

基于混叠效应下听感相似度的普通话声母聚类分析

章斯宇, 孟子厚

(中国传媒大学 传播声学研究所, 北京 100024)

摘要: 通过汉语普通话单音节清晰度听觉混淆实验考察混响所带来的混叠效应对普通话声母听感相似度的影响,并进一步比较在混响和噪声这两种不同的声学传递条件下声母听觉混淆情况间的差异,结果发现二者之间存在明显的差别。混叠效应下声母听感相似度的绝对量值明显高于噪声掩蔽下的值。且在混叠效应下声母的聚类关系大体上是按照发音部位形成的,而噪声掩蔽条件下的聚类关系大体上是按照发音方式形成的。这对思考构建合理的声母区别特征体系有启发意义。

关键词: 混叠效应; 声母; 相似度; 聚类分析

中图分类号: TN 912.34 文献标识码: A

文章编号: 1000-0054(2009)S1-1262-04

Reverberation effect on the clustering of Mandarin initials

ZHANG Siyu, MENG Zihou

(Communication Acoustics Laboratory, Communication University of China, Beijing 100024, China)

Abstract The effect of perceptual similarity of Mandarin initials caused by reverberation is compared with the effect of noise masking in various Mandarin monosyllable clarity simulations. The results show that the perceptual similarity of Mandarin initials having reverberation is significantly higher than with noise masking. The clustering of Mandarin initials due to reverberation is hierarchized according to the physical location of the pronunciation whereas the clustering is hierarchized according to the pronunciation method during noise masking. The results can help identify the distinctive features of the initials.

Key words reverberation effect; Mandarin initials; perceptual similarity; cluster analysis

对于汉语语音的音位系统,传统音韵学既提供了元音辅音分类系统,又提供了声韵调分类系统^[1]。为了能够用现代语音学理论解释这种传统音韵学分类系统,张家骥等采用听感实验的方法,利用语音听觉混淆的聚类分析方法确立了声韵调系统作为汉语语音的音位系统^[2-3],并据此初步提出了汉语普通话的区别特征系统^[3-4]。张家骥通过一个在多种噪声传递条件下的语言清晰度主观评价实验得到声母、韵母的听觉混淆度,又依据 Shepard 定义的语音相似度^[5]进行聚类分析,最终得到声母间和韵母间的易混淆情况^[2]。

语音的听觉混淆以及音位的知觉空间分布与声学传递条件有关。影响语言清晰度的因素主要有 3 个方面: 环境噪声、环境的混响、系统的失真。系统

无明显的失真是一个基本要求。虽然环境噪声往往难以避免,但可以根据环境噪声的模型和可能的量级对清晰度受噪声影响的程度进行预测。且大多数场所的背景噪声是动态的或可控的,在需要的时候可以将背景噪声降到不影响语言清晰度的程度。在这种情形下,对语言清晰度的影响就主要来自环境的混响作用。环境的混响作用一般是不能改变的,因此有必要研究无噪声干扰条件下,环境混响对语言清晰度的影响,对语音听感上混淆情况的影响。

本文参照张家骥的实验方法,以声母为例,通过

收稿日期: 2009-03-19

基金项目: 声望奖励金资助项目

作者简介: 章斯宇(1985-),女(汉),江西,博士研究生。

通讯联系人: 孟子厚,研究员, E-mail: mzh@cuc.edu.cn

一个多混响下的汉语单音节清晰度主观评价实验考察混叠效应下语音听感相似和易混淆情况。将聚类结果与张家骥等的实验结果进行比较分析,为汉语普通话区别特征系统的研究提供一定依据。

1 实验方法

实验参照国家标准《声学: 语言清晰度测试方法》^[6]进行。采用在实验室录音测听,模拟测试的方法。首先录制无混响的语音干信号,分别与模拟的不同混响时间的房间冲击响应函数作卷积来模拟受混响混叠作用后的语音信号,然后将这些语音信号在测听室回放给听音人进行语言清晰度的评测实验。

1.1 声学传递条件

语言清晰度测试分别在多种不同混响时间的声学传递条件下进行。由于本文考察的是混响造成的混叠效应下语音的听觉混淆情况,为避免其余声学因素,如直达声及部分早期反射声的干扰^[7],所用到的房间冲击响应函数均切除了前 10 ms 的直达声和早期反射声。作者利用混响效果处理器 TC-Reverber4000 模拟生成一个后期反射声分布均匀的混响时间为 6.0 s 的厅堂冲击响应函数,又对其进行模拟的均匀“吸声”处理,在不改变反射声序列和不改变与声源距离的前提下降低混响时间,最终得到来自同一房间同一测点上的具有 12 种不同混响时间的冲击响应函数。混响条件如表 1 所示。

表 1 实验中所采用的混响条件					
序号	RT/s	序号	RT/s	序号	RT/s
1	0.5	5	2.6	9	4.5
2	1.0	6	3.0	10	5.0
3	1.5	7	3.6	11	5.5
4	2.0	8	4.0	12	6.0

1.2 测试语料

采用了 8 张汉语语言清晰度测试音节表(KXY 表),各表之间等价性良好^[6]。每张表均有 75 个音节,采用随机组合的方式分成 25 组,即每组 3 个音节,组前需加上引导句。注意每组音节不构成一个单词。依据音节的朗读次序将每组内的第 1 个音节定义为前音,第 2 个音节为中音,第 3 个称为后音。

1.3 测试人员

语料的发音人为 1 男 1 女两名专业播音员。录制其按照标准普通话,以自然平稳的正常语速(4 个单音节/s)朗读的测试音节表。录音室混响时间小于 0.1 s,背景噪声低于 20 dBA。听音人共有 11 名,男女

比例基本平衡,年龄在 22~26 岁之间,均为母语为普通话的在校研究生,无话语和听力缺陷,有着多次普通话评测和测听实验的经历。

1.4 听音实验

录制的语音信号分别与前述 12 种房间冲击响应作卷积后开始测试。实验在混响时间低于 0.3 s,频响曲线基本平直的测听室内进行。语音信号通过 5 声道模拟环绕声系统重放,使听者有较好的包围感,声场中心声压级 65 dBA。正式实验前对听音人进行必要的训练以熟悉实验过程。按照标准规定,实验时以组(每组 3 个音节)为单位听音。每组音节间隔约 10 s,听完一组音节 after 听音人记录下自认为听到的音节拼音。如听到“序号×发你灰”。

由本文前期实验已知混叠效应对前、中、后 3 个音节的声母清晰度影响有明显差异^[8-9]。其中前音节声母受混叠效应影响最小,清晰度最高,而中、后音节声母受混叠效应的影响最明显,清晰度较低。本次实验的前、中、后 3 个音节处的声母清晰度分别是 0.587、0.418 和 0.413。本文的听觉混淆分析只选取了中、后音位置上音节的声母作为考察对象。

2 听觉混淆聚类分析

根据语言清晰度测试的结果分别统计出各名被试在各混响条件下的声母听觉混淆矩阵,整合到一起得到一个 12 种混响条件下的综合混淆矩阵。为了和张家骥的结果进行对比,只分析汉语普通话里不包括零声母在内的 21 个声母的听觉混淆情况。本次实验里,音节的刺激-响应事件共发生了 38 400 次,其中发生在中、后音位置上实际有效音节响应事件次数为 25 422 次,即声母的有效响应次数。

由混淆矩阵计算声母的混淆概率分布 $[p_{ij}]$,按照 Shepard 定义的语音相似度:

$$S_{ij} = (p_{ij} + p_{ji}) / (p_{ii} + p_{jj}). \tag{1}$$

式中: S_{ij} 表示第 i 个语音与第 j 个语音之间的相似度; p_{ij} 和 p_{ji} 分别是第 i 个语音听成第 j 个语音和第 j 个语音听成第 i 个语音的概率; p_{ii} 和 p_{jj} 则是第 i 个语音和第 j 个语音听对了的概率。获得声母的听觉相似度分布 $[S_{ij}]$ 。

采用无需先验知识的凝聚式分层聚类方法处理相似度结果。将 S_{ij} 视作声母 i 与声母 j 之间的度量距离, S_{ij} 越大则距离越短, S_{ij} 越小则距离越大。21 个声母最初各自自成一类,相似度最高的 2 个声母首先聚成一个小类,依据重心法计算这个小类和其余类之间的相似度,即距离,然后继续是相似度最高的

2类聚成一类,照此步骤往复,最终得到分层聚类的结果,见图1。图中右侧纵坐标表示听感相似度绝对尺度,聚类图中的每个节点在坐标轴上的对应值反映该节点两分支之间的听觉混淆水平。图2是张亚琛的噪声条件下的声母听觉相似度聚类图。

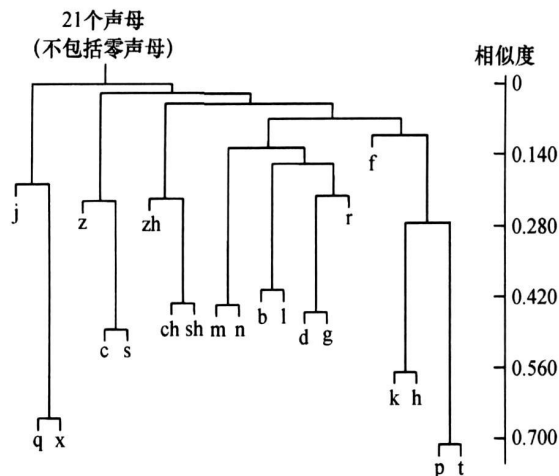


图1 混响条件下的声母听觉相似度聚类图

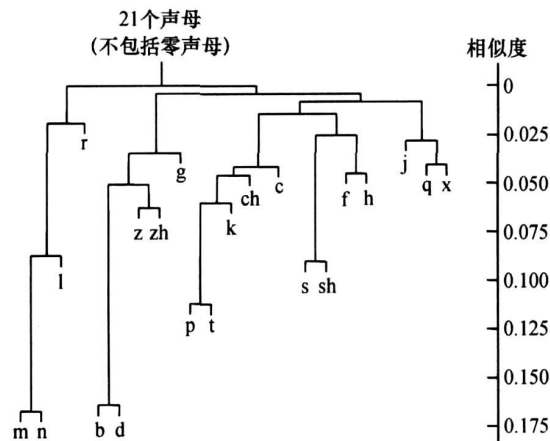


图2 噪声条件下声母听觉相似度聚类图

3 实验结果讨论

比较以上2幅相似度聚类图发现混响条件下的声母知觉空间分布和噪声条件下的存在明显的差异。从聚类图自上而下可以看到噪声条件下声母首先被区分开的是浊音—非浊音这个特征,接着在非浊音内不送气的声母被区分出去,除去舌面音剩下的声母又可分成送气的和摩擦音的2个子类。总体来看除了j, q, x这3个声母外,在噪声条件下基本按照发音方式形成聚类关系。而j, q, x的例外可能是由音位配列学规则决定的^[3]。

与噪声条件相比,混响条件下的声母混淆与语音学上按发音方法和发音部位对声母的分类也有着密切的联系。但首先被区分开的是j, q, x这3个舌

面音和其余的非舌面音,然后在非舌面音内舌尖前音z, c, s和非舌尖前音被区分开,接下来舌尖后音zh, ch, sh从非舌尖前音中分离出来,剩下的声母又可以看成不送气的和送气的2个子类。发现混响条件下存在按发音方式的聚类关系(m, n, d, g; p, t等),也存在按发音部位(舌位)的聚类关系(j, q, x; zh, ch, sh; z, c, s; k, h等),并且逐层分类时更多的是根据发音部位进行声母聚类的。

另外比较二者听感相似度的绝对量值也能发现混响条件下的听感相似度要明显高于噪声条件下的值,即便是同一对易混淆的声母,如p, t, 噪声条件下的相似度要远低于混响条件下的相似度。

如果用聚类图中在每个节点处二分时起主要区分作用的发音方法或发音部位进行标记,可以得到如图3和图4所示的聚类逻辑图。观察这2幅聚类逻辑图的顶部几层,噪声条件下相似度最低即最不易混淆的是浊音和非浊音,就是说最容易识别区分开的是清、浊这一对特征,而在非浊音内最容易被识别区分的是送气、不送气的这对特征。如果把识别1个语音看作是在聚类逻辑图上的路径搜索,那么噪声条件下更多的是搜索语音的发音方法。混响条件下最容易被区别出的是舌面音、非舌面音这对特征,其次是舌尖前音、非舌尖前音。基本上混响条件下的识别语音的路径首先是从发音部位开始的。由此看来噪声条件下听觉对语音的发音方法较为敏感,而混响条件下则对发音部位较为敏感。

4 结论

本文通过在纯混响条件下汉语普通话单音节清晰度的听感混淆实验,对声母的听感相似度聚类关系进行了分析。结果发现混叠效应所造成的听感混淆与噪声掩蔽所造成的听感混淆有明显区别。混叠效应下的听感相似度的绝对量值明显高于噪声掩蔽下的相似度值,且在混叠效应下,声母的聚类逻辑首先是按照发音部位形成的,而噪声掩蔽条件下的聚类逻辑首先是按照发音方式形成的。这对构建一个全面、完备的声母的区别特征体系有启发意义。

由本文结果可以联想到,造成噪声掩蔽和混叠效应这2种不同物理条件下声母聚类逻辑不同的原因有可能主要来自2个方面,一是在传递过程中不同的物理条件造成语音信号的物理畸变过程有所不同;二是不同的物理环境条件作为一种刺激对人的听觉感知系统(包括外耳和听感神经处理)发生了不同的作用。这些问题有待进一步的分析和探索。

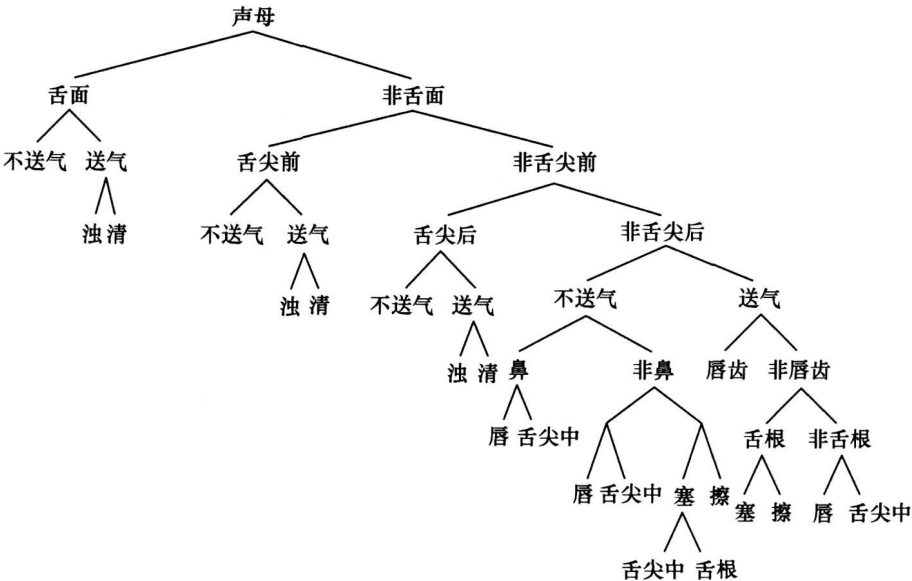


图 3 混响条件下的声母聚类逻辑

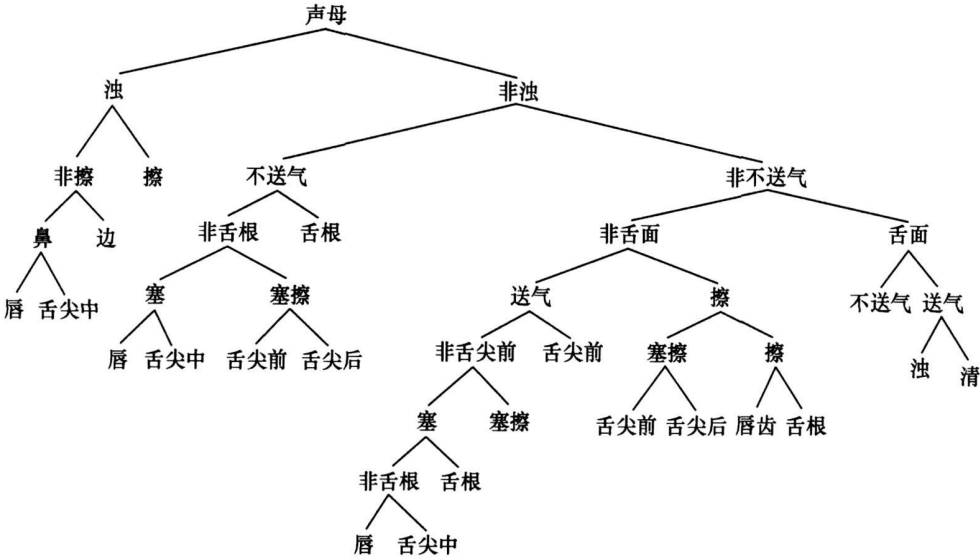


图 4 噪声条件下的声母聚类逻辑

参考文献 (References)

[1] 张家骥, 齐士铃, 吕士楠. 汉语辅音知觉结构初探 [J]. 心理学报, 1981, 13(1): 76 – 85.
ZHANG Jialu, QI Shiqian, LV Shinan. The research on perceptual structure of Chinese consonants [J]. *Acta Psychologica Sinica*, 1981, 13(1): 76– 85. (in Chinese)

[2] ZHANG Jialu, LÜ Shinan, QI Shiqian. A cluster analysis of the perceptual features of Chinese speech sounds [J]. *Journal of Chinese Linguistics*, 1982, 10: 189 – 206.

[3] 张家骥. 汉语普通话区别特征系统 [J]. 声学学报, 2005, 30(6): 506 – 514.
ZHANG Jialu. The distinctive features for standard Chinese (Putonghua) [J]. *Acta Acustica*, 2005, 30(6): 506– 514. (in Chinese)

[4] 张家骥. 汉语普通话区别特征系统树状图 [J]. 声学学报, 2006, 31(3): 193 – 198.

ZHANG Jialu. The distinctive feature trees of standard Chinese (Putonghua) [J]. *Acta Acustica*, 2006, 31(3): 193 – 198. (in Chinese)

[5] Shepard R N. Psychological representation of speech sounds [M]//David E E, Denes P B. Human Communication A Unified View. New York: McGraw Hill, 1972.

[6] GB/T 15508-1995,《声学 语言清晰度测试方法》[S].

[7] 王以真. 实用扩声技术 [M]. 北京: 国防工业出版社, 2004 82 – 84.

[8] 戴璐, 孟子厚. 混响条件影响汉语单音节清晰度与 STIPA关系的初步实验 [C]//中国声学学会 2008年学术交流会. 上海, 2008.

[9] 章斯宇, 戴璐, 孟子厚. 混叠效应对汉语单音节清晰度测试方法的影响 [C]//中国声学学会 2008年学术交流会. 上海, 2008.