基于 LPC 谱的汉语韵母感知度量的研究

黄高扬、贾珈、蔡莲红

摘要：语音信号的频谱分布在语言感知中具有关键作用。本文基于汉语韵母的频域特征，提出了一种韵母感知差异的度量方法。首先选取韵母的三个关键帧，然后计算三帧附近的平均 LPC 谱，进而计算特定频率段上的积分，以此作为每个韵母的特征向量。最后定义韵母之间距离的计算方式，并算出了韵母的感知距离矩阵。通过对比韵母的视位混淆树、单韵母共振峰分布图，看出基于 LPC 谱的汉语韵母距离分布与两者有着很好的一致性。通过对汉语言语听辨实测数据中的韵母进行混淆度计算，并将混淆度与基于 LPC 谱的距离矩阵相对比，发现两者也具有很好的一致性。表明了本文提出的基于 LPC 谱的汉语韵母感知度量方法的合理性和有效性。

关键词: 韵母，听觉感知， LPC 谱，距离度量，层次聚类

1. 引言

在汉语言语感知中，韵母的感知占有重要的地位。周迅溢、杨玉芳等提出：音高与时长是影响音节感知的主要特征[1]；张家騄基于汉语的特点，采用描述性的特征，初步提出汉语普通话声母、韵母、声调的区别特征系统[2]等。上述工作从定性描述韵母感知差异的角度进行研究，缺少参数化的定量分析，因此也不方便应用于听感差异的直接度量。对于如何参数化描述韵母感知差异，使之可以应用于言语工程的定量分析中，本文展开了相关的研究工作。

在语音学中，通常采用第一、二共振峰评估韵母之间的感知差异 [3][4]，发音过程中，单韵母的共振峰值中基本保持不变，而复合韵母的共振峰值则处于时刻的变化中。在听力学中，认为语言频率在人耳对语音信号的感知中具有关键作用。500Hz、1000Hz、2000Hz 这三个频率被称为语言频率，因为它们是人们言语交往的主要频率。聋儿配戴助听器后，如果在 500Hz、1000Hz、2000Hz 这 3 个频率上获得听力补偿，便可听懂约 70%的言语声。由于复合韵母共振峰值的不确定性和变化的复杂性，共振峰用于度量韵母的听感差异难度较大，而人耳在语言频率上的听感特性，却值得我们在量化韵母的听感差异时着重考虑。

线性预测编码(LPC)，是用过去时刻的语音采样值的线性组合，以最小预测误差来预测语音信号下一时刻的采样值，在语音编码和识别领域有着广泛的应用，也是分析语音的频域特征时的常用方法[5][6]。

本文基于汉语韵母声学特征分析，提出了一种韵母的感知度量方法。通过分段计算韵母在特定语言频率段上的 LPC 曲线积分，并将积分结果通过本文定义的度量方式转换为韵母间的距离，计算出了韵母间的感知距离矩阵。将距离矩阵与韵母视位参数分类和单韵母共振峰分布图进行比对，结果表明其具有较好的一致性，从而验证了该度量方法的合理性。最后，本文设计了基于 LPC 谱的韵母感知距离矩阵与汉语言语听辩实测数据中的韵母混淆矩阵的对比实验，实验结果表明两者也具较好的一致性，从而验证了该度量方法的有效性。

2. 频域特征提取

线性预测编码(LPC)相当于设计一个线性滤波器，LPC 系数便相当于这个线性滤波器的系数。在确定了 LPC 系数后，便可计算该滤波器在不同频率下的响应的估计

值，即 LPC 谱。本文在 LPC 谱的基础上进行不同韵母频域特征的提取。

2.1 分析语料库

本文的实验中所用的语料库是由男性发音人录制，共有汉语单音节字 550 个，涵盖了汉语所有的声韵母组合。所有语音文件均由人工进行声韵母边界以及基频序列的标注。

2.2 特征提取的步骤

对语料库中的每个音频文件，特征提取的过程都是一样的。具体步骤如下。

1、分帧与选帧

首先对语音文件进行分帧。从标注文件中读取韵母的起始及终止位置，选择时序上位于韵母总时长的 1/6、3/6、5/6 三个时刻的三帧，再选取与每帧相邻的前后各一帧，总共九帧的波形进行计算。

2、计算各段的平均频谱

对于每个时刻的三帧，进行加窗、高频预加重等处理，计算出 LPC 系数，进而通过程序拟合得到三条 LPC 谱曲线。然后对三条频谱曲线做平均，即对每个采样点对应的幅值进行算术平均。之所以进行这样的平均化，是为了减小选帧不适当所引起的误差。

最后，对于韵母的三个时刻，各有一条平均 LPC 谱曲线，反应各个时刻的频域特征。计算 LPC 系数、拟合出 LPC 谱曲线等步骤均在 MATLAB 中以调用库函数的方式完成。

3、计算 LPC 谱上语言频率段的积分值

为与纯音听力测试方法[7]对比，并不是直接计算 LPC 谱上三个语言频率点对应的值，而是以 500Hz，1000Hz，2000Hz 为中心，分别计算[450，550]、[950，1050]、[1950，2050]三个频率段内平均 LPC 谱曲线的积分。又因为语音的三个时刻分别对应一条平均 LPC 谱曲线，所以最后每个语音文件都对应一个九维的特征向量。

3. 距离度量

经过特征提取，语料库中的每个语音文件都对应于一个九维向量。然而由于同一个韵母所搭配的辅音不同，或者辅音相同但声调不同，一个韵母在语料库中会有多个样本。在多样本的情况下，必须事先定义一种距离度量的方式，使得不同韵母之间的距离能够反映出韵母之间的差异。

3.1 距离度量定义

我们首先定义不同的样本之间的距离为其特征向量间的欧式距离。对于不同的样本集之间的距离，我们采用两种不同的定义方式[8]：

方式一（Centroid）：定义两个样本集间的距离为它们的中心间的距离，样本集的中心即是指集中所有样本点的特征向量的平均值所对应的虚拟出的点。

$$D_1(X,Y)=\sqrt{\sum_{i=1}^{9}\left(\sum_{j=1}^{N_1}\frac{x_{ji}}{N_1}-\sum_{t=1}^{N_2}\frac{y_{ti}}{N_2}\right)^2} \qquad (1)$$

$$(x_{j1},x_{j2},...x_{j9})\in X$$

$$(y_{t1},y_{t2},...y_{t9})\in Y$$

其中 X，Y 为两个韵母的样本集，X 中样本数为 N_1，Y 中样本数为 N_2。

方式二(Ward)：假设两个样本集合并，两个集合中的所有样本点到新的集合中心的距离的平均值。

$$D_2(X,Y)=\sqrt{\sum_{j=1}^{N_1+N_2}\sum_{i=1}^{9}(x_{ji}-\bar{x}_i)^2} \qquad (2)$$

$$\bar{x}_i=\sum_{j=1}^{N_1+N_2}\frac{x_{ji}}{N_1+N_2}(i\in[1,9]) \qquad (3)$$

$$(x_{j1},x_{j2},....x_{j9})\in X\bigcup Y$$

其中 X，Y 为两个韵母的样本集，X 中样本数为 N_1，Y 中样本数为 N_2。

3.2 层次聚类实验

为比较这两种距离定义方式对韵母感知度量的影响，我们对所有样本进行层次聚类[8]。聚类过程为：

1) 初始将每一个样本点置为一类。

$$S_i=\{X_i\},X_i=(x_{i1},x_{i2}...x_{i9}),i<N$$

2) 每次都将类间距离最小的两类归为一个新类。

$$\text{If: } D(S_m,S_n)=\min\{D(S_i,S_j)\}$$

$$\forall i, j < N, i \neq j$$

Then: $S_m = S_m \bigcup S_n, S_n = \varnothing$

其中类间距离 D 的定义即分别采用以上两种方式。

3) 重复步骤 2，直到剩余类数达到要求。所得的聚类结果的正确率如下图 1 及表 1 所示。

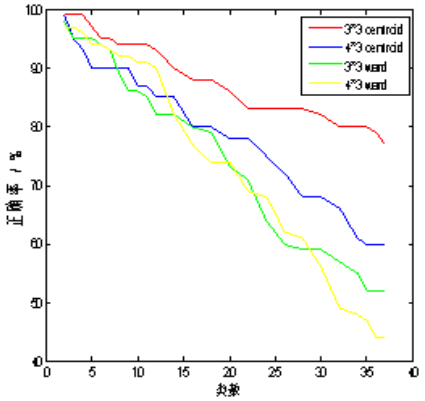

图 1：不同的特征向量以及距离方式在类数递增时，所引起的聚类正确率变化情况。

图 1 和表 1 中，聚类正确率是指每个韵母的全部样本中，如果有 60%以上被分在了同一个类中，则认为该韵母能够被正确分类，该韵母的分类正确率就是此百分比，否则认为该韵母不能被正确分类，正确率为 0。而图中所指正确率是聚类的整体正确率，即所有韵母聚类正确率的相加平均。

表 1: 设定剩余类数为 37 时的聚类正确率。

维度	类间距离度量方式	聚类正确率
9	方式一	78%
12	方式一	60%
9	方式二	52%
12	方式二	43%

实验时还以临界带宽为参考[6]，分别计算语音三个时刻内的 23 个临界带宽频带内的 LPC 谱曲线积分，得到一个 23 维向量。本实验只抽取了其中四个维度，即[400，510]、[920，1080]、[1720，2000]、[2000，2320]四个频段。最后得一个十二维的特征向量，作为上述九维向量的对比。

由实验结果可见，采用以语言频率段为基础的九维向量，且类间距离定义为类中心之间距离的方式，可以使得层次聚类取得较高的正确率。因此，我们将按照方式一(Centroid)中的定义，来度量两个韵母间的距离。

3.3 对比分析

我们采用方式一中的距离定义，计算所有韵母(37 个)两两之间的距离，得到一个距离矩阵，我们称之为感知距离矩阵。将此距离矩阵与视位参数和共振峰分布图做对比，以验证本文提出的距离度量方式的合理性。

3.3.1 与汉语韵母视位参数的对比

清华大学计算机系的王志明博士曾经从可视语音合成的角度，以汉语不同韵母在发音时口形的 FAP 参数做为特征向量，对汉语韵母进行了聚类，建立了视位混淆树[9]（见下图 2）。

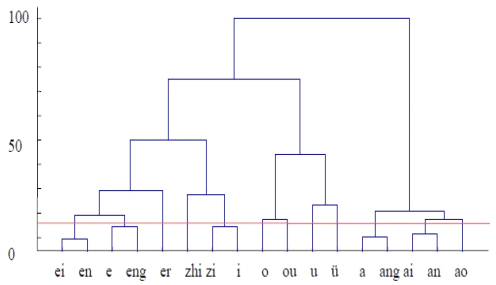

图 2：汉语韵母的视位混淆树。

我们将不同的韵母，基于感知距离矩阵，采用层次聚类方法进行聚类，建立韵母感知混淆树，并将其与视位混淆树进行对比分析。韵母感知混淆树的部分结果如图 3 所示。

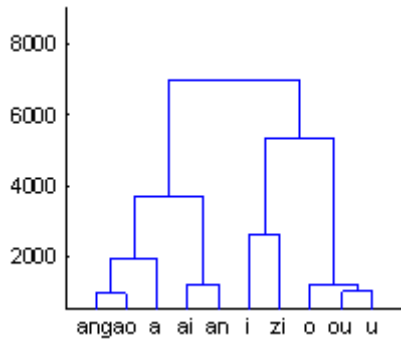

图 3：**汉语**部分**韵母的感知混淆树。**

通过对图 2 和图 3 进行对比可以看出：韵母感知混淆树与视位混淆树具有较好的相似性。视位参数在一定程度上体现了韵母发音过程差异，印证了本文提出的基于 LPC 谱的韵母感知度量的合理性。

3.3.2 与汉语单韵母共振峰分布图的对比实验

通过提取汉语单韵母的第一和第二共振峰，得到汉语单韵母共振峰分布图，如图 4 所示。

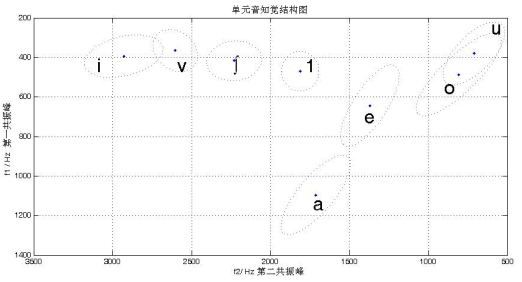

图 4：汉语单韵母共振峰分布图。

从图 4 中进一步总结两个单韵母之间的距离关系。 以单韵母“a”为例，它与其他单韵母的距离由小到大排列依次是：

e<o<i(zi)<u<i(zhi)<v<i

对比本文计算出的韵母感知距离矩阵，“a” 与其他单韵母的距离由小到大排列依次是：

e<o<i(zi)<u<i(zhi)<i<v

两结果具有较好的一致性。观察其他单韵母在共振峰分布图上的距离，与在韵母感知距离矩阵中的分布也有类似的一致性。该对比实验结果显示，基于本文提出的感知度量方法得到的汉语单韵母分布，与经典的韵母共振峰分布具有很好的一致性，从而验证了本文方法的科学性。

4. 实测验证

在以上的工作中，我们首先提取了各个韵母的声学特征，进而定义了不同韵母间的距离的度量方式，而且验证了该距离度量方式的合理性。至于这里的“距离”概念是否能反映出不同韵母听感上的差异，仍需实测数据的支持。

我们设计了一项实验，将感知距离矩阵与汉语言语听辨实测数据中的韵母混淆矩阵相对比，从而验证本文提出的感知度量方法的有效性。

4.1 实验方法

首先由汉语听辨实测数据计算韵母相互之间的混淆值[10]，两个韵母间的混淆值越大，说明人们在言语感知中将它们听混的概率越大，而对应到本文计算出的感知距离矩阵中，它们间的距离也应该越小。因此，在计算出所有韵母相互间的混淆值后，按混淆值由大到小的顺序，将所有韵母进行成对的排序，再观察排序之后的每对韵母间的感知距离会有怎样的规律。在进行这样的纵向比较的同时，也可以进行横向的比较，也就是比较该对韵母到其他韵母的平均感知距离与此对韵母间的感知距离有怎样的关系。

4.2 实验数据

本文中所使用的汉语听辨实测数据，采集自年龄在 20 岁左右的听力正常的人群，人数为 20 人。我们使用清华大学计算机科学与技术系与解放军总医院耳鼻喉头颈外科共同开发的《计算机辅助汉语普通话言语测听系统》[7]，以规范的流程对这些被测者进行普通话单音节言语测听，并在过程中记录他们的错误情况（即当被测者将某音听错成了另一个音时，将这两个音对应着记录下来），最终根据错误结果记录，计算两两韵母之间的听辨混淆概率，从而得到韵母的听辨混淆矩阵。为了保证测听的结果与之前计算结果有可比性，听辨实验中使用的语料即为本文分析用语料。

4.3 实验结果与分析

对于每对韵母，列出其在感知距离矩阵中的距离值，以及这两个韵母与其他韵母之间的平均距离与该距离的比值，作为对比。

这里给出部分的比较结果。表 2 所示为听辨混淆矩阵中混淆度最高的十对韵母的比较情况。其中，比例一是指：韵母一与其他韵母的平均距离与韵母对的感知距离的比

值；比例二是指：韵母二与其他韵母的平均距离值与韵母对的感知距离的比值。

表 2：听辨混淆矩阵中混淆度较高的十对韵母及它们之间的距离对比。

听辨易混淆的韵母对		混淆值	感知距离	比例一	比例二
韵母一	韵母二				
in	ing	0.152	680	5.04	4.82
uai	uan	0.10	1153	2.91	2.59
ei	uei	0.095	1665	2.15	2.02
en	uen	0.077	1210	2.13	2.29
uan	van	0.073	1261	2.37	2.09
o	uo	0.059	881	3.40	3.89
ia	iang	0.056	1647	2.09	1.72
en	eng	0.053	1725	1.49	1.49
ai	an	0.051	1210	2.67	2.34
a	an	0.051	2435	1.5	1.16

结果显示：在纵向的比较中，混淆值高的两个韵母，其感知距离不一定比混淆值低的两个韵母小，但是其总体的趋势是随着混淆值的降低而增加的。这是由于：在实际的语音中，韵母的感知会受到前面的辅音与声调等的影响，如果两个韵母可以搭配的辅音完全不同或只有少数相同，则它们之间混淆的概率就会大大降低，所以纵向的比较并未如预期中有较好的规律可循。而在横向的比较中，听辨混淆度较高的两个韵母之间的距离明显小于这两个韵母与其他韵母的平均距离，这从另一个角度证明了本文所计算出的感知距离能够反映出两个韵母之间的感知差异。

4.4 实验结论

由以上的实验可知：韵母之间的听感差异是能够以感知距离来度量的，听感上相差较小、比较容易混淆的两个音，它们之间的感知距离也相对较小。该结果说明了基于 LPC 谱的感知距离度量与听辨实测结果的一致性，从而验证了本文提出的韵母感知度量方法的有效性。

5. 结论与展望

本文提出了一种度量汉语韵母听感差异的方法。基于语言频率在人耳对语音信号的感知中具有关键作用，提取韵母的 LPC 谱在语言频率段的积分作为每个韵母的特征向量，并以特征向量之间的距离度量韵母之间的听觉感知差异。该方法与汉语韵母听辨实测结果，以及韵母视位参数和第一、第二共振峰的分布都有较好的一致性，实验结果说明了基于 LPC 谱的感知差异度量方式的有效性与可行性。该方法可以应用于言语工程的相关定量分析中，为进一步开展汉语言语感知计算、言语测听[7]的相关研究工作奠定了基础。

6. 致谢

本文研究得到了国家自然科学基金（编号：90920302，60928005，60910130）的经费支持。

7. 参考文献

[1] 周迅溢、王蓓、杨玉芳，关于汉语音节知觉空间的实验研究，声学学报，2003 年第 28 卷第 3 期，235—240 页。

[2] 张家騄，汉语普通话区别特征系统，声学学报，2005 年第 30 卷第 6 期，506—514 页。

[3] 林茂灿，语音知觉研究的几个问题，声学技术，1988 年第 7 卷第 2 期，23—27 页。

[4] 朱晓农，说韵母，语言科学，2008 年第 7 卷第 5 期，459—482 页。

[5] 蔡莲红、黄德智、蔡锐(2003)，现代语音技术基础与应用，北京：清华大学出版社。

[6] 杨行峻、迟惠生等(1995)，语音信号数字处理，电子工业出版社。

[7] 黄高扬、贾珈、蔡莲红、郁昕，计算机辅助汉语言语测听软件的研究与实现，HHME2009 收录。

[8] 张俊妮，数据挖掘与应用，北京大学出版社，2009 年。

[9] 王志明，汉语视位建模及可视语音的研究，清华大学工学博士学位论文，2003 年 4 月。

[10] 张家騄、齐士铃、吕士楠，汉语辅音知觉结构初探，心理学报，1981(1):76—85

黄高扬 清华大学计算机科学与技术系 100084
贾　珈 清华大学计算机科学与技术系 100084
蔡莲红 清华大学计算机科学与技术系 100084