

一种基于声音分类学习的神经模型研究

王莉莉, 张少白

(南京邮电大学 计算机学院, 江苏 南京 210003)

摘要:语音感知能力是人类感知系统的一个重要方面。在婴儿时期,随着婴儿对母语音位的不断接触和学习,会产生分类感知和知觉磁效应现象,这与自组织特征映射的经典表述相矛盾。文中提出了一种能对这种现象进行解释的神经模型,并通过模型的仿真实验证明了该模型的可行性和正确性。模型进一步预测磁效应的产生是因为原型元音比非原型元音有更小的听觉皮层表示,通过功能磁共振成像实验对模型提出的预测进行了论证。

关键词:声音分类学习;知觉磁效应;自组织特征映射

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2011)12-0127-04

Study of a Neural Model Based on Sound Category Learning

WANG Li-li, ZHANG Shao-bai

(Computer College, Nanjing University of Posts & Telecommunications, Nanjing 210003, China)

Abstract: Speech perception is an important part of the perception system. The category learning during infancy leads to the phenomena such as categorical perception and the perceptual magnet effect, which conflicts with the classical formulation of a self-organizing feature map. It proposes a model to explain this phenomenon and the simulation of model proves the feasibility and correctness of this model. The model further predicts that the magnet effect arises because prototypical vowels have smaller auditory cortical representations than non-prototypical vowels, which was supported by a functional magnetic resonance imaging (fMRI) experiment.

Key words: sound category learning; perceptual magnet effect; self-organizing feature map

0 引言

将各种刺激分为不同类别的能力是人类感知系统中一个非常重要的方面,比如在生活中人们会将一个不断变化的语音信号解释为一连串离散的音位、音节以及单词。当一个婴儿出生时,她就像一个“世界公民”一样,做好了在各种自然语言中分辨不同语音差异的准备。但是,随着年龄的增长及学习语言过程的展开,这种天生的状态也会被外界语言信息的输入所改变。婴儿学习语言的一个重要任务就是形成母语的音位范畴表征。而这种母语音位范畴化的能力对于婴幼儿的言语知觉、语言模仿和语言获得都是非常重要的。

世界上每一种语言的元音基本上都可以通过声音信号的前三个共振峰频率被区分开来。这些共振峰频率可以被看做一个三维的声音空间,不同的语言会把这个空间分成不同的元音系统。如英语在频率空间的

元音区域近似有12个纯元音类别,而西班牙语和日语在同样的空间只有5种元音类别^[1]。当一个婴儿进入特定语言者的世界时,他必须以某种方式学会将语言信号解析为母语的音位组。例如,她必须学会将共振空间分割成与她当地语言的元音数相对应的恰当数量的离散区域。

1 知觉磁效应现象

在声音空间,人们对空间中不同部分的声音的辨别能力是不同的。比如,在声音分类学习过程中,对受试者施加人工合成的语音刺激时,受试者对于一个类别内的两个刺激的辨别能力很弱,反之对于横跨类别边界的两个刺激的辨别能力较强,这种现象就称之为分类感知。而知觉磁效应指出,相对于原型元音,听者对非原型元音的辨别能力更强,尽管一般情况下原型元音比非原型类元音经历的要多的多^[2]。

有关感官映射在神经物理学方面的大量研究显示,过多的遭受某种特定类型的刺激,通常会导致该刺激更大的皮层表示。比如,饲养在只具有竖条纹的视觉环境中的小猫,相比在正常环境中饲养的小猫具有更多的被调整用于竖直形状的视觉皮层细胞^[3]。

收稿日期:2011-05-23;修回日期:2011-08-27

基金项目:国家自然科学基金(61073115)

作者简介:王莉莉(1989-),女,江西九江人,硕士研究生,研究方向为模式识别与智能系统;张少白,硕士研究生导师,研究方向为人工智能与认知科学,信息获取、处理与识别等。

在自组织特征映射的经典表述中,更多的遭受某种刺激会导致对该刺激更大的皮层表示。而且人们普遍相信,更大的感官皮层表示会使得对其表示的刺激有更好的鉴别能力^[4]。尽管对于听者,原型元音可能比非原型类元音声音经历的要多的多,但是知觉磁效应指出听者更不擅长于区分原型元音。这显然与自组织特征映射的经典表述相矛盾,因为在经典的自组织特征映射中,对原型元音将会有有一个更大的皮层表示,相比具有较小皮层表示的非原型的类元音声音,应该有更好的鉴别能力^[5]。文中所提出的神经模型能对这种知觉磁效应进行很好的解释。

2 分类学习的神经模型

为了能对知觉磁效应进行解释,文中提出了一种基于自组织特征映射体系结构的神经模型。该模型的基本思想是:在婴儿时期的分类学习会改变听觉皮层映射中神经元的激活优先级的分布,从而改变对声音空间不同部分的声音分辨能力。

模型的示意图如图 1 所示。模型包括两层:共振表示层和听觉映射层,通过自适应权值相互连接。假设听觉系统在早期处理语音信号输入,并提取出重要的听觉信息,如模型的第一层表示的共振峰频率。每一个共振层细胞投影到听觉映射层的每个细胞,听觉映射上的细胞被看作共振频率空间上的一个特定点,并对于特定输入产生最大程度的激活^[6]。

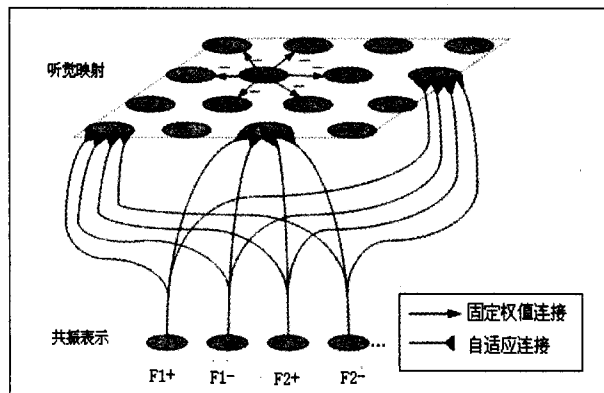


图 1 模型示意图

对一个声音信号而言,其共振峰的神经元表示如下:

$$x_i^+ = \frac{F_i - F_{i\min}}{\sqrt{(F_i - F_{i\min})^2 + (F_{i\max} - F_i)^2}} \quad (1)$$

$$x_i^- = \frac{F_{i\max} - F_i}{\sqrt{(F_i - F_{i\min})^2 + (F_{i\max} - F_i)^2}}$$

式中 i 为声音信号的共振峰序号, x_i^+ 和 x_i^- 是对第 i 个共振峰进行编码的细胞活力,成对的共振峰表示是为了确保输入声音的共振峰值的共振表示的总活力相

同。 F_i 为第 i 个共振峰频率值, $F_{i\max}$ 和 $F_{i\min}$ 分别为第 i 个共振峰频率的最大值和最小值,这两个参数保证细胞活力 x_i^+ 和 x_i^- 在 0 到 1 之间变化。

$$m_j = \frac{\sum_i x_i^+ z_{ij}^+ + x_i^- z_{ij}^-}{M} \quad (2)$$

m_j 为听觉映射层第 j 个映射细胞的活力, z_{ij}^+ 和 z_{ij}^- 为共振表示层第 i 对细胞投影到第 j 个映射细胞的可变突触权值, M 为映射层的细胞数。

在下面的模型仿真中,听觉映射的学习依赖于 Kohonen 自组织特征映射的运算法则的改变。在 Kohonen 的公式中,学习速率 η 是一个标量,它以指数级时间衰退,并对发生在特定时间步骤内的突触权值的变化总量进行调整^[1]。模型中, η 也通过当前输入与最近的类别中心的欧氏距离来调整。当输入与某个类别中心一致时,学习速率呈现最小值,当输入靠近类别边界时,学习速率呈现最大值^[7]。

模型在听觉映射层的输出是一个群体响应。假设在与当前输入的一定距离 r 内,所有带调制中心的细胞都会对表示感知语音的群体代码有所贡献。群体响应是活性细胞的调制中心的加权平均,而每个细胞的贡献是以该细胞的激活水平来衡量的^[8]。

根据这个模型可知,是婴儿经历的语音训练的类型,而不仅仅是训练刺激的分布导致了语音刺激下神经映射的缩小。分类训练导致对频率越高的刺激进行编码的细胞数量越少,而辨别训练的结果恰恰相反。

这个模型对知觉磁效应的解释非常简单:一个类别的原型样例比非原型样例更难以相互区分,因为它们听觉皮层中的表示更小。该模型进一步假设,减少的皮层表示是源于婴儿时期的音位分类学习,特别是将来自声音空间的特定区域的声音看做同一类别成员的学习过程会导致类别中心附近声音的听觉皮层表示范围的缩小。

3 仿真实验

实验 1. 模型的仿真。

针对对应于美式英语音节 /r/ 和 /l/ 以及日语音节 /r/ 的共振频率空间区域,分别对模型进行了仿真。图 2 的上半部分显示了美式英语发音 /r/ 和 /l/ 的仿真结果,仿真过程中,从一个近似于美式英语 /r/ 和 /l/ 的声音分布中抽取声音样例施加于该模型。这个仿真的共振峰表示包括第二和第三个共振峰频率 (F2 和 F3)。由仿真结果可以看出用于表示原型声音的细胞更少(在 F2-F3 平面用菱形表示),而类别边界附近的声音有更大的皮层表示。而图 2 的下半部分显示了在共振频率空间中,从近似日语 /r/ 的声音分布中选取声

音刺激,施加于模型后的仿真结果。

对比两个仿真结果图,不难发现在图 2 的上图有一个对应于日语/r/的类别中心的回升点,这个回升发生在近似于美式英语/r/和/l/的边界位置。这就解释了为什么日本学习英语的成人难以区分英语中的/r/和/l/,因为它们反复接触日语音节/r/,所以它们会有相对较少的与这个边界的共振空间区域相对应的听觉皮层细胞。

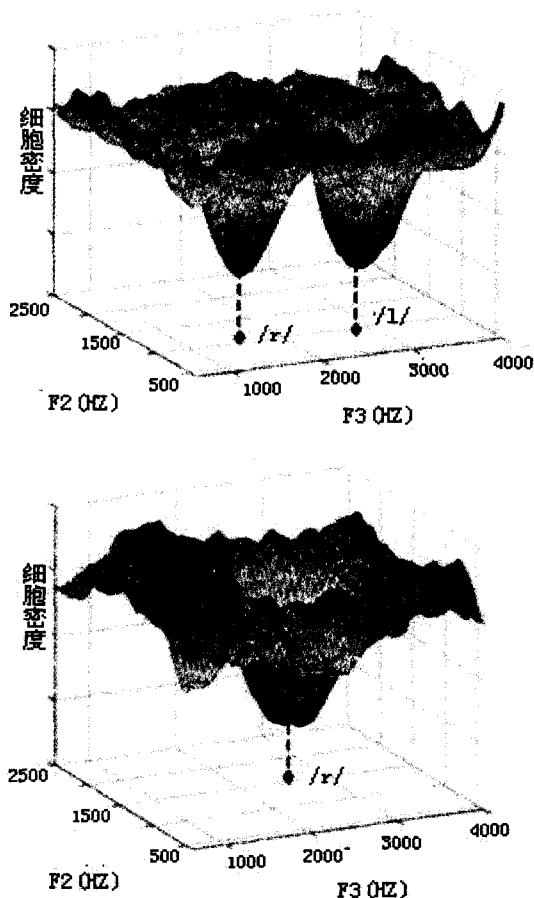


图 2 分别用美式英语发音/r/和/l/ (左图),以及日语发音/r/ (右图)对模型进行仿真训练
实验 2. 用听觉神经物理学实验验证模型。

Guenther 等人在 1999 进行了一个听觉神经物理学实验,在这个实验中,受试者要在 45 分钟内执行一个分类学习任务,每个受试者训练前后对声音的辨别能力用 d' 测量的方法进行评估^[9]。在频率空间的两个区域:控制区域和训练区域中的每个区域都有 7 个刺激产生,然后分别测量训练前后对每个范围的中心刺激和这个范围的其他刺激的辨别能力。控制区域的声音刺激是没有在训练期间出现的。实验结果如图 3 上图所示。分类训练导致对应于训练类别中心的频率空间区域的刺激的辨别能力下降,并出现元音的知觉磁效应现象。尽管在实验过程中所受的训练区域的刺激比控制区域的刺激更频繁,但结果显示,相比受试者

辨别控制区域内刺激的能力,他们对训练区域刺激的辨别能力有所减弱。

在另一个辨别训练实验中,对受试者施加分类训练实验中相同的刺激。在每一次试验中,受试者被给予来自训练区域的两个刺激,并被要求说出这两个刺激是否相同,试验结果如图 3 的下半部分所示。分类训练使得对训练刺激的辨别能力下降,而采用同样一组刺激的辨别训练则使得辨别能力增强。这表明,是训练任务的本质,而不只是训练刺激的分布导致了第一个实验中的知觉磁效应。

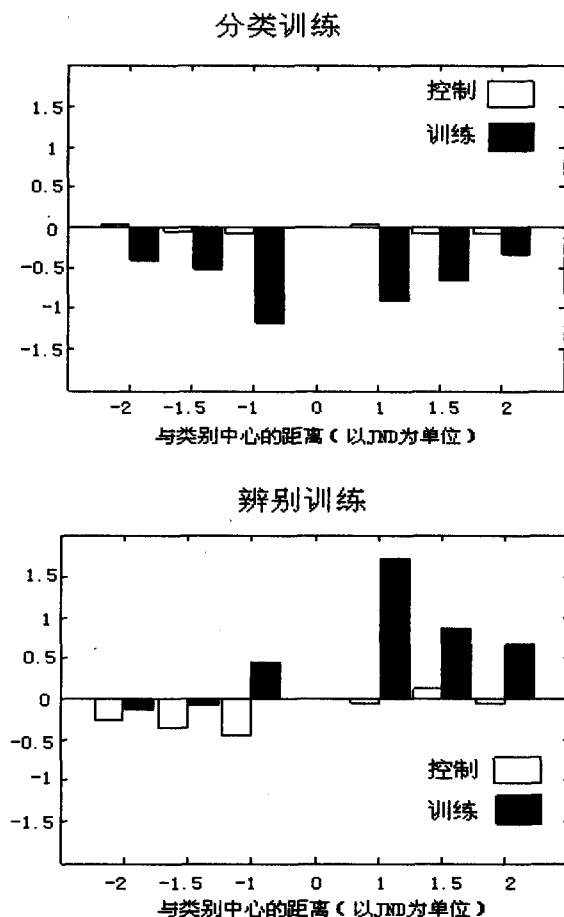


图 3 分别进行分类训练和辨别训练

对模型进行仿真并与 Guenther 等人的实验数据加以比较。在仿真中,第一层用于传导神经生理学实验中受试者遭受的窄带宽噪音刺激的中心频率。当该模型进行听觉映射的学习时,根据实验规定,用与原型间的不同距离 (以实验开始测量的 JND 的单元数来表示) 来表示不同刺激^[10]。

据 zohary 所说,在群体代码响应中噪音的标准差与群体响应中用到的细胞数的平方根是成反比的^[8]。在此处的仿真中,我们假设在与输入刺激的固定距离 r 内,所有与刺激一致的细胞构成了该刺激的群体代码响应。因为类别原型附近频率空间区域的刺激的细胞表示更少,所以原型附近刺激的神经元表示比远离

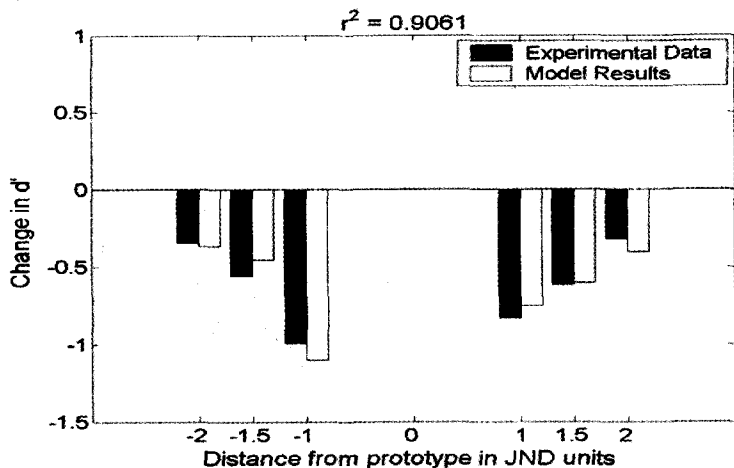


图 4 仿真结果与实验数据的对比

类别中心的刺激的神经元表示对噪音更加敏感^[11]。

与神经生理学实验一样,在训练后的测试阶段对这个模型施加成对的刺激,如 zohary 所描述的加入高斯噪声,然后计算全部刺激的噪音群体响应,如果他们都在某个敏感性度量的范围内,这些刺激就被视为相同,否则就会判为不同^[8]。结果的 d' 值被计算出来,并与用频率空间细胞的均匀分布映射所获得的值进行对比。 d' 值的变化与用线性最小二乘拟合(见图 4)所得的实验结果是一致的, r^2 值为 0.9061,说明这个模型能对实验结果中 90% 的数据进行解释。这不仅论证了该模型的可行性和正确性,也说明该模型是具有神经生理学意义以及实际应用价值的。

实验 3. 用大脑成像实验验证模型的预测。

前面提到,根据该模型的基本思想,可以进一步预测,知觉磁吸效应之所以会产生,是因为在听觉皮层中一个音节的原型样例的皮层表示更小。模型的这个预测可通过一个功能磁共振成像实验来检验^[12],实验中受试者分别听美式英语元音 /i/ 的原型和非原型样例。

实验结果显示,相比非原型样例,听取元音的原型样例会使得听觉皮层的激活区域更少。在听觉皮层区域的大部分区域,原型元音比非原型元音有更少的激活区域。尤其在右半球听觉皮层区中这一点特别明显^[13]。这个结果支持了该模型对知觉磁效应的解释:原型元音比非原型元音更难相互辨别,是因为它们有更小的皮层表示,而更小的皮层表示导致对神经信号中的噪音更加敏感。

4 结束语

文中提出了一种基于声音分类学习的神经模型,它对语音感知中出现的知觉磁效应现象给出了解释:在婴儿时期的音位类别学习会导致知觉磁效应的产生,因为分类训练使得一个元音类别的原型样例在听

觉皮层的皮层表示范围减小,从而使得相较非原型样例,对元音的原型样例的辨别能力更弱。模型的仿真实验证明,这个模型与听觉神经物理学以及皮层神经生理学的相关数据是相符的,说明该模型是具有神经生理学意义和实际应用价值的。

参考文献:

- [1] Kohonen T. Self-organized formation of topologically correct feature maps[J]. Biological Cybernetics, 1982,43:59-69.
- [2] Guenther F H, Gjaja M N. The perceptual magnet effect as an emergent property of neural map formation[J]. Journal of the Acoustical Society of America, 1996,100:1111-1121.
- [3] Rauschecker J P, Singer W. The effects of early visual experience on the cat's visual cortex and their possible explanation by Hebb synapses[J]. Journal of Physiology, 1981,310:215-239.
- [4] Grossberg S. Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors[J]. Biological Cybernetics, 1976,23:121-134.
- [5] 傅雪,张少白. 一种生长型自组织神经网络的聚类研究[J]. 计算机技术与发展, 2011,21(3):64-66.
- [6] 贺金戈,胡桂明,黄海英. 一种基于自组织神经网络的语音识别系统[J]. 电声技术, 2006(7):56-59.
- [7] 刘雅琴,智爱娟. 几种语音识别特征参数的研究[J]. 计算机技术与发展, 2009,19(12):67-70.
- [8] Zohary E. Population coding of visual stimuli by cortical neurons tuned to more than one dimension[J]. Biological Cybernetics, 1992,66:265-272.
- [9] Guenther F H, Husain F T, Cohen M A, et al. Effects of categorization and discrimination training on auditory perceptual space[J]. Journal of the Acoustical Society of America, 1999,106:2900-2912.
- [10] 易荣庆,李文辉,王铎. 基于自组织神经网络的特征识别[J]. 吉林大学学报(工学版), 2009(1):148-153.
- [11] 苏洪全,朱义胜. 自组织神经网络的参数自适应方法[J]. 计算机工程与应用, 2009(31):22-24.
- [12] 李鑫环,陈立潮,赵红艳,等. 基于多小波分析与 SOFM 的 MR 图像分割算法研究[J]. 计算机技术与发展, 2009,19(9):104-107.
- [13] Guenther F H, Nieto-Castanon A, Ghosh S S, et al. Representation of sound categories in auditory cortical maps[J]. Journal of Speech, Language, and Hearing Research, 2004,47(1):46-57.