语语音端点检测的准确性甚至在某种程度上直接决定了整个藏语语音识别系统的成败，没有足够准确的端点检测(尤其是起点)，精密优选特征类型或识别方法的工作往往劳而无功。

2 藏语语音学知识

藏语以音节为单位成词，句。寻找语音段的起止点实际上就是寻找音节的起点和终点，研究藏语的音节结构、声韵特性将有助于有效的选择所需的特征。

2.1 藏文音节结构

藏文自左向右横写，音节之间用音节符隔开。音节内部有些字母可以上下叠写。这是藏文不同于一般拼音文字的一个特点。藏文的音节最少由一个字母构成，最多用六个字母组成。每个音节还可以加元音符号。若无元音符号，则读 a。 举例如下:

一个字母构成的音节: ག ga

两个字母构成的音节: གད ga

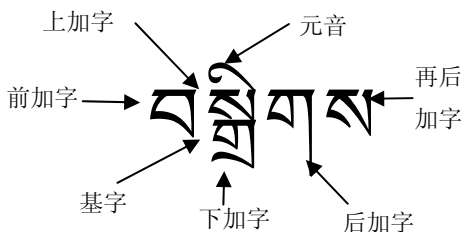
三个字母构成的音节: གདན gaNs

四个字母构成的音节: དམངས dmaNs

五个字母构成的音节: བསྒྲགས bsgrags

六个字母构成的音节: བསྒྲིགས bsgrigs

其中六个字母构成的音节: བསྒྲིགས “结构”，若按藏文方式排列如下所示。



多字母构成的音节，必有字母作核心，叫基字。藏文 30 个辅音字母均可充当基字，分别可带元音、上加字、下加字、前加字、后加字和再后加字。

语言中最常见的音节结构类型是辅音加元音，即 CV 结构，但基于这种基本结构之下，不同语言有不同的音节结构模式，藏语音节结构模式为(C)(C)V(C)，此音节结构模式是在节首和韵尾上，藏语节首可以有 0-2 个辅音或辅音组合（复辅音），韵尾有 0-1 个辅音。按照节首音位配列的特征可以分为三种。(1)VC 结构：藏语起首的音节元音前均带有轻微的喉塞

音/？/）(2) CVC 结构：藏语中除浊塞音、浊塞擦音一般不能作词首音节声母外，其它单辅音音位可以作为节首(3) CCVC 结构：安多方言的阿力克话复辅音声母较多而且比较完整，在现代藏语中有一定的代表性。它由p、、F、、 r、、m、、 n、、 w、6 种前置辅音与基本辅音构成的 92 种复辅音声母，另外还有一个由基本辅音与后置辅音构成的hw。此外安多藏语的道孚话还有 CCCV 结构（如道孚话zbrg“蛇”）和 CCCVC 结构（如道孚话 zprgn“云”）。

2.2 藏语安多方言（半农半牧区）语音特点

藏语安多方言语音上的重要特征概括起来就是声母分清浊，声调无区别词义的作用，复辅音较多。安多声母系统的主要特征是清浊音对立，送气音与不送气音对立，声母数目较藏语其他方言较多。韵母系统的主要特征都是单元音，没有复元音韵母。

藏文上的前加字或上加字大体上相当于复辅音声母中的前置辅音，基字相当于基本辅音，下加字相当于后置辅音。从现行藏文来看，古藏语的声母由一个到四个辅音组成。藏文三十个字母，不带任何上加字、前加字或下加字作单词出现时，这就是我们所说的有单辅音构成的声母。藏文所反映的复辅音系统存在二合复辅音声母、三合复辅音声母、四合复辅音声母。在安多方言（半农半牧区）中复辅音声母已比较简化，能作前置辅音的归纳起来只有鼻音n和喉擦音h(ï)两个，结合起来也是固定的，由此构成只有二合复辅音声母共 29 个，已没有三合和四合复辅音声母。本文据文献对藏语安多方言清浊音对立情况总结如表 1。

表 1　声母清浊分类

	清音	浊音
双唇塞音	p　pH	b
舌尖塞音	t　tH	d
舌根塞音	k　kH	g
舌尖前塞擦音	ts　tsH	dz
舌尖后塞擦音	t½　t½H	dÈ
舌面前塞擦音	t　tH	dþ
舌尖前擦音	s	z
舌面前擦音	_	þ
鼻音		m　n　ù　N
边音		l
半元音		j

2.2.1 单辅音声母

在安多藏语中，清塞音t、k，清塞擦音ts、t½、t，清擦音s、_都能单独作声母，即单辅音声母。清送气音tH、kH、tsH、t½H、tH一般都是单辅音作声母。鼻音 m、n、ù、N，边音 l都能单独作声母。

2.2.2 二合复辅音声母

清（喉）擦音h能同清塞音t、k，清塞擦音ts、t½、t，清擦音s、_相结合。

浊（喉）擦音ï能同鼻音 m、n、ù、N，边音 l，半元音j相结合。

鼻音n能同浊塞音b、d、g，浊塞擦音dz、dÈ、dþ相结合（其发音部位实际上跟基本辅音的发音部位相一致）。

据文献总结藏语安多方言二合复辅音结合规律情况如表 2。

表 2 二合复辅音结合表

前置辅音	基本辅音	
h	清音	t、k、ts、tʃ、t、s、
ï	浊音	d、g、dz、dʒ、dɖ、z、ɮ　　m、n、ù、N
	边音	l
	半元音	j
n	浊音	b、d、g、dz、dʒ、dɖ

2.2.3 习惯调

藏语三大方言中，卫藏和康两大方言有声调，安多藏语没有声调。安多藏语现在并不见有的音节结构方面，通过声调高低或曲折来区别词义的现象，但在语言中存在着一种“习惯调”。一般是清声母字读高调，浊声母字读低调。在双音节词中，前一音节仍然是“清高浊低”。后一音节不论其清浊，一般读成高调，但调值比清声母的要略低些。

2.2.4 韵母系统

安多方言只有单元音韵母，没有复元音韵母。有ཨ　a[?a]、ི(i)[i]、ུ(u)[u]、ེ(e)[e]、ོ(o)[o]。

3 藏语语音信号特征参数选取

3.1 线性预测倒谱系数(LPCC)

线性预测倒谱系数（LPCC）是线性预测系数（LPC）在倒谱域中的表示。该特征是基于语音信号为自回归信号的假设，利用线性预测分析获得倒谱系数。典型的藏语语音 LPCC 参数求解流程如图 1 所示：

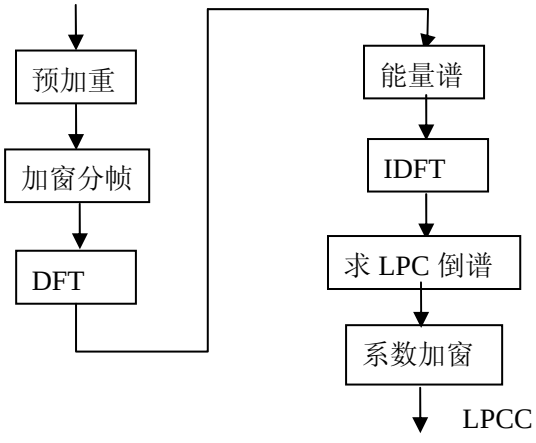


图 1　LPCC 求解流程

LPC系数可用来估计语音信号的倒谱，这也是语音信号短时倒谱分析中一种特殊的处理方法。在线性预测(LPC)分析中，声道模型系统函数为：

$$H(Z)=1/(1-\sum_{k=1}^{p}a_kZ^{-K}) \qquad (1)$$

其冲击响应为h(n)，设$\hat{h}(n)$ 表示$h(n)$的倒谱，则有：

$$\hat{H}(Z)=\sum_{n=1}^{\infty}\hat{h}(n)Z^{-n} \tag{2}$$

将(2)式代入并将其两边求导数，有：

$$(1+\sum_{k=1}^{p}a_kZ^{-K})\sum_{n=1}^{\infty}n\hat{h}(n)Z^{-n+1}=\sum_{k=1}^{p}ka_kZ^{-K+1} \tag{3}$$

令上式左右两边的常数项和各次幂的系数分别相等，从而可由a_k求得$\hat{h}(n)$：

$$\hat{h}(0)=0$$

$$\hat{h}(1)=a_1$$

$$\hat{h}(n)=a_n+\sum_{n=1}^{\infty}(1-k/n)a_k\hat{h}(n-k)\qquad(1\le n\le p) \tag{4}$$

$$\hat{h}(n)=\sum_{n=1}^{\infty}(1-k/n)a_k\hat{h}(n-k)\qquad(n\ge p)$$

按(4)式求得的倒谱称为LPC复倒谱(LPCC)，由$c(n)=0.5(\hat{h}(n)+\hat{h}(-n))$求得LPC实倒谱。式中p为LPC阶数(10=<p<=16)，n为LPCC阶数。LPC倒谱(LPCC)由于利用了线性预测中声道系统函数的最小相位特性，避免了相位卷积，求复对数的复杂，LPCC参数的优点是计算量小，易于实现。通过分析激励信号的语音特点以及声道传输函数的零极点分布情况，可知激励信号倒谱的分布范围很宽，语音信号倒谱从低时域延伸到高时域，而$\hat{h}(n)$主要分布于低时域中。而语音信号所携带的语义信息主要体现在声道传输函数上，因而在语音识别中通常取语音信号倒谱的低时域构成LPC倒谱特征。

3.2 Mel 频标倒谱系数(MFCC)

美尔频标倒谱系数（MFCC）考虑了人耳的听觉特性，将频谱转化为基于MEL频标的非线性频谱，然后转换到频谱域上。由于充分考虑了人的听觉特性，而且没有任何前提假设，MFCC参数具有良好的识别性能和抗噪声能力。MFCC是采用滤波器组的方法计算出来的，这组滤波器在频率的美尔坐标上是等带宽的。如图2给出频率与美尔频率的特性曲线。

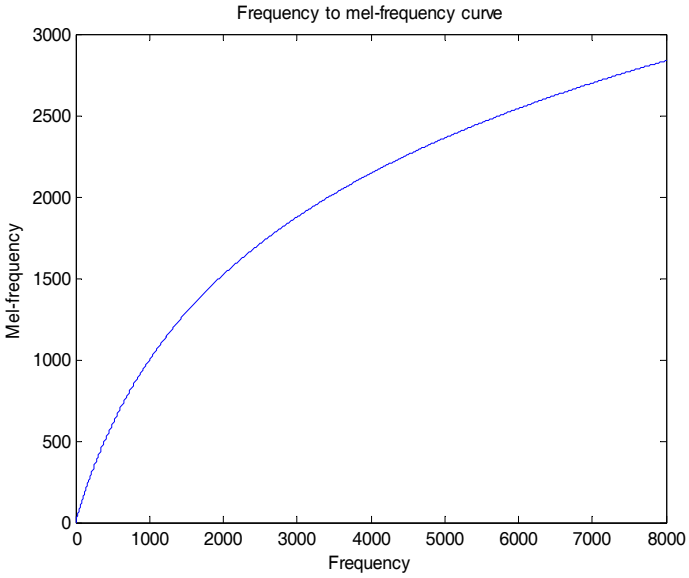


图 2　MFCC 计算过程

藏语语音 MFCC 参数计算过程如图 3，具体计算步骤如下：

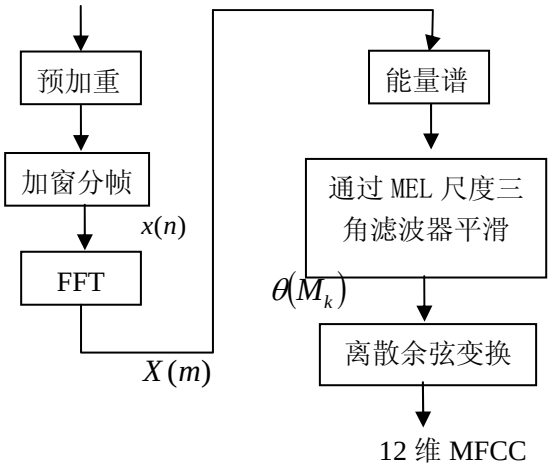


图 3　MFCC 计算过程

(1) 语音信号在经过加窗处理后变为短时信号，用 FFT 将这些时域信号 x(n)转化为频域信号 X(m)，并由此可以计算它的短时能量谱 P(f)。

(2) 将 P(f)由在频率轴上的频谱转化为在美尔坐标上的 P(M),其中 M 表示美尔频率，由下式可以完成该转换，并且美尔频率考虑了人耳的听觉特性。

$$F_{mel} = 3322.23\lg(1+0.001) f_{Hz} \tag{5}$$

(3) 在美尔频域内将三角带通滤波器加于美尔坐标得到滤波器组 $H_m(k)$，然后计算美尔坐标上的能量谱 P(M)经过此滤波器组的输出：

$$\theta(M_k) = \ln \sum_{k=1}^{K} |X(k)|^2 H_m(k) \quad k=1,2\cdots,K \tag{6}$$

式中，k 表示第 k 个滤波器，K 表示滤波器个数。

(4)通过一个具有 40 个滤波器的滤波器组。前 13 个滤波器在 1000Hz 以下是线性划分的，后 27 个滤波器在 1000Hz 以上是在美尔坐标上线性划分的。

(5)如果$\theta(M_k)$表示第 k 个滤波器的输出能量，则美尔频率倒谱$C_{mel}(n)$在美尔刻度谱上可以采用修改的离散余弦反变换(IDCT)求得：

$$C_{mel}(n)=\sum_{k=1}^{K}\theta(M_k)\cos(n(k-0.5)\pi/k) \qquad n=1,2,\cdots,p \qquad (7)$$

式中，p 为 MFCC 参数的阶数。

4 基于倒谱特征的藏语语音端点检测的方法框图

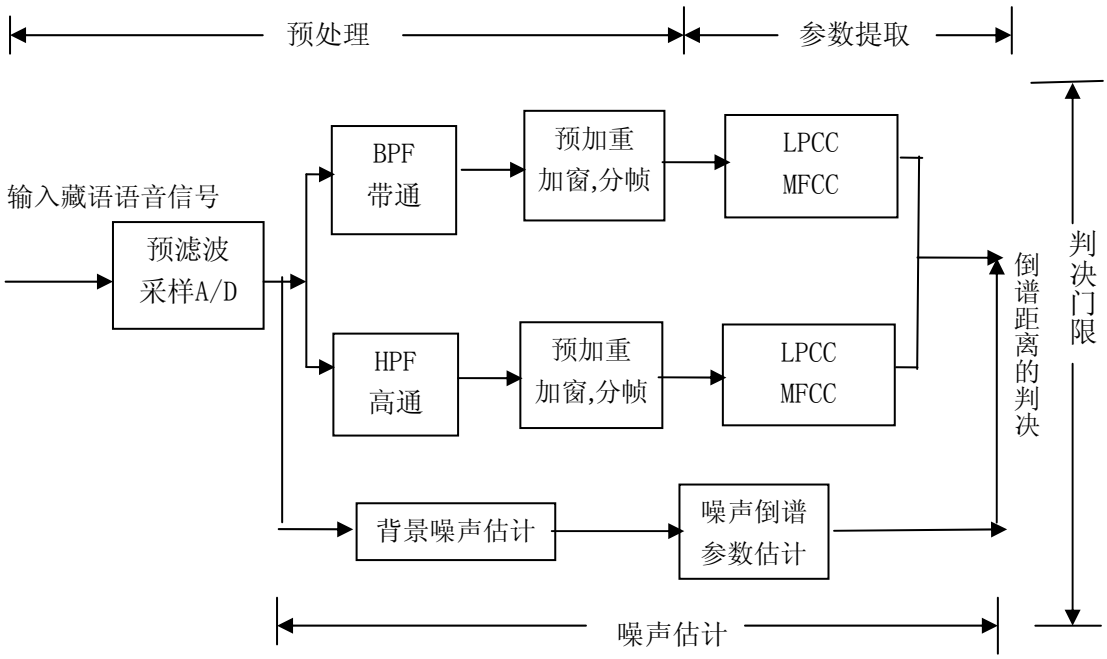


图 4　基于倒谱特征的藏语语音端点检测的方法框图

基于倒谱特征的藏语语音端点检测的方法框图说明如下：

(1)藏语语音预处理。先将经A/D转换采样后的藏语语音信号经数字滤波分成高、低频2个子带，频带间允许重叠。然后对滤波形成的2个子带藏语语音信号分别进行分帧、加窗，根据高频段清音信息丰富，变化较快；低频段浊音信息丰富，且变化较慢的特点，选择不同的帧长、帧移和窗长。

(2)参数提取。分别求取2个子带中每一帧藏语语音信号的LPCC、MFCC参数。

(3)噪声估计及判决过程。将 2 个子带前十几帧的参数视为对背景噪声初始参数的估计，分别计算 2 个子带所有帧藏语语音信号与背景噪声之间的倒谱距离，当距离接近门限时，对噪声参数进行更新，根据得到的距离轨迹可检测藏语语音端点。

5 仿真试验

5.1 试验条件

测试藏语语音信号在安静环境下录制，以 8kHz 采样，16 位量化，并人为地以不同程度加入白噪声和汽车噪声形成带噪藏语语音。藏语语音样本取四组语料，分别由常用字母、音

节、短语、句子组成，形成安多方言藏语语音库。噪声样本选用 NOISEX-92 专业噪声库中的 white noise、volvo noise 等。按照不同的信噪比将藏语语音和噪声混合形成带噪藏语语音样本，试验中帧长 30ms（240 点），试验中藏语语音信号被分为 30ms 的帧，相邻帧有 1/2 重叠。倒谱系数采用 12 阶 LPCC、MFCC 倒谱系数。各段采样藏语语音信号通过人耳区分手工标号可作为测试各种语音端点检测正确率的标准。我们从藏语语音库（安多方言）中的字母语音库选择 34 个，其中辅音 30 个，元音 4 个。对每一信噪比的藏语语音信号作 200 次检测试验。

5.2 试验结果

1. 下面给出藏语语音 LPCC 参数的倒谱距离曲线与 MFCC 参数的倒谱距离曲线的比较

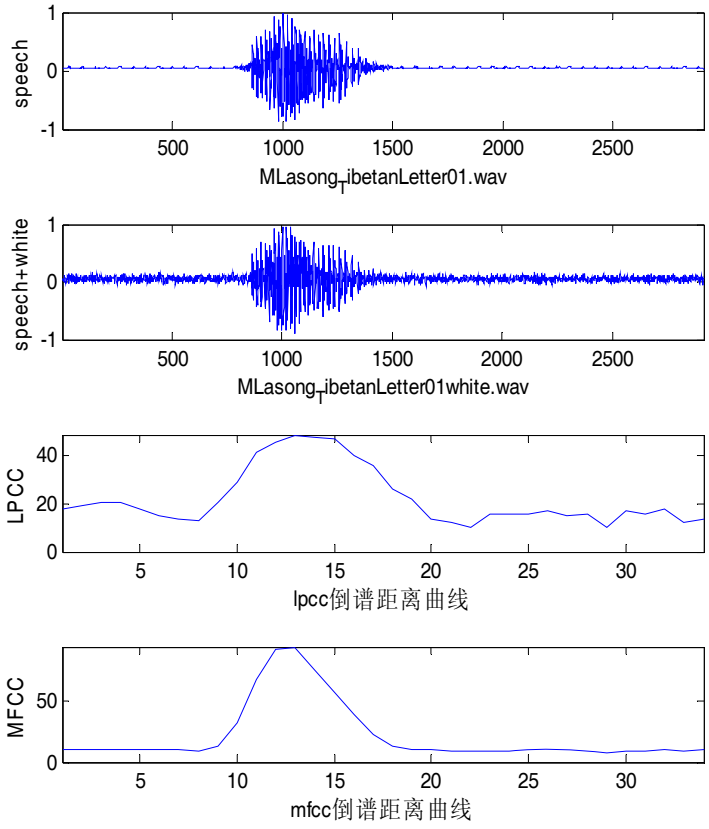


图5 分别为纯净藏语字母“ཀ”语音波形speech，加白噪声（SNR=5dB）的藏语字母“ཀ”语音波形speech+white，LPCC倒谱距离轨迹LPCC，MFCC倒谱距离轨迹MFCC

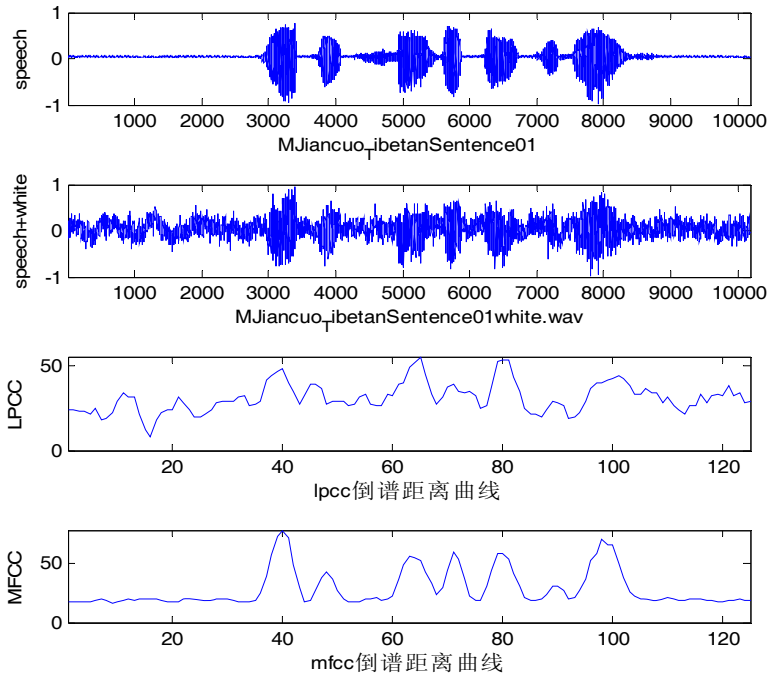


图6 分别为纯净藏语句子“ང་ཚོ་ཚང་མ་སྨྲད་ཀྱ་ཆེན་མོ་བ་ཡིན།”语音波形speech，

加白噪声（SNR=5dB）的藏语句子语音波形speech+white，

LPCC倒谱距离轨迹LPCC，MFCC倒谱距离轨迹MFCC

2. 测试的藏语语音信号采用倒谱系数的算法、传统的能量法在不同信噪比的白噪声和从NOISEX-92专业噪声库中选用volvo汽车噪声干扰下藏语语音端点测试情况见表3。表中：Energy为对数短时能量判决法；LPCC、MFCC为倒谱距离测量法；white表示白噪声；volvo表示汽车噪声。

表3 藏语端点检测试验结果

端点检测方法	检测次数	15 dB（white）	5 dB（white）	0 dB（white）	-5 dB（volvo）
Energy	检测为语音	196	152	128	140
	检测为噪音	198	120	102	100
LPCC	检测为语音	197	186	176	185
	检测为噪音	198	150	130	144
MFCC	检测为语音	198	192	184	184
	检测为噪音	198	160	140	152

从上述结果显示图和试验结果表中，可以看出藏语语音MFCC倒谱特征比LPCC倒谱特征效果更好。

6. 结论

本文进行 LPCC 和 MFCC 倒谱特征藏语语音端点检测算法设计及实现，做出 LPCC 和 MFCC 参数的倒谱距离轨迹，进行曲线的比较，反复调整参数及算法，使其取得较好的效果。

本算法主要是验证藏语语音特征优越性,LPCC特征参数没有考虑人类听觉系统对藏语语音处理的特点，而Mel频带划分是对人耳听觉特性的一种工程化模拟，MFCC在一定程度上模拟了人耳对藏语语音处理的特点。试验证明，本文研究的藏语语音MFCC参数的提取算法提取的特征参数的顽健性较好。另外，通过试验也说明了藏语语音的MFCC参数比LPCC参数更好地提高藏语语音端点检测的正确率。

参考文献

[1] Rabiner L，Juang Biing-Hwang. Fundamentals of Speech Recognition. Prentice Hall，1993，北京：清华大学出版社(影印版)，1999

[2] Huang Xuedong，Acero A，Hon H W. Spoken Language Processing. Prentice Hall、2001

[3] Wendt S，Fink G A，Kummea F. Forward Masking for Increased Robustness in Automatic Speech Recognition. in：Proc. of European Conf. on Speech Communication and Technology， Aalborg，Dhnemark，2001. 1：615-618

[4] 马学良.汉藏语概论.民族出版社,2003

[5] 敏生智、耿显宗.安多藏语会话读本.西宁:青海民族出版社,2003

[6] 胡光锐,韦晓东.基于倒谱特征的带噪语音端点检测[J].电子学报,2000,28(10):95-97