

# 汉语语音的聚类分段研究

宋原章 王仁华

(中国科技大学)

## 摘 要

本文提出了一种新的聚类分段算法,这个算法以段内平均离散度最小、段间平均离散度最大为准则,采用聚类的方法逐次迭代选择最佳分段断点和分段段数,能正确地对汉语语音进行音素分段,它和以往分段方法相比在性能上有很大提高。文中还给出了应用该算法对汉语单音所作的部分实验统计结果,可供进一步开展基于音素或音位的汉语语音识别研究参考。

**关键词:** 音素识别,离散度,聚类分段。

## 一、引 言

正确的语音分段是以音素序列的汉语语音识别为基础的。由于语音信号的特征变化呈现出有序性和短时平稳性,在音素内部尤其是元音音素内部的特征参数(例如短时谱包络或 LPC 系数)变化较慢,即帧与帧之间还存在很大的冗余度;而音素与音素的过渡区域,以及一些辅音音素内部特征变化很快。通过分段,突出变化快的语音部分,压缩变化慢的部分,可以最大限度地压缩语音特征在时间上变化的冗余度,同时又保留时间上有序性,从而提高语音识别的效果。

语音信号的分段主要有时域分段<sup>[1]</sup>、频域分段<sup>[2]</sup>,以及时域、频域相结合的分段<sup>[3]</sup>。在时域分段中利用时域参数,如功率时间分布,过零率;在频域分段中则利用能频分布,或者采用线性预测参数按一定距离准则进行分段等。但是以往的分段方法都有一个基本的弱点,即它们只考虑在给定的参数产生最大变化处作为分段断点,或分段时断点的判定只考虑断点相邻数帧之间的变化情况,这显然是不够的。为了使分割所得到的段达到段内平稳或近近平稳,理应考虑整个段的情况,应该从段与段之间的比较出发决定断点的位置,这样的分段显然更具有物理意义。基于这一思想,本文提出了一种新的聚类分段算法,对汉语单音节所作的实验表明,其分段性纯有较大的提高,并由此得到了汉语单音节中音素间可分性、过渡段存在性和稳定性的一些初步规律,为进一步开展基于音素或音位的汉语语音识别研究提供了基础。

## 二、聚类分段的原理

在一定距离定义下,决定一个音节所分割的段数和各段之间的断点,使得不依赖于分段段数和音长的某一平均误差度最小,定义这一误差度为平均段内离散度和平均段间离散度之比。平均段内离散度表示每段内各帧参数矢量之间的平均距离对段数的平均,平均段间离散度则表示各段的平均矢量之间的平均距离。采用聚类分段在语音识别中有以下二个明显的优点:

1) 识别时对一个音无任何先验知识,不知道其音素、音位结构,人为地规定如何分段(如分几段,分出辅音等)是不合适的。聚类分段则从聚类的角度上,按一定的准则自动决定分段数和断点,显然要合理得多;

2) 每个人发音特点不同,同一个人两次发同一音也有一定差别,尤其反映在过渡信息上,有时过渡明显,有时不明显,过渡段的变化在一定范围内有随机性,聚类分段能较好地反映这种随机性。

在已知一个音节的参数矢量序列后,聚类分段算法的基本步骤为<sup>[4]</sup>:

- 1) 计算均值矩阵和离散度矩阵,运用递推关系可使这一计算得到简化;
- 2) 建立分段总离散度矩阵,记录对应最小离散度的分段断点位置,在给定分段段数下决定最佳分割断点;
- 3) 根据平均段间离散度和平均段内离散度之比最大的准则,选择最佳的分段段数。

## 三、初步实验结果和统计分析

本文工作中使用的语音数据采集系统如图 1 所示。其中 ASA 16 信号处理器是由频率分布在 200 Hz—7000 Hz 的 16 个带通滤波器组成,故经 A/D 采样后得到的是 16 维频谱包络矢量序列。



图 1 语音采集系统框图

首先对汉语每个复合元音的多个样本进行聚类分段实验,并在以下几方面统计和估价:

- 1) 音素与音素之间的可分性;
- 2) 音素间是否存在明显的过渡段;
- 3) 若明显存在过渡段,则稳定性如何;
- 4) 过渡段的分割特性,包括能否单独分出或归并于前后的音素,或被分为若干段等;
- 5) 各个可分的音素和属于同一音位的其它音素之间差异程度;
- 6) 根据以上统计,确定每个复合元音的最小可分基元和基元间的过渡情况。

最终结果示于表 1。表 1 给出了复合元音中可用于音素、音位(序列)识别的最小基元的一种参考。

从表 1 可以得到以下一些初步推论：

- 1) 大部分双元音中的单元音可以作为识别的最小基元,而且它们之间的过渡特性较稳定;
- 2) 三元音中元素被前后音素同化的现象普遍,这与语音学中所说的语流音变现象吻合;
- 3) 有些元音(如 i) 存在可分的韵尾,韵尾特征相当稳定;
- 4) 某些音素组合如 ei、en 等,彼此融合不可分,应作为一个基元。但这类基元内部并不很平稳,识别时仍不能完全扔掉其内部的有序信息,对于这类基元,应作特殊的处理。

表 1 复合元音最小可分基元表

基元 结构		发音形状		开口呼		齐齿呼		合口呼		撮口呼	
复合韵母组合	ai	$a^1-T-i^1,$ $a^1-i^2$		ia	$i^1-T-a^1$		ua	$u^1-T-a^1$		üe	$(üe)^1$
	ei	$(ei)^1$		ie	$(ie)^1$		uai	$u^1-T-a^1-i^2$ -尾音			
	ao	$a^1-T-o^2,$ $a^1-o^1$		iao	$i^1-T-a^1-o^1,$ $i^1-T-a^1-T-o^1$		uo	$u^1-o^1,$ $u^1-T-o^1$			
	ou	$o^2-u^1,$ $o^2-T-u^1$		iou	$i^1-o^3-u^1,$ $i^1-T-u^1$		uei	$u^1-T-i^1$ -尾音, $u^1-(ei)^1$ -尾音			
鼻音尾韵母组合	an	$a^1-T-n^1$		ian	$i^1-T-(an)^2$		uan	$u^1-T-a^2-n^1,$ $u^1-T-a^2-T-n^1$		üan	$ü^1-(an)^2,$ $ü^1-T-(an)^2$
	en	$(en)^1$		in	$i^1-n^2$		uen	$u^1-T-(en)^1$		ün	$ü^1-n^1,$ $ü^1-T-n^1$
	eng	$e^1-ng^1,$ $e^1-T-ng^1$		ing	$i^1-T-ng^1$		ueng	$u^1-e^2-ng^1$		iong	韵头- $i^1-o^1-T-ng^1$
	ang	$a^1-T-ng^1,$ $a^1-ng^1$		iang	$i^1-T-(ang)^2$		uang	$u^1-T-a^1-ng^1$			
							ong	$o^3-T-ng^1,$ $o^3-ng^1$			

注: T 表示过渡阶段, 字母上角表示不同的基元。

对辅音与单元音以及辅音与复合元音构成的汉语单音节, 也进行了类似的分段实验和统计分析, 同样得到了辅元组合中最小可分基元及过渡段情况的一些有价值的结果。

#### 四、总结与讨论

本文提出的聚类分段算法, 在汉语孤立字音素分段中获得了较满意的结果, 比以往的

方法性能上有较大的提高。以此为基础进行的实验统计,得到了用于汉语语音识别的一些最小识别基元,可供基于音素、音位(序列)的汉语语音识别参考。

进一步的工作除了对过渡段随机性需要作大量实验总结规律之外,还包括发展一些其它手段来提高聚类分段的效果。例如,现在的分段算法中,离散度定义仅基于谱包络矢量之间的欧氏距离,也可以采用其它加权距离定义,或者将一些特征参数如谱矩、共振峰位置等加到离散度的定义中,使得离散度包含有多项指标,从而进一步提高聚类分段的可靠性。此外,对基于模糊等价关系的离散度性能也作过一些探索,结果表明,对某些音的分割和过渡段稳定性均有明显的效果,这些问题都正在深入研究之中。

### 参 考 文 献

- [ 1 ] Reddy, D. R., Segmentation of Speech Sounds, *J. of the Acous. Soc. of America*, 40(1966), pp307—312.
- [ 2 ] Shozo Makino and Keniti Kido, A Speaker-Independent Word Recognition Based on Phoneme Recognition for A Large Size (212) Words Vocabulary, ICASSP 17. 8. 1—17. 8. 4, 1984.
- [ 3 ] Rabiner, L. R. and Schafer, R. W., *Digital Processing of Speech Signals*, Prentice-Hall Inc., 1978.
- [ 4 ] 丁士晟,多元分析方法及其应用. 吉林人民出版社,1981.

## CLUSTERING SEGMENTATION OF CHINESE SPEECH

ZONG YUANZHANG    WANG RENHUA

(University of Science and Technology of China)

### ABSTRACT

This paper proposes a novel clustering segmentation algorithm. Based on the criterion of the minimization of average dispersion within segments and the maximization of average dispersion between segments, the optimum speech segmentation has been approached by this clustering method. Compared with former classical methods, this algorithm has improved the performance in the phoneme segmentation of Chinese syllables significantly.

The statistical results of some preliminary experiments for isolated Chinese words are reported, which can be used as reference in the phonemes recognition of Chinese speech.

**Key words** ——Phonem recognition; dispersion; clustering segmentation.