文章编号: 1004-4353(2019)02-0128-08

朝鲜语语音音节自动切分算法的研究

李洺宇, 金小峰*
(延边大学 工学院, 吉林 延吉 133002)

摘要: 针对目前语音语料人工标注效率低的问题,提出了一种朝鲜语连续语音语料的音节自动切分方法. 该方法首先采用 Seneff 听觉模型提取音频的包络检测响应和广义同步检测响应等特征参数,其次结合朝鲜语发音特点确定音节的候选边界位置,最后通过静音段和摩擦音检测消除虚假边界,以提高边界检测的准确率. 实验结果表明,该朝鲜语语音语料音节自动切分方法的准确率(93.56%)比传统的基于 Seneff 听觉模型的分割算法提高了 14.59%,召回率(86.43%)比传统的基于 Seneff 听觉模型的分割算法降低了 1.69%;因此,本文算法总体优于传统的基于 Seneff 听觉模型的分割算法.

关键词: 朝鲜语语音语料; 语料自动标注; Seneff 听觉模型; 语音音节分割
中图分类号: TP391.4 **文献标识码**: A

Research on automatic segmentation algorithm of Korean speech syllables

LI Mingyu, JIN Xiaofeng*
(*College of Engineering*, *Yanbian University*, *Yanji* 133002, *China*)

Abstract: Aiming at the current low efficiency of manual annotation of speech corpus, an automatic syllable segmentation method for Korean continuous speech corpus is proposed. First, Seneff auditory model is used to extract the audio characteristic parameters, such as the envelope detection response and generalized synchronous detection response, etc. Secondly, the candidate boundary position of syllables is defined according to the Korean pronunciation characteristics. Finally, the false boundary is eliminated by silent segment and fricative detection to improve the boundary detection accuracy. The experimental results show that the accuracy of the proposed Korean syllable segmentation method is 93.56%, increased by 14.59% than that of traditional segmentation algorithms based on Sneff auditory model, meanwhile, the recall rate reaches to 86.43%, decreased by 1.69%. Therefore, the proposed algorithm in this paper is overall superior to traditional segmentation algorithms based on Sneff auditory model.

Keywords: Korean speech corpus; automatic segmentation; Seneff auditory model; syllable segmentation

0 引言

近年来,语音技术得到了快速的发展和应用. 语料库作为语音技术研究的底层,对语音识别、语音合成、语音信号处理等具有重要的支撑作用. 传统的语音语料标注采用的是人工标注方法,需耗费大量人力和时间[1],且已无法满足语音语料日益增多的需求,因此研究语音自动标注算法变得尤为必要. 目前,语音语料自动标注的方法主要分为基于语音识别的方法和非语音识别的方法. 例如:王丽娟等[2]提出了

收稿日期: 2019-02-17 * **通信作者**: 金小峰(1970—),男,教授,研究方向为智能信息处理.
基金项目: 吉林省教育厅"十三五"科学技术项目(JJKH20191126KJ);延边大学世界一流学科建设培育项目(18YLPY14)

一种基于 HMM 模型的语音单元边界自动切分方法；李诗心[3]针对傣语语音合成系统提出了一种自动分词技术与音子自动切分技术；韩虎[4]研究了一种汉语连续语音的音节自动标注算法；Tolegen Gulmira 等[5]和 Jaruwat Pailai 等[6]将音素的边界用其他标签表示，然后以标签作为检测目标，即将边界检测任务视为序列标注任务. 上述基于语音识别的方法，其音节切分准确率较高，但依赖于训练好的语音模型.

除了基于语音识别的方法外，研究人员还从构成语音单元(元音和辅音)的特征区分角度，提出了多特征参数组合的语音单元边界检测方法. 例如：文献[7-8]依据不同的特征参数对语音单元进行区分，取得了较好的分割效果；王艳等[9]提出了一种基于元音检测的汉语连续语音声韵母的分割方法，这种方法具有较好的抗噪性；基于元音的能量远大于辅音的特点，姚徐等[10]利用双门限的方法设计并实现了音段自动切分系统，但该方法对元音的识别效果很大程度受阈值选取的影响；陈斌等[11]104 提出了一种基于 Seneff 听觉谱特征的汉语连续语音声韵母边界检测方法，该方法对语音单元尤其是音素切分有较好的效果. 基于文献[11]的研究，本文采用 Seneff 听觉模型，结合朝鲜语的发音特点提出一种朝鲜语语音语料音节自动切分方法，并通过实验验证本文方法的有效性.

1　朝鲜语发音特点

朝鲜语文字组字规则以音节为单位，一个音节组成一个朝鲜语文字. 朝鲜语的音节由初声子音(声母)、中声音(韵母)和终声子音(韵尾) 3 个部分组成，分为声母＋韵母＋韵尾、声母＋韵母、韵母 3 种形式[12]. 在朝鲜语语音语流中，一个语音往往受邻近语音的影响而发生音变[13]，这些音变往往会给朝鲜语语音音节的检测和分割带来负面影响. 常见的朝鲜语发音变化主要分为如下几种：

1)连音化. 语流中闭音节跟元音开头的音节相连时，前面闭音节的韵尾(韵尾ㅇ/ㆆ除外)为后面音节的头音. 如果是单韵尾闭音节，韵尾直接成为后面元音开头音节的头音；如果是复韵尾闭音节，复韵尾的右边辅音成为后面元音开头音节的头音，左边辅音不管原先是否发音，都成为该闭音节的发音韵尾.

2)紧音化. 以代表音读为ㄱ的ㄲ/ㅋ/ㄳ/ㄺ、ㄷ的ㅅ/ㅆ/ㅈ/ㅊ/ㅌ、ㅂ的ㅍ/ㅥ/ㅦ/ㅄ的收音之后出现的ㄱ/ㄷ/ㅂ/ㅅ/ㅈ要发生紧音化；词干的收音ㄴ(ㄵ)/ㅁ(ㄻ)之后所接的词尾部分的第 1 个音节的辅音是ㄱ/ㄷ/ㅅ/ㅈ时，这些松音辅音会发生紧音化；词干的收音ㅥ/ㅀ之后所接的词尾部分的第 1 个音节的辅音是ㄱ/ㄷ/ㅅ/ㅈ时，这些松音辅音会发生紧音化；收音ㆆ(包括ㄶ和ㅀ)在碰到辅音ㅅ时，ㆆ会与ㅅ结合成为紧音ㅆ；在词尾中出现的收音ㄹ会引起后面的ㄱ/ㄷ/ㅂ/ㅅ/ㅈ发生紧音化.

3)送气音化. 送气音也叫吐气音，分为两种情况：收音ㆆ/ㄶ/ㅀ与以ㄱ/ㄷ/ㅈ为头音的音节相连时，ㄱ/ㄷ/ㅈ分别变成送气音ㅋ/ㅌ/ㅊ；收音ㄱ/ㄷ/ㅂ/ㅈ与以ㆆ为头音的音节相连时，受ㆆ的影响变为送气音ㅋ/ㅌ/ㅍ/ㅊ.

4)腭化. 以ㄷ/ㅌ/ㅀ为收音的词素后接以ㅣ为头音的助词或词缀时，变为ㅈ/ㅊ.

5)辅音同化. 两个不同的辅音遇到一起时，变为相似或者相同的音，这种同化现象分为鼻音化和ㄹ音化现象两类. ①鼻音化. 当发音为ㄱ/ㄷ/ㅂ的收音后接以辅音ㅁ/ㄴ开头的音节时，收音的发音要相应地变为ㅇ/ㄴ/ㅁ. 当收音ㅁ/ㅇ与以辅音ㄹ开头的音节相遇时，辅音ㄹ的发音自动变为ㄴ. 当收音ㅂ/ㄱ与以辅音ㄹ开头的音节相遇时，辅音ㄹ的发音自动变为ㄴ；前面的ㅂ/ㄱ相应地变成鼻音ㅁ/ㅇ. ②ㄹ音化现象. 在收音ㄴ的后面或辅音ㄴ的前面出现ㄹ时，ㄴ的发音自动变为收音ㄹ或是辅音ㄹ.

2　Seneff 听觉模型

Seneff 听觉模型模拟了人耳耳蜗结构，具备良好的语音频带处理性能，被广泛应用于说话人识别、语音识别、语音信号处理等领域[11]105. Seneff 听觉模型框架如图 1 所示.

Seneff 听觉模型在处理语音信号时,首先利用滤波器预处理听觉信号(去除高低频),然后将预处理后的输出信号输入到个数为 35 的临界频带滤波器组.因为临界频带滤波器在高频段具有良好的时间分辨率,在低频段具有良好的频率分辨率,因此可以提高共振峰信息提取的准确率.临界频带滤波器的主要参数为临界频带带宽的频率尺度,其求解方式为通过非线性映射函数将频率尺度转换为 Bark 尺度.一个 Bark 的频率差为一个临界带宽,每个相邻滤波器的频率上下限采用式(1)计算:

$$B(f)=\begin{cases}0.01f, & 0\leqslant f<500;\\ 0.007f+1.5, & 500\leqslant f<1\,200;\\ 6\ln f-32.6, & f\geqslant 1\,200.\end{cases}\quad(1)$$

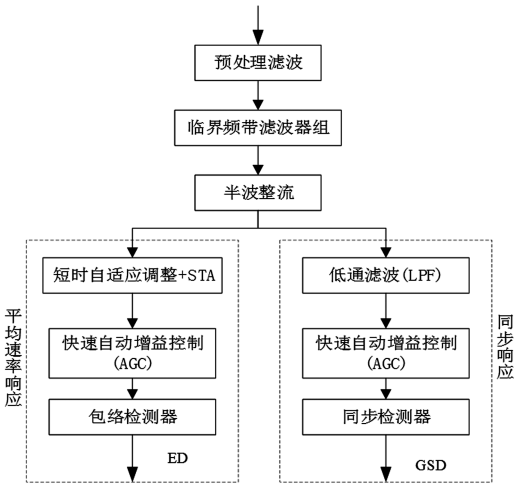

图 1 Seneff 听觉模型框架

利用公式(1)可求得临界频带滤波器组的参数值(中心频率),具体计算过程如下:以中心频率 f_0 先求出 $B_0=B(f_0)$,然后倒转得到频率尺度中的滤波器上下限 $f(B_0-1/2)$ 和 $f(B_0+1/2)$,每个上下限相邻间隔大约为半个临界带宽.计算得到的中心频率见表 1.

表 1 临界频带滤波器中心频率取值 Hz

编号	1	2	3	4	5	6	7	8	9
中心频率	228.6	275.9	326.5	381.0	432.4	484.8	533.3	592.6	640.0
编号	10	11	12	13	14	15	16	17	18
中心频率	695.7	761.9	842.1	941.2	1 000.0	1 066.7	1 142.9	1 230.8	1 333.3
编号	19	20	21	22	23	24	25	26	27
中心频率	1 391.3	1 454.5	1 523.8	1 600.0	1 684.2	1 777.8	1 882.4	2 000.0	2 133.3
编号	28	29	30	31	32	33	34	35	
中心频率	2 285.7	2 461.5	2 666.7	2 904.1	3 177.7	3 495.2	3 866.6	4 303.4	

利用式(2)对通过各临界频带滤波器的输出信号进行具有饱和非线性的半波整流.

$$R_i(n)=\begin{cases}G\{1+A\tan^{-1}[B\cdot CB_i(n)]\}, & CB_i(n)>0;\\ G\,\mathrm{e}^{A\cdot B\cdot CB_i(n)}, & CB_i(n)\leqslant 0.\end{cases}\quad(2)$$

式(2)中 $CB_i(n)$ 为临界频带滤波器的输出,$G=2.35$,$A=10$,$B=65$.显然,从公式(2)的分段情况可知:对于小输入值,可进行线性处理;对于大输入值,可进行压缩处理.

信号经半波整流后,系统被分成两个分支:一个分支用以求解平均速率响应,另一个分支用以求解同步响应.平均速率响应从短期自适应和正向掩蔽(STA,short term adaptation)模块开始,然后依次为自动增益(AGC,automatic gain control)模块、包络检测器(ED,envelope detector).同步响应路径依次为低通滤波器(LPF,low-pass filter)、AGC 和广义同步检测器.

STA 模块模拟的是在耳蜗反应中发生的短期适应效应和正向掩蔽效应,这两种效应影响神经递质浓度的机制可用如下公式表示:

$$\frac{\mathrm{d}C_i(n)}{\mathrm{d}n}=\begin{cases}\mu_a[R_i(n)-C_i(n)]-\mu_b C_i(n), & C_i(n)<R_i(n);\\ -\mu_b C_i(n), & C_i(n)\geqslant R_i(n);\end{cases}\quad(3)$$

$$STA_i(n)=\max\{0,\mu_a[R_i(n)-C_i(n-1)]\}.\quad(4)$$

其中,$C_i(n)$ 为区域内神经递质的浓度,$R_i(n)$ 为输入(源区域)的浓度,常数 $\mu_a=8.3$ s,$\mu_b=58.3$ s,初始值 $C_i(0)=0$.STA 模块仅用于平均速率响应分支中.若将 STA 模块加入到同步响应分支中,则会消除元

音的共振峰结构. 上述两种机制对同步响应分支仅产生轻微影响.

AGC 用于动态范围压缩以适应各种幅度的输入，其计算公式为 $HC_i(n)=\dfrac{y(n)}{1+K\langle y(n)\rangle}$. 式中 K 通常取 0.002[14]，$\langle y(n)\rangle$ 为输入信号的包络. 在平均速率响应支路中，$STA_i(n)$ 为输入信号.

在产生平均速率响应的分支上，信号最后通过的是包络检测器（低通滤波器）. 包络检测器的作用是避免高频信号产生脉冲响应，并平滑半波整流的输出. 包络检测器的输出即为平均速率响应，表示为 $ED_i(n)$. 包络检测器的转移函数为：

$$H_1(z)=\left[\frac{1-\alpha_1}{1-\alpha_1 z^{-1}}\right]^2, \tag{5}$$

式中 $\alpha_1^2=\exp[-1/\tau_1]$，时间常数 $\tau_1=4$ ms.

在产生同步响应的分支上，信号首先通过的是低通滤波器. 该低通滤波器的作用是模拟由神经延迟和响应抖动而导致在高频段发生的同步抑制现象，其输出信号用 $LPF_i(n)$ 表示. 该低通滤波器的传递函数为：

$$H_2(z)=\left[\frac{1-\alpha_2}{1-\alpha_2 z^{-1}}\right]^4, \tag{6}$$

式中 $\alpha_2^4=\exp[-1/\tau_2]$，时间常数 $\tau_2=0.04$ ms. 在产生同步响应的分支中，AGC 的计算过程与产生平均速率响应分支的计算过程相同，$LPF_i(n)$ 作为 AGC 的输入信号.

本文采用 Seneff 设计的 GSD(generalize synchrony detector)[15] 计算类似于自相关关系的输出来检测时间响应中的周期性，生成每个滤波器输出的和以及差的期望幅值和差值以及延迟的软限制比，每个 GSD 的延迟必须与滤波器的中心频率对应. GSD 的计算公式为：

$$GSD_i=A_s\tan^{-1}\frac{1}{A_s}\left\{\frac{\langle|y(n)+y[n-n_i]|\rangle-\delta}{\langle|y(n)-\beta^{n_i}\cdot y[n-n_i]|\rangle}\right\}, \tag{7}$$

其中，$y(n)$ 为 AGC 的输出 $HC_i(n)$，$A_s=4$，$\delta=0.1$，$\beta=0.99$，$n_i=f_s/f_i$，f_i 为第 i 个滤波器的中心频率. δ 的作用是抑制对小幅度信号的响应，A_s 的作用是控制输入的线性范围.

M. Ahmed 等[16] 研究表明，GSD 包含明显的伪峰，这些伪峰是由基频 F0、噪声及其他因素的谐波引起的. 为了消除伪峰，M. Ahmed 等提出了平均局部同步检测器(ALSD，average local sync detector)，该检测器的转化过程如图 2 所示.

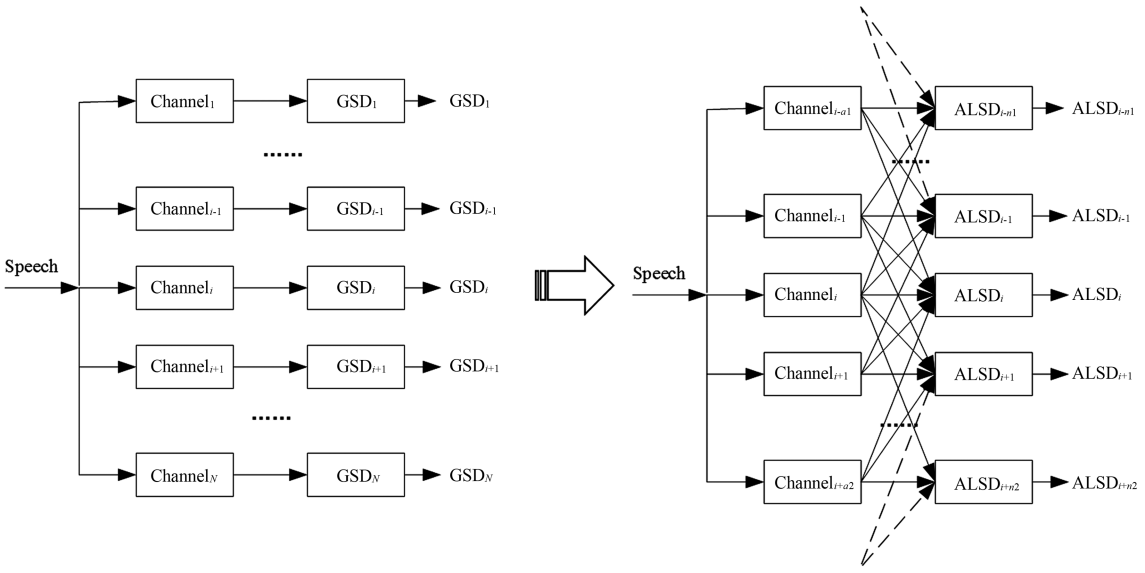

图 2　同步检测器转化为平均局部同步检测器的示意图

图 2 中,将各同步检测器的计算结果取平均值,即可得到第 i 个滤波器的 $ALSD_i$. $ALSD_i$ 的计算公式为:

$$ALSD_i=\frac{1}{n}\sum_{k=i-n_1}^{i+n_2}GSD_i(y_k)\ , \tag{8}$$

其中 $n=n_1+n_2$, $n=3$. 3 的含义为在中心滤波器的每一侧均有一个滤波器.

3 音节分割方法

3.1 基于 Seneff 听觉模型的音节自动分割算法

检测和分割朝鲜语语音音节,首先需要能够区分朝鲜语音节中的响音和阻塞音. 因响音和阻塞音在不同频段上存在差异,因此本文根据响音和阻塞音的频率分布特性,采用 Seneff 听觉模型中的临界频带滤波器对其进行划分. 表 2 给出了部分响音和阻塞音与 Seneff 听觉模型中滤波器的对应关系.

表 2 部分频带范围与 Seneff 听觉模型中滤波器的对应关系

频带划分	频带范围/Hz	滤波器序号
低频带	200～<800	1－12
中频带	800～<1 200	13－16
中高频带	1 200～5 000	17－35
全频带	200～5 000	1－35

发响音时,因声带振动的能量较高,且信号周期性和共振峰特性较为明显,因此检测时本文选择对信号周期性及共振峰特性具有明显效果的 ALSD 特征参数. 具体计算的参数包括低频带 ALSD、全频带 ALSD 和 ED 中低高频带比. 因响音的这 3 个参数值偏大,且 ALSD 谱的重心偏小,所以在确定边界点时,为了能够与其他参数保持趋势一致,采用负值描述 ALSD 谱重心. 各参数计算公式如下:

$$LBE(n)=\sum_{i=1}^{12}ALSD_i(n), \tag{9}$$

$$ABE_{\mathrm{ALSD}}(n)=\sum_{i=1}^{35}ALSD_i(n), \tag{10}$$

$$LHR(n)=\sum_{i=1}^{16}ED_i(n)/\sum_{i=17}^{35}ED_i(n), \tag{11}$$

$$SCG_{\mathrm{ALSD}}(n)=\sum_{i=1}^{35}iALSD_i(n)/\sum_{i=1}^{35}ALSD_i(n). \tag{12}$$

其中 $LBE(n)$ 表示低频带 ALSD; $ABE_{\mathrm{ALSD}}(n)$ 表示全频带 ALSD; $LHR(n)$ 表示 ED 中低高频带比; $SCG_{\mathrm{ALSD}}(n)$ 表示 ALSD 谱重心.

响音包括元音ㅏ/ㅓ/ㅗ/ㅜ/ㅡ/ㅣ、鼻音ㄴ/ㅁ和边音ㄹ,其中鼻音和边音在频率 200～400 Hz 范围内呈现明显的共振峰. 在 800 Hz 时,鼻音共振峰大幅衰减,而且鼻音收音和与它之前相邻的元音的能量发生也骤降:所以,对响音和阻塞音提取参数后,为避免出现相邻响音被划分到同一音节的现象,需要对响音进行进一步细分. 细分时提取的参数有全频带 ALSD、中高频带 ALSD、全频带 ED、中高频带 ED 以及 ED 谱重心,计算公式如下:

$$MHE_{\mathrm{ALSD}}(n)=\sum_{i=17}^{35}ALSD_i(n), \tag{13}$$

$$MHE_{\mathrm{ED}}(n)=\sum_{i=17}^{35}ED_i(n), \tag{14}$$

$$ABE_{\mathrm{ED}}(n)=\sum_{i=1}^{35}ED_i(n), \tag{15}$$

$$SCG_{\mathrm{ED}}(n)=\sum_{i=1}^{35}iED_i(n)/\sum_{i=1}^{35}ED_i(n). \tag{16}$$

其中 $MHE_{\mathrm{ALSD}}(n)$ 表示中高频带 ALSD；$MHE_{\mathrm{ED}}(n)$ 表示中高频带 ED；$ABE_{\mathrm{ED}}(n)$ 表示全频带 ED；$SCG_{\mathrm{ED}}(n)$ 表示 ED 谱重心.

利用式(9)—(16)计算得到 8 个参数后，需要进一步确认准确的突变点，以此确定响音和阻塞音的边界(切分点). 为了消除野点对突变点的影响，采用 Kaiser 滤波器(通带为 4 Hz，阻带为 14 Hz)进行时域平滑，采用高斯滤波器($\mu=0$，$\sigma^2=6$)进行频域平滑[17]. 平滑后通过定位突变点的位置来表征响音和阻塞音的边界位置. 突变点有正负两种类型. 为消除这两种类型的突变点，将同时满足式(17)中 3 个条件的突变点定义为正突变点，并标记为 n_+：

$$n_+=\{n \mid diff(n)>diff(n-1);\ diff(n)>diff(n+1);\ diff(n)>\theta_+\}. \tag{17}$$

其中：

$$diff(n)=x(n+1)-x(n); \tag{18}$$

$$\theta_+(x)=\mu(x)+p\sigma(x),\ p=0.5; \tag{19}$$

$$\mu(x)=\frac{1}{N}\sum_{n=1}^{N}diff(n); \tag{20}$$

$$\sigma(x)=\sqrt{\frac{1}{N}\sum_{n=1}^{N}[diff(n)-\mu(x)]^2}. \tag{21}$$

类似地，负突变点 n_- 定义为：

$$n_-=\{n \mid diff(n)<diff(n-1);\ diff(n)<diff(n+1);\ diff(n)<\theta_-\}. \tag{22}$$

其中，

$$\theta_-(x)=\mu(x)-p\sigma(x),\ p=0.5. \tag{23}$$

以朝鲜语语音“중앙인민방송국입니다”为例进行音节分割，各参数边界检测情况如图 3 所示. 图 3 中“▲”为波峰，“▼”为波谷，“+”为正突变点，“*”为负突变点.

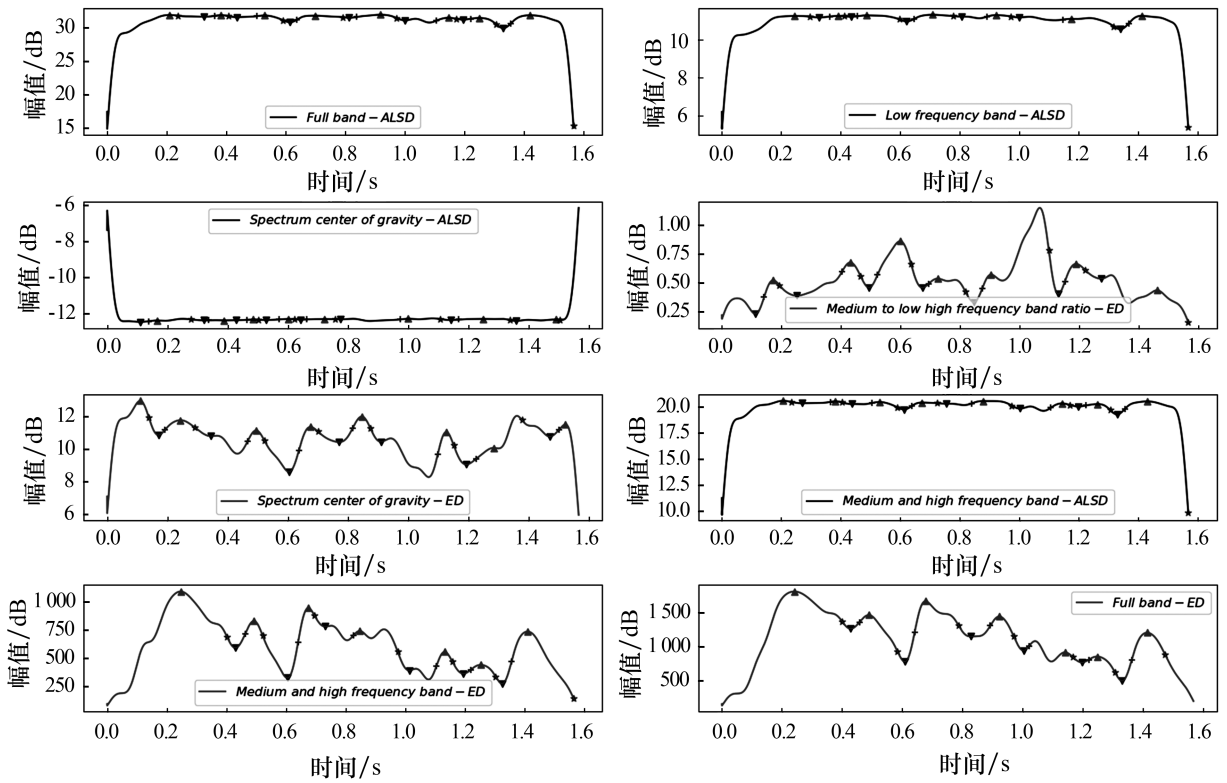

图 3　语音“중앙인민방송구김니다”的特征参数曲线

图 3 中的曲线为能量和共振峰等特征经 Seneff 听觉模型转变为 ED 和 ALSD 后在各频段的分布情况. 通过图 3 中的曲线趋势变化,可求得语音特征曲线的正负突变点,从而实现音节边界点的划分. 由于每个音节通常是由辅音加元音构成,所以若根据各频段中表示能量和共振峰的参数来确定音节的边界点,只需在求得的突变点中找到负突变点的位置即可确定音节的分割点. 结合朝鲜语发音特点,本文提出的基于 Seneff 听觉模型的朝鲜语语言音节自动分割算法(算法 1)的步骤如下:

step 1 由式(17)—(23)初步确定各特征参数的突变点位置.

step 2 查找每个特征参数曲线中两个连续波峰和波谷的时间间隔 t,若 $t>20$ ms,保留该区间的所有正负突变点;否则,保留区间内前面的波峰或波谷的突变点,同时删除区间内后面的突变点. 依据前一次迭代的情况保留波峰或波谷,如前一次迭代保留的是波峰,则本次保留波谷.

step 3 对所有保留下来的各参数的正负突变点以 5 ms 为单位进行分段. 每个波峰选 1 个最大的正突变点和 2 个最大的负突变点,若两个负突变点的时间间隔大于 15 ms,保留位置靠后的负突变点;否则保留斜率绝对值较大的负突变点.

step 4 将所有特征参数曲线的负突变点以 40 ms 为阈值进行聚类整合,获取音节分割的边界.

3.2 音节分割算法的改进设计

采用算法 1 对朝鲜语语音音节进行分割,结果如表 4 所示. 由表 4 可以看出,算法 1 的召回率较为理想,但是准确率偏低. 准确率偏低的主要原因是在检测音节的过程中,静音段中的噪声以及鼻韵尾、摩擦音、塞音等引起的音节检测错误较多. 检测过程中分割错误占比见表 3.

表 3 分割错误占比

错误原因	音素	占比/%
静音中的噪声	无	51.3
摩擦音/塞擦音,塞音	ㅅ/ㅎ/ㅈ/ㅊ,ㅂ/ㅃ/ㅍ/ㅋ	31.2
鼻韵尾	ㄴ/ㅁ/ㅇ	10.0
其他(元音/半元音等)	ㅏ/ㅣ/ㅓ/ㅡ/ㅜ/ㅗ/ㅑ	7.5

为了提高音节检测的准确率,本文提出改进的基于 Seneff 听觉模型的音节自动分割算法(算法 2),具体步骤如下:

step 1 通过双门限端点检测算法检测连续语音中的各个静音段边界,然后从算法 1 得到的候选音节边界列表中删除静音段引起的错误边界.

step 2 设定过零率阈值,并将大于此阈值的候选边界确定为摩擦音及塞擦音;设定阈值 a,若经 step1 筛选后的塞擦音和摩擦音的边界位置 k 满足 $k \leqslant$ 边界位置 $\leqslant k+a$,则删除该边界.

step 3 设定阈值 b,若经 step 1 和 step 2 筛选后剩余的候选边界位置 k 满足 $k \leqslant$ 边界位置 $\leqslant k+b$,则删除该边界.

在改进的音节自动切分算法中,静音段、摩擦音/塞擦音以及塞音等的检测阈值采用的是经验值,若该值采用不当,会误删真实的边界而导致召回率降低;因此,经验值的选取非常关键.

4 实验结果及分析

选取朝鲜语连续语音语料(准书面语)中的 100 段音频作为实验数据,其中包含响音(元音/边音/鼻音)、阻塞音(摩擦音/塞擦音)等各类语音. 语音的采样频率为 16 kHz,量化精度为 16 bit. 音节的真实边界通过人工标注获得,并将算法自动检测出的音节边界和人工标注出的基准边界进行比较. 假设算法得到的边界为 t_s,人工标注的基准边界为 t_p,且定义 $|t_s-t_p| \leqslant 20$ ms 时为检测准确.

算法评估指标采用准确率 P 和召回率 R. 假设算法检测出的边界个数为 N_t,人工标注的边界个数

为 N_h，边界检测错误的个数为 N_c，则准确率和召回率的计算公式为：

$$R=\frac{N_t-N_c}{N_h}\times 100\%,\ P=\frac{N_t-N_c}{N_t}\times 100\%.$$

算法 2 的实验结果见表 4. 由表 4 可看出，虽然算法 2 的召回率较算法 1 略有降低，但是准确率明显提高，说明算法 2 优于算法 1. 另外，若将算法 2 与人工校正相结合，则可在后续的语料标注过程中显著提高标注工作的效率.

表 4　两种算法的音节分割结果　%

边界检测方法	准确率	召回率	$F1$
算法 1	72.12	95.25	82.09
算法 2	86.43	93.56	89.85

5　结论

本文基于朝鲜语语音发音特点，提出了一种基于 Seneff 听觉模型的朝鲜语语音语料音节自动切分算法. 测试结果表明，本文方法不依赖于事先训练好的语音模型，仅仅从语音特征参数即可实现音节的自动切分，且切分效果显著优于传统的基于 Seneff 听觉模型的分割算法. 为提高音节分割的准确率，今后我们将引入机器学习的方法对其进行研究.

参考文献：

[1] 何可嘉. 广播语音的自动标注系统[D]. 北京：北京邮电大学，2010.

[2] 王丽娟，曹志刚. 基于 HMM 模型的语音单元边界的自动切分[J]. 数据采集与处理，2005，20(4)：381-383.

[3] 李诗心. 傣语语音合成系统中自动分词技术与音子自动切分技术研究[D]. 昆明：云南大学，2015.

[4] 韩虎. 汉语连续语音的音节自动标注算法研究及实现[D]. 哈尔滨：哈尔滨工业大学，2008.

[5] TOLEGEN Gulmira，邬春学. 基于深度学习方法的句子及语素边界划分研究[J]. 电子科技，2017，30(9)：20-22.

[6] PAILAI J，KONGKACHANDRA R，SUPNITHI T，et al. A comparative study on different techniques for Thai part-of-speech tagging[C]//Electrical Engineering/Electronics，Computer，Telecommunications and Information Technology (ECTI-CON)，2013 10th International Conference on. Krabi，Thailand：IEEE，2013：245-247.

[7] RUNSHEN C. A modified syllable segmentation method based on multi-feature for mandarin speech[C]//The 2010 4th International Conference on Intelligent Information Technology Application (IITA2010). Qinhuangdao，China：IEEE，2010：487-489.

[8] 郝静，张刚. 基于粒计算的清浊音检测算法[J]. 太原理工大学学报，2008，39(3)：39-40.

[9] 王艳，冯宏伟，张利平，等. 基于元音检测的汉语连续语音声韵母分割[J]. 计算机工程与应用，2011，47(14)：134-136.

[10] 姚徐，于洪志，单广荣. 音段自动切分系统的设计与实现[J]. 电脑知识与技术，2008，2(13)：737-739.

[11] 陈斌，张连海，王波，等. 基于 Seneff 听觉谱特征的汉语连续语音声韵母边界检测[J]. 声学学报，2012，37(1)：104-110.

[12] 王桂荣. 朝鲜语和蒙古语语音对比分析方法研究[D]. 延吉：延边大学，2018.

[13] 张美英. 基础韩国语[M]. 哈尔滨：黑龙江朝鲜民族出版社，2016：55-59.

[14] 陈斌. 汉语连续语音声韵母类别属性检测技术研究[D]. 郑州：解放军信息工程大学，2015.

[15] STEPHANIE S. Pitch and spectral analysis of speech based on an auditory synchrony model[D]. Cambridge：Massachusetts Institute of Technology，1980：83-89.

[16] AHMED M，ABDELLATY A. Robust auditory-based speech processing using the average localized synchrony detection[J]. University of Pennsylvania Scholarly Commons. Departmental Papers (ESE)，2002，10(5)：280-282.

[17] HU Guoning，WANG Deliang. Auditory segmentation based on onset and offset analysis[J]. IEEE Transactions on Audio，Speech，and Language Processing，2007，15(2)：398-399.