

---

# 一种新的针对汉语 TTS 的韵律词自动切分方法

李小明<sup>1</sup> 易立夫<sup>2</sup> 杨静<sup>2</sup> 孙金城<sup>2</sup>

<sup>1</sup>(清华大学热能系, 北京 100084);

<sup>2</sup>(中国科学院声学所, 北京 100080)

E-mail: backtou@sina.com

**摘要:** 以往的语音合成中, 分词单位多为词典词, 而这种分词单位所形成的韵律信息与 TTS 所需的韵律信息常不符合。本文针对 TTS 系统的需要制定了韵律词的标准。分别利用基于规则的方法和基于机器学习的方法对实际文本进行了韵律词级别的自动切分, 讨论了韵律词自动切分的前景和难点。

**关键词:** TTS; 韵律词

## 引言

在汉语语音合成系统中, 韵律信息是相当重要的, 以它为基础建立的韵律模型直接影响了合成语音的自然度和汉语语音合成的应用。如何得到符合 TTS 系统所需的韵律信息, 已经成为众多 TTS 研究人员越来越关注的问题。

以往的 TTS 系统采用的韵律信息都以词典词为单位, 从合成效果来看, 自然度不太理想。研究发现, 这主要是词典词单位太小的缘故。汉语中存在着很多这样的结构, 该结构由若干词典词组成, 读起来连接相当紧凑, 听感与词典词的听感并无区别。以词典词为单位的韵律信息不能体现出这种结构, 也就影响了汉语语音合成的自然度。

本文制定了适用于汉语 TTS 需要的韵律词标准, 使之成为一个介于词典词与韵律短语之间的单位。然后分别利用基于规则的方法和基于机器学习的方法对实际文本进行了韵律词的自动切分, 最后讨论了韵律词自动切分的前景和难点。

## 1 韵律词标准

本文所说的韵律词标准主要是针对人的说话习惯提出的, 人说话时不可能是一口气说出的, 尤其是长句, 其中肯定有不少停顿。我们认为相邻两个停顿之间的所有音节就构成一个韵律词, 这里不考虑停顿的等级。

韵律词有以下一些特点:

1. 韵律词由一个或多个词典词组成。
2. 韵律词读起来比较连贯, 没有停顿。
3. 韵律词间的停顿不是通过无声停顿或者时长延长来体现, 而是通过基频重设来体现<sup>[1]</sup>。
4. 韵律短语由韵律词组成, 而句子由韵律短语组成。

---

我们从语料库中选择了具有代表性的 2200 句, 然后用手工进行了韵律词的切分, 韵律词切分时的主要依据是看是否发生了基频重设。

这 2200 句话共有音节约 3 万个, 每句平均有 15 个音节, 切分后共有韵律词 8434 个, 韵律词平均长度为 3.56 个音节。

## 2 自动切分的一些问题

### 2.1 评价标准

首先说明如何来评价自动切分的结果。韵律词自动切分时要确定的是当前词典词后面是否是韵律词边界, 如果是, 自动切分时就会在当前词后面插入一个标记, 针对这些标记设立如下的参数:

1.  $n\_true$ : 自动切分后的标记中所有正确标记的数目。
2.  $n\_redundance$ : 自动切分后的标记中所有不正确标记的数目。
3.  $n\_total$ : 手工切分后所有标记的数目。

针对以上参数, 使用两个指标来评价切分结果, 这两个指标是召回率  $\alpha$  和正确率  $\beta$ 。其计算公式如下:

$$\alpha = \frac{n\_true}{n\_total} \quad (1)$$

$$\beta = \frac{n\_true}{n\_true + n\_redundance} \quad (2)$$

召回率和正确率综合起来才能反映韵律词切分结果的好坏, 如果其中一个指标偏低, 那么韵律词切分结果就是不可接受的。

### 2.2 预切分

韵律词是由词典词组成的, 所以首先对语句进行词典词级别的预切分, 然后才能进行进一步的韵律词切分。预切分的步骤如下:

1. 采用词典进行词典词切分, 同时排除可能存在的交叉歧义。这里词典是 5 万词的规模, 最长词长为 4。
2. 标注相应的词性, 利用 HMM 对多词性词进行词性排歧。这里采用的是 26 种词性的小词性库。
3. 处理人名和地名。

预切分后, 共有切分标记 15251 个。

## 3 基于规则的自动切分

### 3.1 规则的提出

经过对人工切分结果的分析发现, 语句中的韵律词很少是单音节的, 而且很多韵律词由同词性的词组成。针对这些特点提出了 3 条规则:

1. 同词性词合并规则

如果语句中相邻的词典词词性相同, 而且符合词性要求, 那么就合并为一个韵律词。这里的词性要

求根据对手工切分结果分析得出。

### 2. 单字实词规则

主要处理预切分后句子中残留的单字实词。根据不同词性，分别作为前一个词的后缀或者后一个词的前缀。

### 3. 单字虚词规则

主要处理预切分后句子中残留的单字虚词<sup>[2]</sup>。

## 3.2 自动切分结果

对 3 条规则采用不同的组合形式，其结果如表 1 所示。

表 1 基于规则的自动切分结果

同词性词合并	单字实词	单字虚词	召回率(%)	正确率(%)
N	N	N	100	41
Y	N	N	98	49
N	Y	N	100	46
N	N	Y	100	49
Y	Y	N	97	54
Y	N	Y	97	60
N	Y	Y	100	55
Y	Y	Y	96	65

从表 1 中可以看到采用 3 条规则后正确率提高了 24 个百分点，这说明规则还是起到了一定的作用。3 条规则中同词性合并规则和单字虚词规则所起的作用大致相同，单字实词规则所起的作用稍低。

采用规则的召回率都接近 100%，主要问题是正确率比较低，这可能是如下原因造成的：

#### 1. 词库规模

现在很多 TTS 系统都采用上 10 万的词库规模。这里词库规模相对较小，如果词典词没有切分出来，那会直接影响后续步骤的处理。

#### 2. 词性

对于规则处理来说，词性是越细化越好。很多 TTS 系统采用的词性标记数都在 100 以上，而这里由于词库的原因，只使用了 25 个词性标记。另外，词性标注也是当前 TTS 的一大难题，词性标注的不准确也会影响到规则的处理。

#### 3. 人工切分结果

由于用于参照的标注结果是根据基频重设手工切分的结果，有些语句的基频重设现象比较明显，有些则比较模糊，所以也就影响了手工切分的准确性。

## 4 基于机器学习的自动切分

### 4.1 方法的选用

当今 TTS 中机器学习的应用相当普遍。机器学习的方法包括决策树、神经网络、线性回归、HMM 建模等。

对于韵律词的切分来说，主要就是利用一些上下文的环境信息，得出韵律词的边界，而且这里所提

到的环境信息既有词性这样的离散量，又有词长这样的连续量，所以选用决策树方法进行韵律词的切分。

决策树方法，又称多级分类器，在韵律词切分时的步骤大致如下：

1. 选用一定量的手工切分后的语句作为训练集。
2. 针对选用的语句，建立决策属性集。
3. 确立决策结果，这里决策结果只有两种，即韵律词边界和非韵律词边界。
4. 利用决策属性集训练，得到决策树。
5. 利用决策树对测试集进行自动切分。

## 4.2 决策树训练

这里可以利用的决策属性有两种，一种是词长及位置属性，包括当前词的长度，前一个词的长度，后一个词的长度，整个句子中词典词数目，当前词在句中位置等；另一种是词性属性，比如当前词词性，前一个词词性，后一个词词性等。

实际训练时选用 1000 句进行训练，剩余的 1200 句进行测试。同时也对这两种类型的属性进行了组合，得到了相应的决策树，决策结果如表 2 所示。

表 2 基于机器学习的自动切分结果

词长及位置属性	词性属性	召回率(%)	正确率(%)
Y	N	18	42
N	Y	82	87
Y	Y	75	92

从表 2 可以看出，词性属性在决策树中起到至关重要的作用。词长及位置属性只是起到一个调节召回率和正确率的作用，如果用召回率与正确率的乘积作为参照，可以看到词长及位置属性对决策树的结果并没有什么提高。

## 4.3 与基于规则方法的比较

由于汉语的语言现象相当复杂，规则只能解决相对多数问题，要提高正确率，必须得细化规则，规则的数量也得相应增加。要提高规则的效果，就意味着工作量的大量增加，而且规则受人为的影响很大，所以规则方法是一种比较繁琐的方法，其提高空间也不大。

相对基于规则的自动切分方法，基于机器学习方法的结果在正确率上有所提高，但同时召回率有所下降。总体来看，两者的效果是不相上下的。但是机器学习方法的关键在于决策属性的选择，可供选择的属性是一定的，其工作重点就是选出几种合适的属性进行组合。相对来说，工作量比较小。

## 5 结论

本文针对 TTS 的特点，制定了适用于 TTS 系统的韵律词标准。并且分别利用基于规则的方法和基于机器学习的方法进行了韵律词级别的自动切分，发现两种方法有着各自的特点，而总体效果很相近。

今后将在此基础上进行两方面的研究，一是把两种方法结合起来，使得自动切分结果达到可以实用的地步；一是在韵律词的基础上，进行韵律短语的切分。

参考文献：

- 
- [1] 熊子瑜, 基频重设与语流间断, 第五届全国现代语音学学术会议, 北京, 2001  
[2] 李小明等, 自动分词中的虚词处理, 第五届全国现代语音学学术会议, 北京, 2001  
[3] Quinlan, J. R. C4. 5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann Publishers, Inc., 1993

作者简介: 李小明(1979—), 男, 江苏泰州人, 硕士生, 主要研究领域为语音合成, 汉语韵律模型; 易立夫(1977—), 男, 湖南长沙人, 硕士生, 主要研究领域为信号处理; 杨静(1975—), 女, 北京人, 主要研究领域为语料库建立; 孙金城(1946—), 男, 北京人, 研究员, 主要研究领域为信号处理、语音合成。

## A New Automatic Segmentation Method of Prosodic Word for Chinese TTS

Li Xiaoming<sup>2</sup> Sun Jincheng<sup>1</sup>

<sup>1</sup> (Chinese Academy of Sciences, Institute of Acoustics, Beijing 100080, China)

<sup>2</sup> (Department of Thermal Engineering, Tsinghua University, Beijing 100084, China)

E-mail: backtou@sina.com

**Abstract:** In former TTS system, the most popular segmentation unit is word from lexicon. But prosodic information based on this segmentation unit is not fit for TTS system. The article redefines the prosodic word for TTS system. The method based on regulars and the method based on machine learning are used to proceed automatic segmentation. The foreground and the difficulty of automatic segmentation of prosodic word are discussed.

**Key words:** TTS; prosodic word