第 59 卷 第 10 期/2022 年 5 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于对数包络的汉语音节切分算法

唐维康，邵玉斌*，龙华

昆明理工大学信息工程与自动化学院，云南 昆明 650500

摘要 为提升目前连续汉语音节切分算法在噪声环境中的切分效果，基于汉语语音的对数包络特征提出一种音节切分算法，用曲线插值法获取信号包络，再经滤波和对数运算获取对数时域包络，然后采用门限判决从包络上获取极值点，根据极值点的分布特性确定音节的切分边界。所提方法对无噪汉语语音的音节切分效果优于传统方法，且在低信噪比情况下仍具有较高的切分准确率。

关键词 信号处理；音节切分；对数包络；极值点；语音包络

中图分类号 TN912.34 **文献标志码** A **DOI**：10.3788/LOP202259.1007001

Chinese Syllable Segmentation Algorithm with Logarithmic Envelope

Tang Weikang, Shao Yubin*, Long Hua

School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, *Yunnan, China*

Abstract A syllable segmentation algorithm with the logarithmic envelope feature of Chinese speech is proposed to improve the segmentation effect of existing continuous Chinese syllable segmentation algorithms in noisy environment. The voice envelope is obtained by curve interpolation, and then the logarithmic time-domain envelope is obtained by filtering and logarithmic operation. Extreme points are obtained with the threshold judgment. Finally, the syllable segmentation boundary is determined according to the distribution of extreme points. The proposed algorithm is more effective than traditional ones for syllable segmentation of Chinese speech without noises, and still has a high segmentation accuracy at low signal-to-noise ratio.

Key words signal processing; syllable segmentation; logarithmic envelope; extreme points; voice envelope

1 引 言

在语音识别领域，汉语音节的切分方法主要有语音信号突变起始点的检测方法[1]和语音信号静音段和非静音段的位置估算方法等。切分技术大致可以划分为单元边界的切分算法[2]、单元对齐的切分算法[3]，以及将单元边界和单元对齐相结合的切分算法[4]。单元对齐切分模型主要有隐马尔可夫模型（HMM）[5]和前后音素的边界模型CDBM（Context-Dependent Boundary Model）[6]；单元边界切分模型主要有神经网络模型（NN）和多层感知（MLP）模型等[7]；同时利用单元对齐和单元边界技术的模型在不同程度上提升了切分正确率[8]。语音信号的单元切分分为音素切分和音节切分，切分的语料分为合成语料和自然语料。目前，国外对于音节切分的研究不多，这主要是由西方语言的发音特性所决定的[9]。

在实际生活中，语音信号会受到人声干扰、传

收稿日期：2021-04-09；**修回日期**：2021-04-28；**录用日期**：2021-05-25

基金项目：国家自然科学基金（61761025）

通信作者：*shaoyubin@kust.edu.cn

输媒介、背景噪声的干扰，造成语音的质量和可懂度降低[10-11]，在低信噪比（SNR）的情况下，传统的音节切分算法很难从连续语音中将音节准确地切分出来[12]。如何在复杂环境中准确确认音节的切分边界成为了现代语音处理的重要研究问题之一。

近年来研究人员在语音切分方面的成果有：文献[13]利用短时能量谱中的峰值点来确定音节切分的边界，文献[14]利用局部奇异性进行音节切分，文献[15]利用小波变换法进行去噪，然后计算出每帧的分形维数，根据分形维数轨迹确定切分点。本文提出了一种基于对数包络的汉语音节切分算法，该算法先对时域包络波形进行对数计算，在对数包络上找出极值点的分布情况，剔除不需要的极值点，最终确定音节的切分边界，实现了在正常语速下的连续语音音节的切分。实验结果说明，本文方法在无噪声语音条件下的切分正确率高于短时能量、局部奇异性和分形维数等传统方法，其音节的切分正确率普遍得到较大提升，即使在加噪环境下语音音节切分正确率也较高。

2　语音信号的对数包络

汉语语音信号的时域包络曲线可以反映语音信号的特征，包络曲线可以很好地描述语音幅值的高低，也能反映出语音波形的分布情况以及语音波形起伏变化[16]，即可利用时域包络进行语音音节的切分。但在连续语音和加噪环境下，时域包络曲线的切分边界难以被找到。

根据人耳的听觉特性，语音信号的包络峰值点会映射在基底膜上，其映射的位置根据频率大小而定。频率值越大，映射位置在基底膜中越深。研究表明，人耳基底膜的音高及音强响应分布呈现对数特性[17]。虽然发声系统是按照波形来展示语音信息的，但音节是根据人耳的听觉特性来分辨的。因此，通过对时域包络进行对数运算，就可以使语音信号的振幅分布很平稳，振幅较大的地方得以压缩，振幅较小的地方得以拉升，从而使结果更符合人耳听觉系统的对数灵敏度分布。

3　音节的切分算法

如图 1 所示，本文切分算法的主要实现模块有三个：1）模块一负责时域包络的获取；2）模块二负责对数包络的获取；3）模块三负责音节切分边界的确定。

切分算法的主要步骤如下：

1）模块一：利用极值法和光滑插值法获取语音信号的时域包络。

2）模块二：利用低通滤波和取对数运算获取时域对数包络。

3）模块三：在对数包络上获取极值点，再利用单门限法和阈值法剔除不必要的极值点，获得汉语语音音节的最佳切分边界。

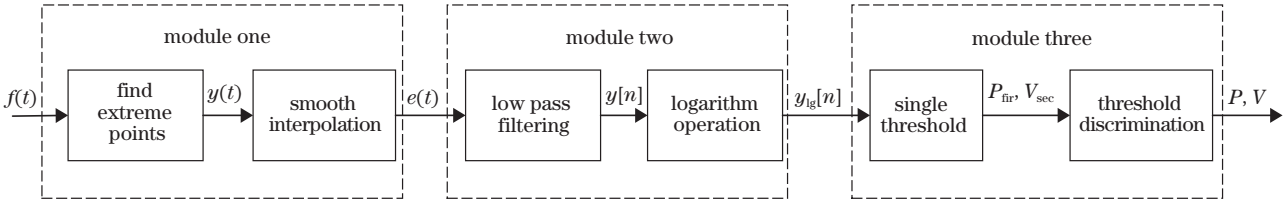

图 1　音节切分算法步骤

Fig. 1　Steps of syllable segmentation algorithm

3.1　时域包络的获取算法

语音原始波形 $f(t)$ 可以看成是一系列时变过程，由于声道和声门的相互作用，语音信号表现出了非线性特性。包络 $e(t)$ 可以看成是紧贴语音信号 $f(t)$ 上半部分的曲线，将语音包于波形 $e(t)$ 之下。如图 2 所示，先对语音波形 $f(t)$ 进行一阶求导，一阶导数的零点就是极值点，通过对极大值点进行插值，即可得到上包络波形 $e(t)$，本文只针对语音信号的上包络进行分析。

如图 2 所示，语音波形 $f(t)$ 在 $[a,b]$ 区间上连续，可表示为

$$y(t)=\frac{\mathrm{d}f(t)}{\mathrm{d}t}, \tag{1}$$

式（1）的零点就是 $f(t)$ 的极值点。在求解 $y(t)$ 的过程中，$y(t)$ 可以用差商来表达，这可使复杂的求导计算简化，减少计算量。利用式（1）即可求出语音原始波形 $f(t)$ 的极值点：当满足 $f(t+\Delta t)-2f(t)+f(t-\Delta t)\leqslant 0$ 条件时，t 为极大值点；当满足

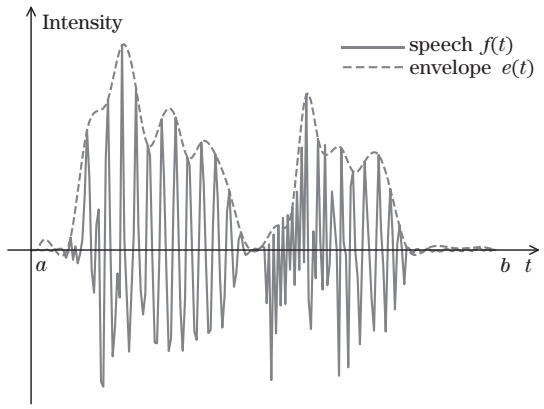

图 2　语音信号的时域波形和上包络波形

Fig. 2　Waveform and its upper envelope of speech signal

$f(t+\Delta t)-2f(t)+f(t-\Delta t)>0$ 条件时，t 为极小值点。

通过计算求出原始语音波形 $f(t)$ 的极大值点为 $t_1, t_2, \cdots, t_n$，与之对应的函数值分别为 $f(t_1)$，$f(t_2), \cdots, f(t_n)$。这里需要估算出区间 $[t_k, t_{k+1}]$（$k=1, 2, \cdots, n-1$）上的插值函数，使其接近真实函数 $f(t)$ 的包络，插值函数常用阶梯插值、线性插值和曲线插值确定，如图 3 所示，得到的语音波形为普通话“联合”两字发音波形的正半部分。

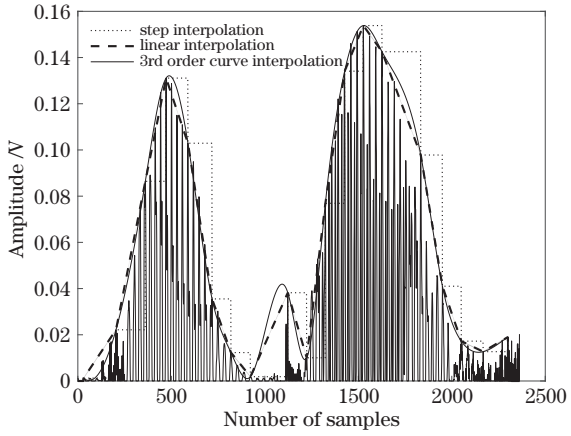

图 3　多种插值法形成的包络线

Fig. 3　Some envelopes formed by various interpolation methods

实验结果的准确性与极值点的提取密切相关，理想情况是一个音节对应一个极大值点，将该极大值点左右两侧的极小值点作为分割边界，插值方法的不同将会影响极值点的准确提取。由图 3 可以看出，阶梯插值出现了很多跳变段，并没有紧贴语音信号的上半部分。对于一个音节，采用阶梯插值将会得到多个极大值点，使切分边界的判断有误。线性插值形成的包络由各个折线段连接，虽然使得曲线收敛了，但光滑性较差，拐点较多。三阶曲线插值使用低阶多项式减小了误差，插值后的包络线更加光滑，这有利于极值点的提取，而更高阶的曲线插值虽然也能使包络线更加光滑，但计算量较大，同时还会出现吉布斯效应。因此本文算法采用了三阶曲线插值。

在区间 $[t_k, t_{k+1}]$ 上可以确定两个值 $f(t_k)=h_k$ 和 $f(t_{k+1})=h_{k+1}$，同时可以选择一个三次多项式函数：

$$e(t)=e_0+e_1(t-t_k)+e_2(t-t_k)^2+e_3(t-t_k)^3, \tag{2}$$

式中：e_0、e_1、e_2 和 e_3 为系数。

根据 Akima 曲线插值法[18]，h_k 和 h_{k+1} 可分别表示为

$$\begin{cases} h_k=\dfrac{\left|g_{k+1}-g_k\right|g_{k-1}+\left|g_{k-1}-g_{k-2}\right|g_k}{\left|g_{k+1}-g_k\right|+\left|g_{k-1}-g_{k-2}\right|} \\ h_{k+1}=\dfrac{\left|g_{k+2}-g_{k+1}\right|g_k+\left|g_k-g_{k-1}\right|g_{k+1}}{\left|g_{k+2}-g_{k+1}\right|+\left|g_k-g_{k-1}\right|} \end{cases}, \tag{3}$$

其中

$$g_k=\frac{f(t_{k+1})-f(t_k)}{t_{k+1}-t_k}。 \tag{4}$$

将端点代入式(4)可得

$$\begin{cases} g_0=2g_1-g_2 \\ g_{-1}=2g_0-g_1 \\ g_n=2g_{n-1}-g_{n-2} \\ g_{n+1}=2g_n-g_{n-1} \end{cases}。 \tag{5}$$

由于式(3)中分母不能为零，所以 Akima 曲线插值法规定：当 $g_{k+1}-g_k=0$ 与 $g_{k-1}-g_{k-2}=0$ 时，$h_k=(g_{k-1}+g_k)/2$；当 $g_{k+2}-g_{k+1}=0$ 与 $g_k-g_{k-1}=0$ 时，$h_{k+1}=(g_k+g_{k+1})/2$。最终得出三次多项式在区间 $[t_k, t_{k+1}]$（$k=1, 2, \cdots, n-1$）的多项式系数为

$$\begin{cases} e_0=f(t_k) \\ e_1=h_k \\ e_2=(3g_k-2h_k-h_{k+1})/(t_{k+1}-t_k) \\ e_3=(h_{k+1}+h_k-2g_k)/(t_{k+1}-t_k)^2 \end{cases}。 \tag{6}$$

图 4(a)为原始语音波形。通过式(2)、(6)即可获取在区间$[t_k, t_{k+1}]$($k=1, 2, \cdots, n-1$)的任意一位置t_i($i=1, 2, \cdots, n-1$)的近似值$e(t)$,$e(t)$就是经过曲线插值后形成的时域包络图,如图 4(b)所示。

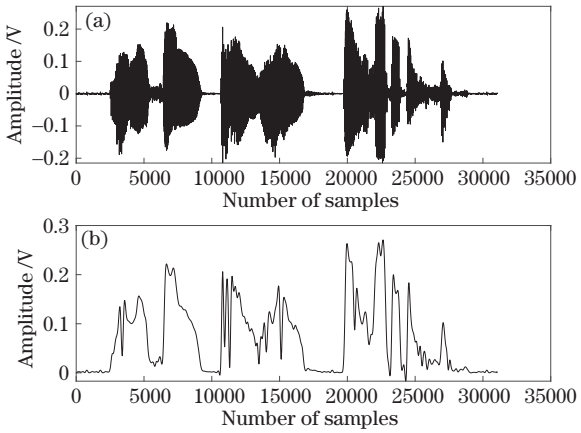

图 4　原始语音波形和插值后的包络波形。(a)原始语音波形;(b)插值后的包络波形

Fig. 4　Original speech waveform and envelope after interpolation. (a) Original speech waveform; (b) envelope after interpolation

从图 4(b)可以看出,$e(t)$可以很好地反映原始语音信号的特征,但$e(t)$会出现很多毛刺,这对本文的实验会造成很大干扰,需要进行进一步的处理。

3.2　对数包络的获取算法

对$e(t)$进行插值以后,还需要进行滤波处理,巴特沃斯滤波器具备通带最平坦和阻带下降光滑特性。本文采用了巴特沃斯滤波器,其系统传递函数[19]为

$$H(z)=\frac{b_0+b_1z^{-1}+\cdots+b_Mz^{-M}}{a_0+a_1z^{-1}+\cdots+a_Nz^{-N}},\tag{7}$$

式中:a_0~a_N和b_0~b_M为系数。对式(7)作z域逆变换,得到差分方程为

$$a_0y[n]+a_1y[n-1]+\cdots+a_Ny[n-N]=b_0e[n]+b_1e[n-1]+\cdots+b_Me[n-M],\tag{8}$$

式中:$e[n]$是$e(t)$的采样离散序列。令$a_0=1$,得到数字滤波后的输出为

$$y[n]=-\sum_{K=1}^{N}a_Ky[n-K]+\sum_{K=0}^{M}b_Ke[n-K],\tag{9}$$

式中:$y[n]$是滤波后的输出,如图 5(a)所示。$y[n]$经对数运算后形成的对数包络为$y_{lg}[n]=20\lg y[n]$,如图 5(b)所示。

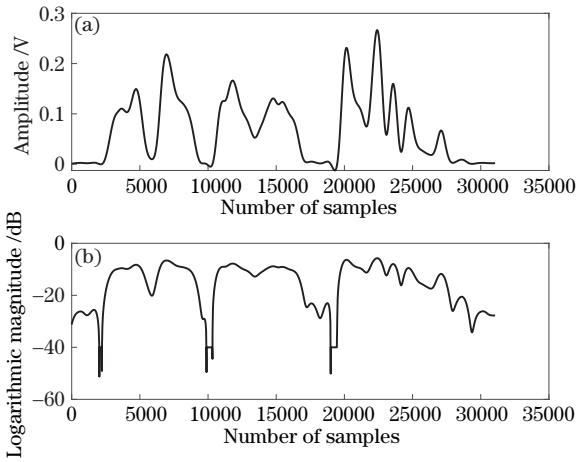

图 5　滤波包络和对数包络。(a)滤波包络;(b)对数包络

Fig. 5　Filtered envelope and logarithmic envelope. (a) Filtered envelope; (b) logarithmic envelope

3.3　音节切分边界的确定

求解出对数包络$y_{lg}[n]$的极大值序列$P_{fir}=\{p_{fir1}, p_{fir2}, \cdots, p_{firm}\}$和极小值序列$V_{fir}=\{v_{fir1}, v_{fir2}, \cdots, v_{firn}\}$,如图 6(a)所示,极大值用加号标识,极小值用圆点标识。由图 6(a)可以看出,该语音的对数包络在−20 dB 以下会出现极小值点。本文提出单门限法,设门限值为T_{min},经多次实验得出T_{min}为对数包络的平均分贝值的 20% 为宜,只有满足$v_{firi}\geqslant T_{min}$($i=1, 2, \cdots, n$)的极小值点才保留,生成的

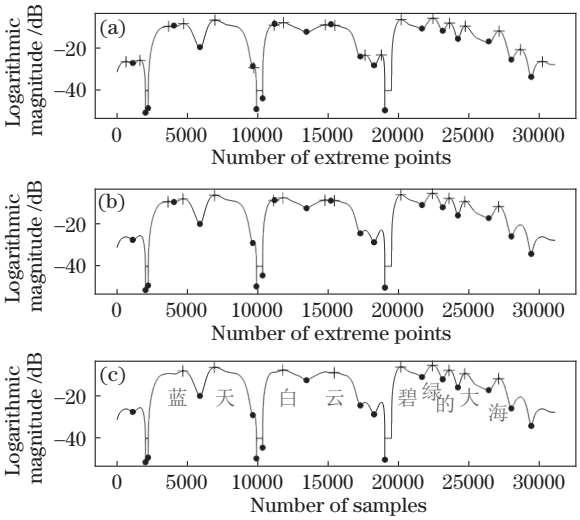

图 6　对数包络上的极值点调整情况。(a)对数包络上的极值点;(b)单门限处理后的极值点;(c)阈值处理后的极值点

Fig. 6　Adjustment of extreme points on logarithmic envelope. (a) Extreme points on logarithmic envelope; (b) extreme points after single threshold processing; (c) extreme points after thresholding

极小值序列 $V_{sec}=\{v_{sec1}, v_{sec2}, \cdots, v_{secn}\}$，如图 6(b)所示，同时也可以看出：1)多个极大值点对应一个音节；2)部分极大值点与极小值点太过于接近。接下来本文提出阈值处理法，设阈值为 T_{dis}，极大值点 P_{fir} 的纵轴坐标集合为 $H_p=\{h_{p1}, h_{p2}, \cdots, h_{pn}\}$，极小值点 V_{sec} 的纵轴坐标集合为 $H_v=\{h_{v1}, h_{v2}, \cdots, h_{vn}\}$，极大值与极小值的纵轴距离 $D_i=|h_{pi}-h_{vi}|(i=1, 2, \cdots, n)$。当 $D_i \leqslant T_{dis}$，将极大值点与极小值点同时删除，最终生成极大值点序列 $P=\{p_1, p_2, \cdots, p_n\}$，极小值点序列 $V=\{v_1, v_2, \cdots, v_n\}$，如图 6(c)所示。

从图 6(c)可以看出一个极大值点就对应一个音节，极大值点左右两边的极小值点为切分边界。图 6 的横坐标是采样点数，采样点数并不能作为切分边界，需要转换为时间坐标，切分边界的时间由采样点数除以采样率的值确定。

4　实验分析

本文采用中央人民广播电台“中国之声”的语音数据建立数据库，用于分析，该数据库由 176 段汉语语音组成，每段语音的时长为 12 s。实验中采用采样频率为 8 kHz、量化精度为 16 bit 的单通道 wav 格式音频文件。对于不同的算法，采用相同的实验数据进行测试。

4.1　实验一

为选取合适的滤波器对插值后的包络数据进行滤波，固定滤波器的阶数为 3 阶，根据音节包络的波动情况(一般语速为每秒 3～5 个音节)，可选择滤波器的截止频率范围在 5～15 Hz 之间，本文采用的截止频率为 8 Hz，此时有较好的效果。在不同信噪比下采用了切比雪夫、贝塞尔和巴特沃斯滤波器进行切分正确率的实验。

本实验的语音数据均有与之对应的文本信息，用于验证切分后语音的正确率。语音音节切分的正确率定义为

$$\eta=\frac{|n_1-n_2|}{n_2}, \tag{10}$$

式中：η 为汉语语音音节的切分正确率；n_1 为切分后的语音音节数量；n_2 为语料中真实存在的音节数量。

从表 1 可以看出，切比雪夫滤波器的切分正确率最低，其原因在于：通带中出现波纹，从而产生多余的极值点。贝塞尔滤波器的阻带下降波动对噪声没有进行很好的抑制。巴特沃斯滤波器通带最平坦，这有利于极值点的提取，同时其阻带下降比贝塞尔滤波器的光滑，可以较好地抑制噪声，切分的正确率也最高。

表 1　不同信噪比下不同滤波器的切分正确率

Table 1　Segmentation accuracy of different filters for different signal-to-noise ratio　　unit: %

Filter type	SNR /dB			
	20	15	10	5
Chebyshev filtering	79.6	78.3	77.1	69.4
Bessel filtering	81.4	80.6	79.3	74.8
Butterworth filtering	91.7	91.0	90.6	87.5

4.2　实验二

为了更好地分析切分边界的准确性，本文从测试语料中截取了一段语音，语音内容为“由国外以航空货运方式寄物回国”。本实验展示了在无噪语音和噪声干扰环境下的音节切分情况，如图 7 所示。图 7(a)是原始的语音波形图，图 7(b)是无噪语音的切分边界，图 7(c)是加入了高斯白噪声后信噪比为 5 dB 的波形图，图 7(d)是本文方法生成的切分

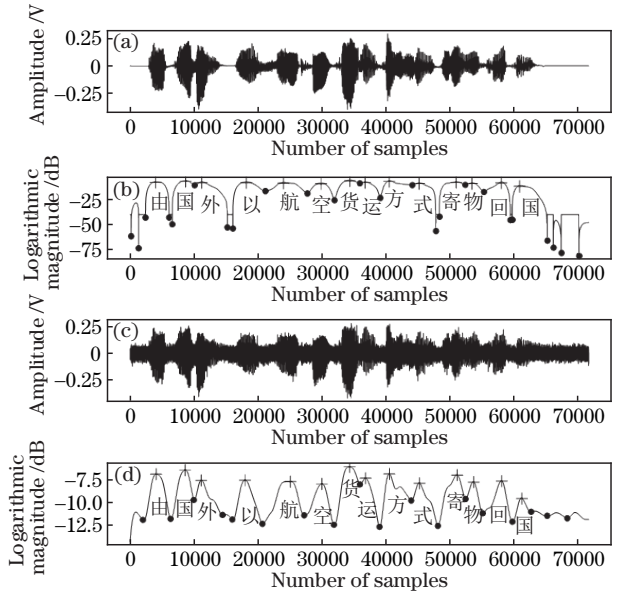

图 7　加噪环境下的汉语语音音节切分。(a)原始语音波形；(b)纯净语音对数包络上的极值点分布；(c)加噪后的语音波形(SNR 为 5 dB)；(d)加噪语音对数包络上的极值点分布

Fig. 7　Chinese phonetic syllable segmentation under noisy environment. (a) Original speech; (b) distribution of extreme points on logarithmic envelope of speech without noise; (c) speech with noises (SNR is 5 dB); (d) distribution of extreme points on logarithmic envelope of speech with noise

边界。

从图 7(b)中可以看出，在无噪语音环境下，根据对数包络上的极值点分布情况可以正确地将音节进行切分。从图 7(c)中可以看出，当SNR为5 dB时，噪声已经对原始的汉语语音信号产生较大干扰，传统方法在信噪比比较低的情况下会把部分噪音误判为汉语语音的音节。从图 7(d)中可以看出，对数包络上的极值点作为语音信号音节的切分边界与人工手动切分的音节边界相差不大，本文方法取得了较好的语音音节切分效果，在低信噪比环境下也能够将连续语音信号的音节进行切分，从而满足了工程上的需求。

4.3　实验三

为使实验结果更加有对比性：一方面，在无噪语音的条件下，分别采用奇异指数、局部奇异性、短时能量、分形维数和本文方法进行切分正确率的验证；另一方面，在语音信号中添加高斯白噪声，使其信噪比为 10 dB、5 dB、0 dB、−5 dB 和−10 dB，再分别采用多种音节切分方法与本文方法进行切分正确率的验证。

在无噪语音条件下，采用的多种音节切分算法与本文方法得到的音节切分正确率如表 2 所示。

表 2　无噪语音条件下不同算法的切分正确率

Table 2　Segmentation accuracy of different algorithms without noise

Segmentation algorithm	Accuracy /%
Singularity index[15]	26.9
Local singularity[13]	50.3
Short-term energy[12]	76.9
Fractal dimension[14]	82.3
Method of this article	92.1

从表 2 可以看出：在无噪声的条件下奇异指数算法的音节切分正确率低于 30%，短时能量法和分形维数法的音节切分正确率在 80% 左右，本文方法与奇异指数、局部奇异性、短时能量、分形维数法相比较，其汉语语音的音节切分正确率都有较大提升，本文方法的音节切分正确率在 90% 左右。在多种信噪比环境下，不同算法获得的语音音节的切分正确率如表 3 所示。

从表 3 可以看出，当SNR为10 dB和5 dB时，本文方法的音节切分正确率相比传统切分算法有较大的提升。当SNR为0 dB和−5 dB时，分形维数法的音节切分正确率高于本文方法。在不同的信噪比下，本文方法的语音音节切分正确率均保持在 70% 以上，在语音信号的信噪比较低的情况下，本文方法的音节切分效果仍然不错。

表 3　不同信噪比下不同算法的语音切分正确率

Table 3　Speech segmentation accuracy of different algorithms under different signal-to-noise ratio

unit: %

Segmentation algorithm	SNR /dB			
	10	5	0	−5
Singular index[15]	24.7	20.6	18.7	16.9
Local singularity[13]	49.3	47.7	42.6	39.1
Short-term energy[12]	69.6	60.3	58.4	54.2
Fractal dimension[14]	83.4	80.1	79.7	71.2
Method of this article	90.6	87.5	76.3	70.6

4.4　实验四

为了验证算法的运行时间，在同一信噪比(SNR为10 dB)、同一实验平台下，将切分后的音节生成 wav 文件保存，计算本文方法和多种音节切分方法所用的时间，所得结果如表 4 所示。

表 4　不同时长下不同算法的切分时间

Table 4　Segmentation time of different algorithms under different durations

unit: s

Segmentation algorithm	Voice duration /s			
	4	6	8	10
Singular index[15]	1.3	2.2	3.9	5.1
Local singularity[13]	2.9	4.1	5.7	7.4
Short-term energy[12]	3.1	4.6	5.9	7.6
Fractal dimension[14]	1.9	2.7	4.5	6.0
Method of this article	2.5	3.7	5.4	6.9

从表 4 可以看出，本文算法的运行时长短于局部奇异性和短时能量算法切分的时长，其原因在于局部奇异性和短时能量切分算法先进行了端点检测，区分出静音段和非静音段后才进行切分。本文算法的运行时长长于奇异指数和分形维数切分算法，其原因在于本文算法求出音节的极值点后还进行了单门限法和阈值法的处理，而奇异指数和分形维数切分算法确定音节的特征轨迹后并没有进行进一步的处理。

根据以上 4 个实验结果，本文方法在低信噪比的噪声环境下进行语音切分是有效的，其切分正确率和程序运行时长较好地满足了工程应用。本文算法较其他算法在切分准确率上有所提升，并且易于实现。

5 结　论

在实际说话环境中，传统汉语语音切分技术对正常语速的汉语音节的切分精度不高。针对这一问题，本文算法利用光滑插值法提取汉语语音的时域包络，并依据人耳听觉的对数特性，在时域包络上进行滤波并取对数运算，得到对数包络，最后利用门限法和阈值法在对数包络上找出合适的极值点，从而由这些极值点的分布确定汉语音节的起点和终点。在无噪语音和低信噪比的语音环境中的实验表明，本文算法能较好地满足工程应用需求，有着较高的切分正确率。

参　考　文　献

[1] Faraji N, Ahadi S M, Sheikhzadeh H. Sequential method for speech segmentation based on Random Matrix theory[J]. IET Signal Processing, 2013, 7(7): 625-633.

[2] Baby A, Prakash J J, Subramanian A S, et al. Significance of spectral cues in automatic speech segmentation for Indian language speech synthesizers [J]. Speech Communication, 2020, 123: 10-25.

[3] Kiss G, Sztahó D, Vicsi K. Language independent automatic speech segmentation into phoneme-like units on the base of acoustic distinctive features[C]// 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom), December 2-5, 2013, Budapest, Hungary. New York: IEEE Press, 2013: 579-582.

[4] Jeyalakshmi K, Anitha J. Ontology based data unit similarity with combining tag and value for data extraction and alignment[J]. International Journal of Innovative Research & Development, 2013, 2(10): 120-143.

[5] 任凯龙, 汪毅, 陈晓冬, 等. 用于腹腔镜扶持器控制的特定人语音识别算法[J]. 激光与光电子学进展, 2020, 57(18): 181702.
Ren K L, Wang Y, Chen X D, et al. Speaker-dependent speech recognition algorithm for laparoscopic supporter control[J]. Laser & Optoelectronics Progress, 2020, 57(18): 181702.

[6] 王丽娟, 曹志刚. TTS语音单元边界的自动切分[J]. 微电子学与计算机, 2005, 22(12): 8-11.
Wang L J, Cao Z G. Automatic segmentation for TTS units[J]. Microelectronics & Computer, 2005, 22(12): 8-11.

[7] Sivaram G S V S, Hermansky H. Sparse multilayer perceptron for phoneme recognition[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(1): 23-29.

[8] Akdemir E, Ciloglu T. HMM topology for boundary refinement in automatic speech segmentation[J]. Electronics Letters, 2010, 46(15): 1086-1087.

[9] 杜守栓. 方言口音普通话语音自动切分算法研究[D]. 北京: 中国科学院计算技术研究所, 2006: 40-51.
Du S S. Research on robust automatic segmentation of dialectal speech[D]. Beijing: Institute of Computing Technology, Chinese Academy of Sciences, 2006: 40-51.

[10] 方斌, 陈家益. 去除脉冲噪声的小波阈值去噪算法[J]. 激光与光电子学进展, 2021, 58(22): 2210016.
Fang B, Chen J Y. Wavelet threshold denoising algorithm for removing impulse noise[J]. Laser & Optoelectronics Progress, 2021, 58(22): 2210016.

[11] 化春键, 马金科, 陈莹. 基于差异哈希算法的改进非局部均值去噪算法[J]. 激光与光电子学进展, 2020, 57(14): 141007.
Hua C J, Ma J K, Chen Y. Improved non-local mean denoising algorithm based on difference hash algorithm[J]. Laser & Optoelectronics Progress, 2020, 57(14): 141007.

[12] 夏令祥. 低信噪比环境下语音端点检测方法的研究[D]. 徐州: 中国矿业大学, 2019: 20-28.
Xia L X. Study on the voice activity detection method in low SNR environment[D]. Xuzhou: China University of Mining and Technology, 2019: 20-28.

[13] Mary L, Antony A P, Babu B P, et al. Automatic syllabification of speech signal using short time energy and vowel onset points[J]. International Journal of Speech Technology, 2018, 21(3): 571-579.

[14] Khanagha V, Daoudi K, Pont O, et al. Phonetic segmentation of speech signal using local singularity analysis[J]. Digital Signal Processing, 2014, 35: 86-94.

[15] Ţălu Ş, Kulesza S, Bramowicz M, et al. Fractal geometry of internal thread surfaces manufactured by cutting tap and rolling tap[J]. Manufacturing Letters, 2020, 23: 34-38.

[16] Somers B, Verschueren E, Francart T. Neural tracking of the speech envelope in cochlear implant users[J]. Journal of Neural Engineering, 2019, 16(1): 016003.

[17] Ben Messaoud M A, Bouzid A. Pitch estimation of speech and music sound based on multi-scale product with auditory feature extraction[J]. International Journal of Speech Technology, 2016, 19(1): 65-73.

[18] 竺明星, 尹倩, 龚维明, 等. 基于Akima插值理论的水平试桩数据处理方法研究[J]. 岩土工程学报, 2020, 42(S1): 80-84.
Zhu M X, Yin Q, Gong W M, et al. Data processing method for laterally loaded trial piles based on Akima interpolation theory[J]. Chinese Journal of Geotechnical Engineering, 2020, 42(S1): 80-84.

[19] 汪宇, 查明, 李纵, 等. 巴特沃思型低通滤波器的归一化设计[J]. 舰船电子工程, 2018, 38(1): 61-64.
Wang Y, Zha M, Li Z, et al. Normalized design and application of butterworth low-pass filter[J]. Ship Electronic Engineering, 2018, 38(1): 61-64.