CHAPTER 3



第3章　基于发育网络的大脑听觉神经系统模拟

听觉是一种非常重要的感觉功能，生物体通过听觉从外界获取的信息量仅次于通过视觉获取的信息量。对于人类而言，语言的发展和语言交流更是离不开听觉。人的听觉系统对噪声有很强的抵抗能力，人与人的语音交流即使在噪声非常复杂的环境中依然可以正常进行。听觉系统可以从语音中提取各种信息，还可以精确地分析声音的各种属性，并能将语音信号特征信息传递进入大脑而被识别，这是计算机无法比拟的。

人的听觉系统相当复杂，就像一个天然的信息处理系统，其对语音信号的处理和识别能力远超目前计算机的信号处理水平。鉴于人类听觉系统在听音后可以快速、准确地识别声音的内容和方向，许多研究学者通过分析人类听觉系统的结构和信息加工方式，深入探讨了听觉系统在语音识别方面的信息处理机制[44]。以听神经为分水岭，听觉系统可以分为听觉外周和听觉中枢两个组成部分。本章主要介绍人类听觉系统的结构，分析听觉系统各个组成结构的功能以及整个听觉系统对语音信号的加工通路，为后面章节建立仿人听觉系统的发育网络模型提供理论基础和背景知识。

3.1　听觉外周及信息传输通道

听觉外周包括外耳、中耳、内耳，主要负责语音信号的接收和传导，是一个复杂的换能器，能够将机械形式的信号转换为传输到大脑的神经冲动。听觉外周的结构如图3.1所示。

3.1.1　外耳

外耳是听觉外周的第一层，由耳廓、外耳道和鼓膜组成，主要起收集声音的作用。耳廓的功能[45]类似于天线，可以收集并反射声音，使声音有效地聚集到外耳道，对于声音的空间定位具有重要作用。外耳道可以被理想化为一端封闭而另一端开口的管子，是声音传导的通道，其作用是传声和扩音。鼓膜接收来自外耳道的声音信号，并产生振动。

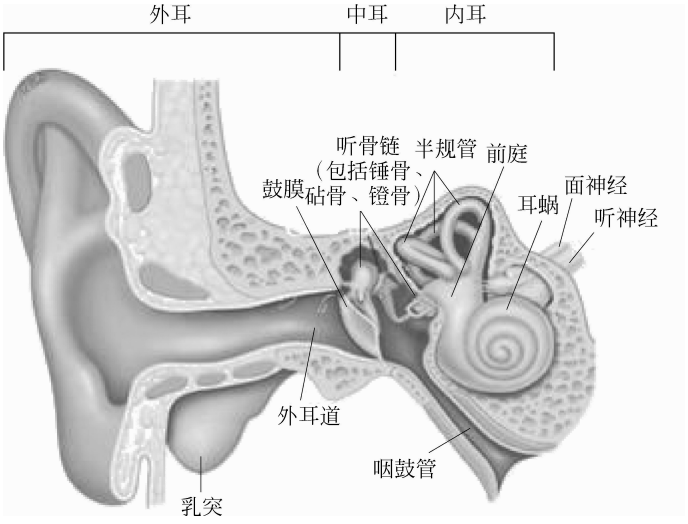


图 3.1　听觉外周结构图

3.1.2　中耳

中耳[46]由鼓室、听骨链和咽鼓管等结构组成。鼓室又称中耳腔，其内部充满了空气；而听骨链由3块听小骨组成，分别为锤骨、砧骨、镫骨；咽鼓管具有维持鼓室内气体压力平衡的能力。中耳一般被认为是传声装置，在声音信号处理方面具有以下三大作用：

（1）传递声音。声音以声波的方式从外耳道传向鼓膜，振动的空气粒子产生压力变化使得鼓膜振动，鼓膜的振动使听骨链做前后移动，从而使声能通过中耳转换成机械能。

（2）声阻抗匹配。由于内耳液体的密度远远大于空气的密度，一般情况下声音是无法经阻抗较小的空气向阻抗较大的液体传播的。如果没有中耳的声阻抗功能，声音能量几乎都会被反射。

（3）声压放大。中耳独特的结构，如同一个杠杆系统，可改变信号的增益。总体来说，中耳对于声波的传输具有重要意义。研究结果表明，中耳如同一个高通滤波器，可以决定声音经过听觉系统的低频阈值。

3.1.3　内耳

内耳[47]包括感受声音信息的耳蜗和感受位置信息的前庭器官。耳蜗是内耳的重要组成部分，也是听觉外周最重要的一个部分。耳蜗是人类听觉系统的感受器，它可将声音的振动信号转换为大脑可以感受和进行处理的神经脉冲信号。人的耳蜗形状如同蜗牛，由螺旋形管道绕中间的蜗轴盘旋而成。耳蜗内含有基底膜，而位于基底膜上的柯蒂氏器包含有毛细胞[48]。研究结果表明，基底膜的运动可以用行波理论[49, 50]来描述，即沿着基底膜分布着不同的特征频率，其值是由基底膜上各个位点之间的参数值决定的；但是对于声音通过基底膜引起的振动而言，要想在基底膜某处产生共振，必须满足振动频率与该处的特征频率相一致的条件。这就说

明了基底膜具有分解声音频率的功能，基底膜可以形成一个频率拓扑以区分输入信号的频率，其功能就如同一组不同中心频率的带通滤波器组。耳蜗毛细胞是听觉感受器细胞，受到声音振动的刺激，听觉毛细胞的膜电位发生变化并释放神经递质，最后使得支配毛细胞的听神经产生兴奋和冲动，将声音信息传到听觉中枢。

从表面上看，耳蜗只是一个简单的换能器，将声音信号转换成电-化学-神经冲动的形式，但实际上，耳蜗不只是简单和被动地对声音做出反应，而是在外毛细胞的主动运动和耳蜗传出神经系统的参与下积极地对外界声音做出极其精细复杂的反应，从而使得耳蜗对声音具有难以置信的频率分辨能力和灵敏度。根据解剖学的研究结果，内毛细胞主要完成听觉信息向中枢的传递，而外毛细胞主要参与听觉信号转换中的离心调节和主动过程，决定了听觉灵敏度和精细辨别能力。内毛细胞和外毛细胞之间形成突触连接到听神经纤维，可将声音由听觉外周传入听觉中枢。

3.1.4 听觉外周信息传导通路

听觉外周主要对声音信号进行传递，以及对频率、强度等信息进行特征编码，也可以理解为对声音特征信息的提取。外界声波通过介质传导到外耳道，再传到鼓膜，鼓膜振动，通过听小骨传到内耳，刺激耳蜗内的纤毛细胞而产生神经冲动。听觉外周的信息传导通路如图3.2所示。

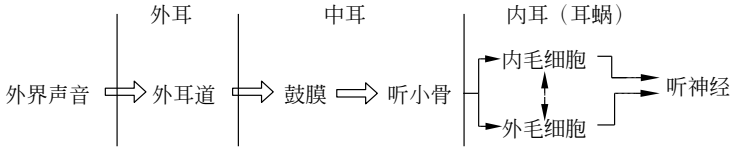


图 3.2 听觉外周的信息传导通路示意图

3.2 听觉中枢及神经传输通道

声音信号经过听觉外周的收集、传导与转换后，以听神经能接受的形式向听觉中枢神经系统传导。听神经是连接听觉外周与听觉中枢神经系统的纽带，也是听觉中枢唯一的信息来源。听神经是声音信号（携带了声音的各种特征信息）的载体，声音信号进入听觉中枢后将进一步被传导、加工、处理与分析。

3.2.1 听觉中枢

听觉中枢主要由耳蜗核（cochlear nucleus）、上橄榄核（superior olivary nucleus）、下丘（inferior colliculus）、内侧膝状体（medial geniculate body）和听觉皮层（auditory cortex）组成。与视觉中枢通路相比，听觉中枢通路[51]上的多个神经核团之间相互交叉连接而呈现出更加复杂的结构。

耳蜗核是听神经进入听觉神经中枢的第一站，它最先对声音信号的频谱特征进行提取，对于听觉信息特征的提取起关键作用。耳蜗核一般可以分为腹侧耳蜗核（VCN）和背侧耳蜗核（DCN）两个部分。这两部分在听神经的传入模式和传出途径上有差别，各自负责处理听觉不同方面的信息，分工合作。听神经进入耳蜗核后会形成上行和下行两个分支：上行听神经进入前腹侧核，下行将听神经投射到后腹侧核和背侧核。耳蜗核内含有大量神经元，这些神经元可保存听神经输入的频谱信息，并具有一定的抑制性和选择性。耳蜗核内的神经处理是听觉中枢对听觉信息加工处理的第一步[52]。

上橄榄核是听觉神经中枢的第二站，它接收耳蜗核的传出纤维，它的发出纤维与耳蜗核的发出纤维组成外侧丘系。上橄榄核可以将声音信息的空间维数进行重组，能够检测声音的位置信息，核内的神经元对信号频率具有锁相作用以及选择性[53]。

下丘位于中脑下部背侧，主要由下丘核及周边的薄层灰质构成。下丘核接收来自外侧丘系的神经纤维，而它发出的神经纤维到达间脑的内侧膝状体。下丘核是听觉中枢传导通路上的重要中继站，并具有音频定位的功能。下丘中的神经元同样具备频率拓扑特性，但神经元的发放率可能与前级神经元有所不同，研究者推测是因为下丘长时间对刺激进行特征表示[54]。

内侧膝状体与下丘的神经元轴突相连并接收下丘传来的听觉纤维，经中继站后发出听辐射，投射到听觉皮层，是在听觉皮层之前整个听觉通路最后一次对神经纤维进行加工[55]。

听觉皮层是听觉神经中枢的最后一站，它接收来自内侧膝状体的点对点输入并呈频率拓扑结构，是听觉信息的处理中枢，可结合大脑对声音进行识别[56]。

3.2.2　神经传导通路

声音是由振源振动并带动周围介质振动产生的，这种声波振动通过外耳和中耳组成的传音系统传递到内耳，经内耳将声波的机械能转变为听神经上的神经冲动，再由听神经传送到大脑皮层听觉中枢，从而产生主观听觉，进而识别声音的内容。根据听神经信息传递的方向，听觉中枢由上行通路和下行通路组成。一条简化的听觉中枢神经传导通路[57]如图3.3所示。

从图3.3可以看出，听觉中枢神经上行传导通路始于耳蜗核，来自耳蜗编码后的听神经信息全部投射并终止于耳蜗核。而耳蜗核又分为耳蜗背核、耳蜗前复核和耳蜗后复核，导致听神经投射从耳蜗核出发分为许多分支而使得通路变得很复杂。由于听觉中枢上行传导通路要穿过多个中枢，故听神经在上行投射中并非一定是逐级向上传递的，也可能是跨级投射或者是跨侧投射。耳蜗复核的多数轴突经过斜方体交叉投射到脑干对侧的上橄榄核，经过换元后向上投射到外侧丘系。再往上的一个重要的听觉核团下丘主要接收来自外侧丘系的投射，部分接收来自耳蜗核的直接投射。下丘是听觉中枢的一个十分重要的中继站，几乎所有的听神经投射都经过下丘再往上投射到更高级的中枢。从下丘出发的神经纤维继续向上投射到丘脑的内侧膝状体，该神经核团是皮层下的高级听觉中枢，从它的复核出发的神经纤维投射到听觉皮层，即听觉中枢传导通路可表示为：听神经 → 耳蜗核 → 上橄榄核 → 下丘 → 内侧膝状体 → 听觉皮层。由

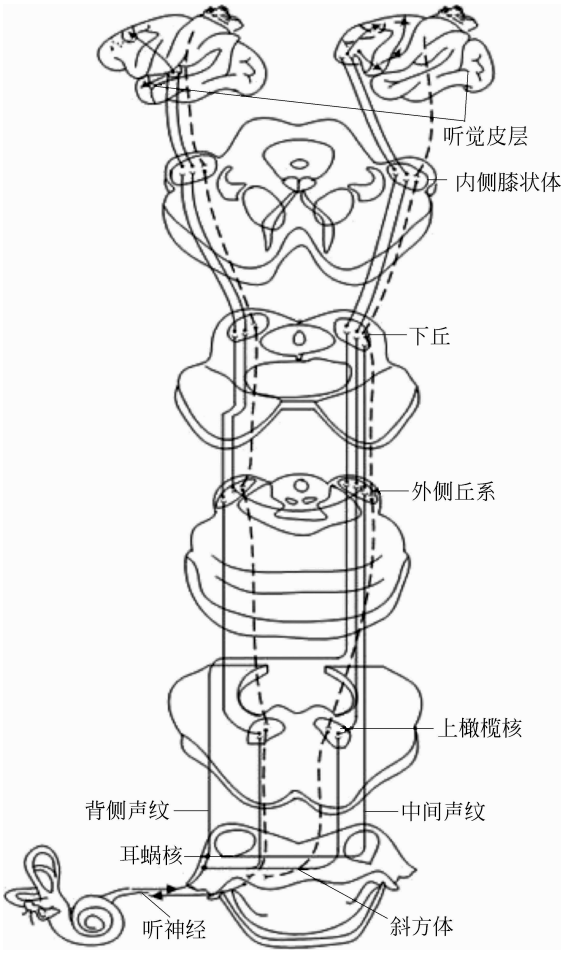


图 3.3　听觉中枢神经传导通路示意图

于听神经进入脑内后不会完全交叉，故听觉信息在上行通路中传递带有双侧性，也就构成了听觉系统信息加工的What通路（内容信息处理通路）和Where通路[58, 59]（空间信息处理通路）。

下行听觉中枢通路起源于听觉皮层或起源于许多其他的听觉核团，终止于各个其他核团。从听觉皮层出发下行投射通路经过多个核团一直往下，最后终止于耳蜗，这些下行投射多数为抑制性的。听觉中枢传导通路对于声音的频率特征表达是以空间的方式在各个结构中形成频率拓扑，频率拓扑结构贯穿于整个听觉中枢，也就是对声音特征信息的表征。听觉中枢传导通路中神经元具有很强的频率锁相特性，但是随着听觉信息沿听觉通路向上传导，神经元发放与频率周期之间的关系变得越来越不明显。听觉中枢存在某种机制，可提高听觉系统的频率选择性。神经元反应的多种多样的模式也说明了听觉系统要完成对听觉信息进行特征提取的各种加工。

3.3　仿人听觉系统的发育网络模型及实现

本节通过模拟听觉信息加工通路的“What”（内容信息）通路，结合发育网络粗略构建仿人听觉系统的发育网络听觉模型，该模型以梅尔频率倒谱系数[60]（Mel Frequency Cepstrum Coefficient，MFCC）模拟听觉外周（耳蜗）的听觉特性作为模型的感知端。

3.3.1　听觉外周的模拟

梅尔频率倒谱系数计算模型是基于人的听觉机理构建的特征计算模型，其模拟了听觉外周对声音频率的感知。为此，以梅尔频率倒谱系数计算模型作为发育网络模型感知端 X 区域的特征信息输入，为 Y 区域的神经元提供学习的特征。本节以语音识别的方式对发育网络模型的性能进行验证和分析，下面先简单介绍梅尔频率倒谱系数特征信息提取的过程。MFCC特征参数提取过程如图3.4所示。

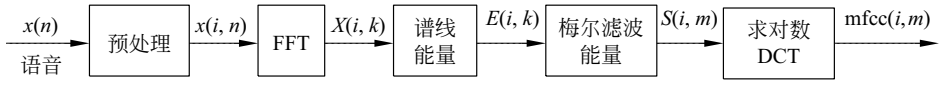


图 3.4　MFCC特征参数提取过程

在语音预处理过程中，第一步是端点检测，它直接影响语音识别性能。在这项工作中，采用的端点检测方法是短时平均能量和短时过零率（ZCR）。

把声音信息记作 s，那么第 n 帧声音信息就是 $s(n)$；把声音信息帧的能量记作 E，第 n 帧声音信息的能量 E_n 按以下公式进行计算：

$$E_n = \sum_{m=-\infty}^{\infty} [s(m)w(n-m)]^2 = \sum_{m=n-N+1}^{n} [s(m)w(n-m)]^2 \tag{3.1}$$

令 $h(n) = w^2(m)$，则 E_n 可以写成如下形式：

$$E_n = \sum_{m=-\infty}^{\infty} s^2(m)h(n-m) = s^2(m) * h(n) \tag{3.2}$$

$h(n)$ 可以看作一个线性滤波器的单位响应。下面给出短时过零率的计算方法：

$$Z_n = \sum_{m=-\infty}^{\infty} |\mathrm{sgn}[s(n)] - \mathrm{sgn}[s(n-1)]|w(n-m) \tag{3.3}$$

$$w(n) = \begin{cases} \dfrac{1}{2N}, & 0 \leqslant n \leqslant N-1 \\ 0, & \text{其他} \end{cases} \tag{3.4}$$

$$\mathrm{sgn}[s(n)] = \begin{cases} 1, & s(n) \geqslant 0 \\ -1, & s(n) < 0 \end{cases} \tag{3.5}$$

图3.5显示了单词“big”的语音信号、相应的短时平均能量（由平均能量表示）和短时间过零率。图3.5显示，短时平均能量和短时过零率切割的语音信号长度不同。短时平均能量容

易将无声声音视为噪声，短时过零率容易将高频噪声误认为语音。为了更好地进行端点检测，这里采用基于短时平均能量和短时过零率的双阈值方法来检测端点。

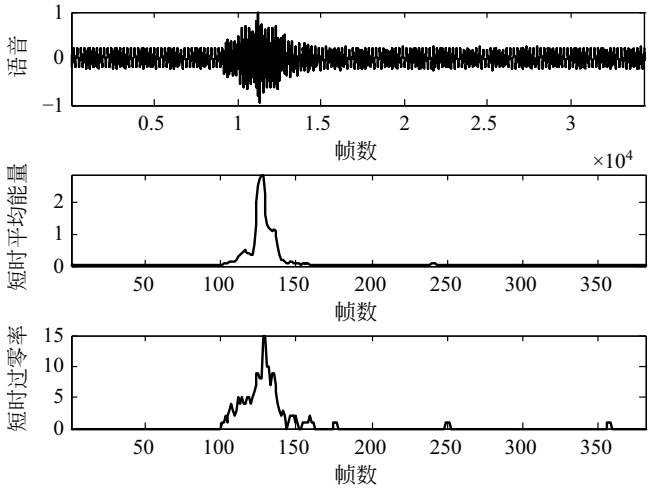


图 3.5　单词“big”的语音信号、短时平均能量和短时过零率

然后，语音信号用一阶数字滤波器进行预加重，其传递函数为$H(Z)=1-\mu Z^{-1}$，其中μ为预加重因子。接下来，执行加帧（256 帧）和加窗（使用汉明窗）、快速傅里叶变换、对数离散余弦变换等操作，最后得到输入语音信号的MFCC参数。由于标准MFCC参数是静态参数，因此只能描述语音信号的静态特征。但人耳不仅对静态特征敏感，对动态特征也敏感，因此本节采用12维MFCC参数及其一阶差分来构建24维MFCC的参数。图3.6提供了单词“big”的可视化MFCC特征。

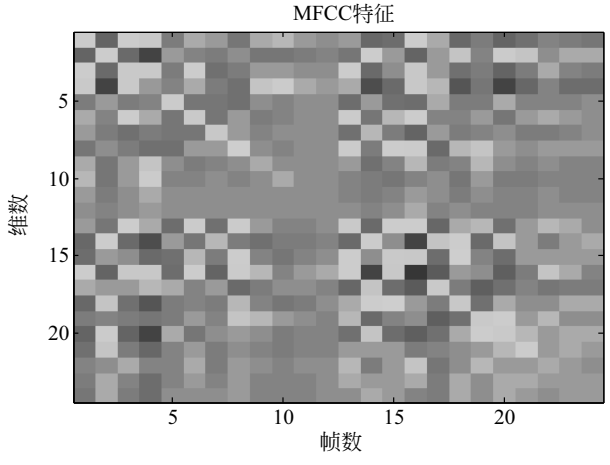


图 3.6　单词“big”的可视化MFCC特征

3.3.2　听觉中枢的模拟

听觉外周可以对声音信息进行传递、转换并且可以对浅层特征进行提取，可作为听觉系统

的感知端。听觉中枢由多个神经核团组成，听神经产生的信息经过听觉中枢的层层加工和整合，最终在听觉皮层被解析和识别，其传导通路为：听神经（耳蜗）→耳蜗核→上橄榄核→下丘→内侧膝状体→听觉皮层。表3.1为听觉中枢核团的初始连接关系表[61]，简单描述了信号在听觉中枢各部分的传输路径。

表 3.1　听觉中枢核团初始连接关系表

听觉细胞核	耳蜗核	上橄榄核	下丘	内侧膝状体	听觉皮层
耳蜗核		√	√		
上橄榄核	∨		√		
下丘	∨	∨		√	
内侧膝状体	∨	∨	∨		√
听觉皮层	∨	∨	∨	∨	

说明：上行信号用√表示；下行信号用∨表示。

从表3.1可以看出，听觉中枢上行传导通路穿过多个中枢的过程中，听神经在进行上行投射时不但出现了逐级向上传递，还出现了跨级投射。也就是说，信号由耳蜗传入听觉中枢时，先传递给耳蜗核，耳蜗核接收信号并对信号进行加工。耳蜗核将信号向上传递给上橄榄核，部分跨级传给下丘，下丘也接收上橄榄核的信号。经过下丘处理的信号传递给内侧膝状体，内侧膝状体处理之后直接传给听觉皮层，最终止于听觉皮层。下行传递起源于听觉皮层或许多其他的听觉核团，是比较复杂的，本节对于下行通路不作过多的讨论。表3.1中的部分连接关系可以转换为听觉中枢上行传导通路，听觉中枢上行传导通路图如图3.7所示。

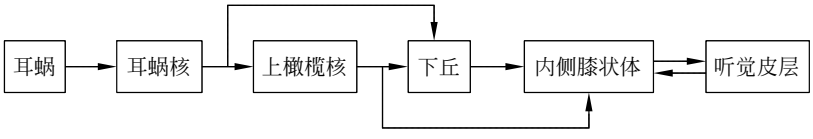


图 3.7　听觉中枢上行传导通路图

通过观察听觉中枢的上行传导通路以及三层结构的发育网络，可以将人类听觉系统粗略地分成3个区域，对应于三层结构发育网络的X、Y、Z区域，进而模拟听觉信息加工通路的“What”（内容信息）通路。可以用X区域表示耳蜗，而耳蜗主要对声音信号进行特征信息处理，作为听觉系统的感知端，与X区域作为网络的感知端相符。将耳蜗核、上橄榄核、下丘、内侧膝状体分别用Y_1、Y_2、Y_3、Y_4表示，这几个中枢可构成Y区域，相当于发育网络的Y区域。在人类听觉系统中，这几个区域和外界环境没有直接的接触，只能通过感知端和执行端对外界信息进行“预测”。而将听觉皮层作为执行端，可以实现对声音信号的识别，对应发育网络的Z区域，如图3.8 所示。

最新的神经科学研究表明，从下丘传出的神经纤维除了传递到内侧膝状体之外，还传递到丘脑中的上丘，而上丘和内侧膝状体都位于丘脑中，它们在处理声音信息方面存在信息交互[61]。针对基本模型中Y区域中Y_1、Y_2、Y_3、Y_4各层之间的连接关系，提出对各层信息进

行综合分析的设想。随后，在Y区域中添加Y_5层，对应于丘脑中的上丘。Y_5层的主要作用是对模型Y区域在学习概念时对概念前后两条特征信息的综合分析，即对概念上下文环境的理解，旨在提高发育网络模型对于新概念的学习能力。而由于Y_5层的存在，整个Y区域依然是“脑壳”封闭的，仍然具有自主发育学习的能力。最终的听觉发育网络模型如图3.8所示。

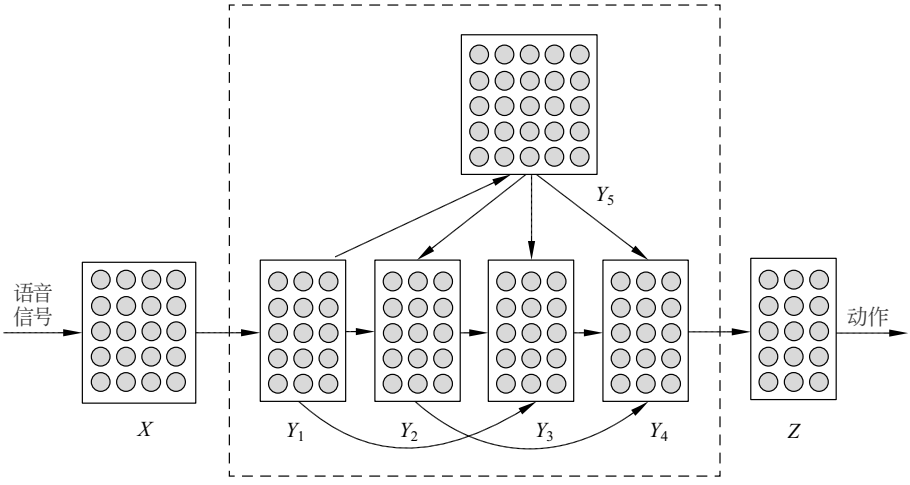


图 3.8　听觉发育网络模型

综上所述，仿人听觉系统的发育网络模型由MFCC和发育网络共同组成。MFCC对听觉外周的特征提取进行模拟，听觉中枢则是采用三层神经网络，MFCC的输出层作为发育网络的X区域输入且Y区域有5层，各层之间以及部分跨层之间也有连接，Z区域作为监督层进行有监督的学习。

3.3.3　听觉系统模拟实验及分析

对于发育网络模型性能的检验，本节以语音识别的方式进行。实验部分主要包括听觉发育网络模型对英文单词识别与短语识别的性能分析；本节通过对英文单词和英文短句的识别，来检验发育网络的语音识别性能。采用自行录制的实验语音样本库，单词和短语录制时分别选择6名和8名志愿者。语音录制时，周边环境略有噪声，录制者正常发音，即人耳可以辨别他们的发音内容，可作为常规的语音样本库。实验中通过自行编辑的程序在MATLAB中运行实现语音的采集，设置的采样频率为10 000Hz，并采用了16位编码的音频格式。录制采集的英文单词有what、where、name、size、small 和big，在6个人中，每个人对每个英文单词连续发音25遍，则总共构成900条英文单词语音数据。在短语录制方面，同样使用相同的程序，以及同样的方法对短语进行录制，录制的短语有good morning、happy new year、how are you、see you、bye bye以及I am fine。每个人对每个短语的发音也是连续25遍，总共构成1200条短语语音数据。实验中，对英文单词和短语语音数据的25遍中的前18遍作为模型训练的样本集，第19~25遍作为模型测试的样本集。

1．训练阶段

首先介绍匹配阈值。对于一个层的每个输入，该层中的每个神经元都有与该输入对应的权重，因此该层中每个神经元都具有与该输入对应的响应。本节采用Top-k竞争机制更新响应值最大的神经元的权重和年龄。然而，若在响应值都不大的情况下，依然对响应值最大的神经元更新其权重和年龄，则会出现误差。因此，设置一个匹配阈值，如果某个神经元的最大响应值大于该阈值，则更新其年龄和权重。如果存在年龄为0的神经元，则选择一个年龄为0的神经元并更新其权重和年龄。如果最大响应值小于该阈值，并且没有年龄为0的神经元，则直接更新响应值最大的神经元的权重和年龄。由前述规则可知，设置的匹配阈值越大，神经元学习到的相关特征越多。

训练流程方案如图3.9所示，具体的训练过程：

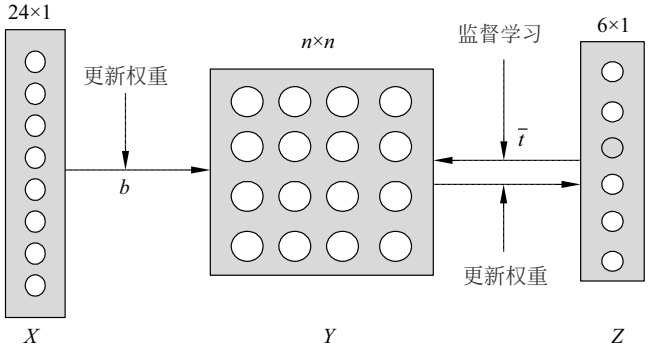


图 3.9　训练流程方案

（1）在训练集中取出一条语音数据，将其归一化后进行MFCC特征提取，输入发育网络的X区域（X区域神经元数目与MFCC特征矩阵相对应）。

（2）计算Y_1层中各个神经元的响应值，并得到响应值最大的神经元的位置m及其最大值r。

（3）如果最大响应值r大于阈值lim，则对该神经元的权重及年龄进行更新。如果上面所得的最大响应值r小于相似度阈值lim，且此时Y_1层中仍有未使用过的神经元（年龄为0），则开辟一个新的神经元当成激活神经元；如果Y_1层中所有神经元都已经更新过一遍，则不用再考虑r和lim的关系，直接激活能量值最大的神经元。将激活神经元的能量值置1，其余神经元置0，得到Y_1层的激活矩阵。

（4）按照权重更新公式，进行神经元权重的更新。激活的神经元的年龄加1，未激活的神经元的年龄不变。

（5）Y_1层的响应值作为Y_2层的输入值，按照步骤(2)～步骤(4)更新Y_2层的神经元权重和年龄。

（6）Y_2层的响应值以及Y_1层的响应值共同作为Y_3层的输入，相同的按照步骤(2)～步骤(4)更新Y_3层的神经元权重和年龄。

（7）Y_3层的响应值、Y_2层的响应值以及Z层的监督值共同作为Y_4层的输入，相同的按照步骤(2)～步骤(4)更新Y_4层的神经元权重和年龄。

（8）Y_4层的响应值作为Z区域的输入值，按照步骤(2)～步骤(4)更新Z区域的神经元权重和年龄。

（9）继续训练下一条语音。

2．测试阶段

如图3.10所示，在测试阶段网络处于“冻结”状态，保留了训练后的权重和年龄。此时，不再向Y区域输入Z区域的监督值，只有X区域向Y区域的输入，根据网络功能自主地预测结果。

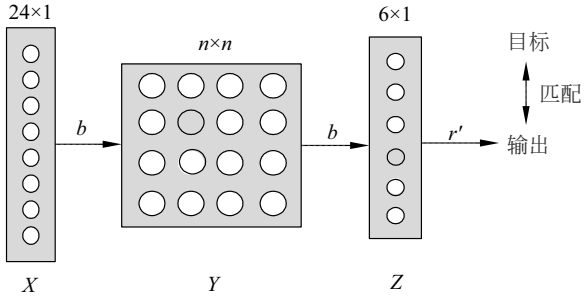


图 3.10　测试流程方案

具体的测试过程如下：

（1）在测试集取出一条语音，归一化以后进行MFCC特征提取，作为网络的输入。

（2）计算Y_1层各个神经元的能量值。

（3）找到能量值最大的神经元的值（这里以k=1为例）及其所在位置，激活此神经元。

（4）将激活的神经元的输出设置为1，其余为0，得到输出矩阵。

（5）将Y_1层的输出矩阵作为Y_2层的输入，按照步骤(2)～步骤(4)计算Y_2层的输出矩阵。

（6）将Y_1层的输出矩阵以及Y_2层的输出矩阵共同作为Y_3层的输入，按照步骤(2)～步骤(4)计算Y_3层的输出矩阵。

（7）将Y_2层的输出矩阵以及Y_3层的输出矩阵共同作为Y_4层的输入，按照步骤(2)～步骤(4)计算Y_4层的输出矩阵。

（8）将Y_4层的输出矩阵作为Z区域的输入，按照步骤(2)～步骤(4)计算Z区域各个神经元的能量值。

（9）Z层中能量最大的神经元激活，将该神经元的输出设置为1，其余设置为0。按照事先标记，得出识别结果。

（10）继续下一条语音，直到测试完待测试集的语音。

（11）统计识别结果，得出识别率，进行对比分析。

3.4　比较实验与分析

3.4.1　单词部分

通过对听觉系统发育网络模型的结构参数进行合理的设置，如Y区域各层神经元的数量以及匹配阈值，选择控制变量的方法进行多次实验旨在得出最优的网络结构参数。首先，选择不同阈值得到模型在不同阈值下的识别率，具体如表3.2所示。

表 3.2　发育网络模型对英文单词的识别率（一）

匹配阈值	0.85	0.9	0.95	0.97	0.98
识别率/%	83.73	92.06	99.21	98.42	95.63

由表3.2可知，随着阈值的增大，发育网络模型识别率呈现先增后减的趋势。当匹配阈值达到0.95时，识别率达到最高，此时识别率为99.21%。出现这种现象的原因是当模型的匹配阈值较低时，导致模型误将两个相关性不大的概念特征视为同一概念，所以模型的性能较差。随着模型匹配阈值的增大，模型对于不同概念特征的辨识能力增强，模型的性能也随之改善，并在匹配阈值为0.95时，模型的性能达到最优。之后模型匹配阈值再增大，由于概念特征的相关性没有匹配阈值高，因此使模型出现识别混乱，模型的性能自然会变差。

其次，控制阈值为0.95不变，选择不同的Y层神经元个数，实验结果如表3.3所示。

表 3.3　发育网络模型对英文单词的识别率（二）

Y区域神经元数量	1000	2000	3000	4000	5000
识别率/%	96.83	97.62	99.21	99.21	99.21

由表3.3可知，当Y层神经元数目为3000时，该模型识别率达到最优，此时识别率为99.21%。随着Y区域神经元数目的增加，发育网络模型正确率为先增后趋于平稳。出现这种现象是当Y区域神经元数目较少时，由于神经元数目太少而将两个并不相关的概念视为一个概念，从而导致模型的识别率较低。当神经元数目增加时，而网络模型识别率不再增加则是因为对于学习这些概念，3000个神经元已足够，若再增加神经元识别率也不会再增加。图3.11给出了发育网络模型中Y_4层神经元对英文单词的学习权重。

综上所述，当匹配阈值为0.95、Y区域神经元个数为3000时，发育网络模型对英文单词的识别率最高，达到99.21%。为了展现Y_4层神经元对英文单词的学习状况，用MATLAB程序得到Y_4层神经元对英文单词概念的学习权重图，如图3.11所示，即发育网络模型中Y_4层神经元对6个英文单词（即Z区域学习概念）记忆学习的积累。类似地，从Z区域神经元的年龄图中可以清晰地观察出每个概念被激活的次数。

从图3.11中可以看出，不同概念在Y_4层神经元学习中出现重叠和不重叠的神经元。这是

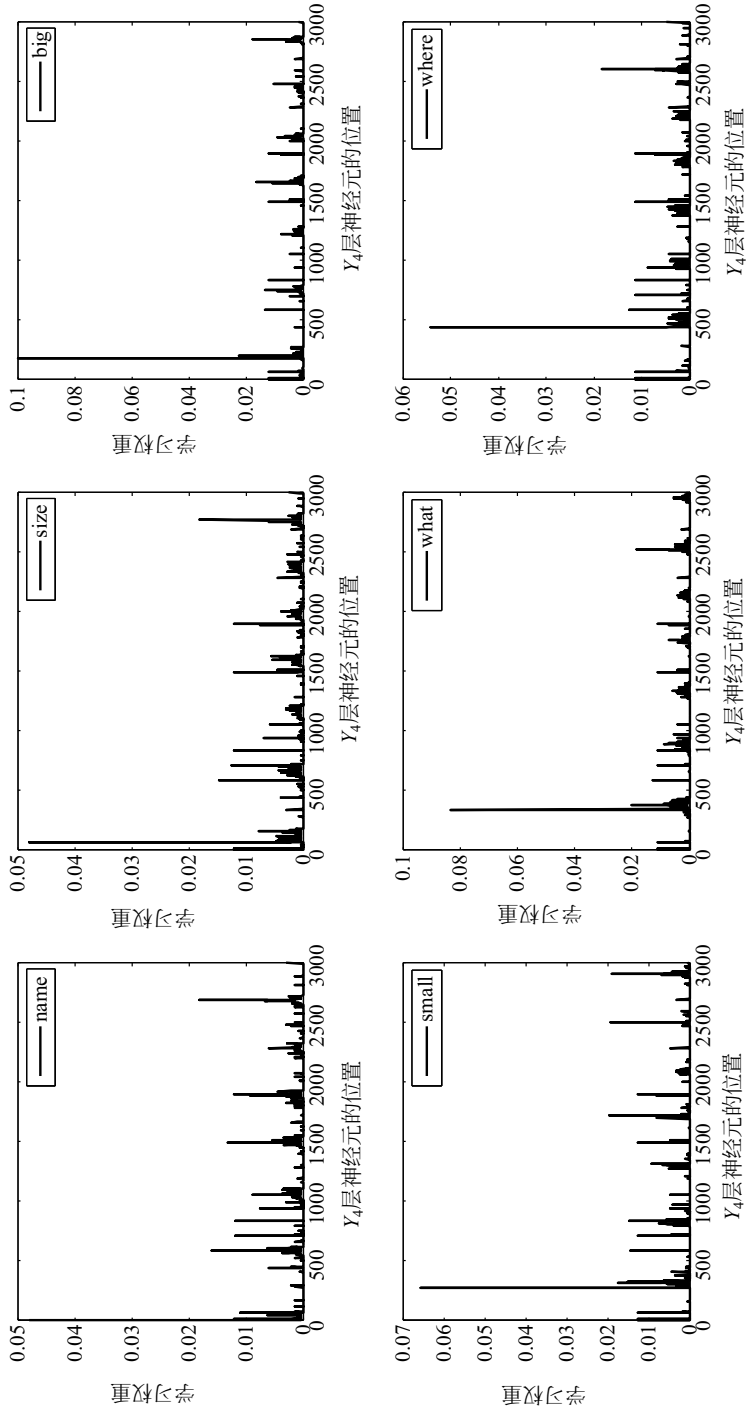


图 3.11 Y_4 层神经元对英文单词概念的学习权重

因为当不同概念的两个特征相关性较大时，某一神经元就会将其归为一类并更新该神经元的权重，这也是模型性能下降的一个因素。从图3.11中还可以看出，几乎每个神经元都被激活并对概念进行学习，有小部分神经元未被使用，这说明实验中设置的神经元数量是充足的。图3.11中某些神经元的权重比较大，造成这种现象的原因是对于不同单词，其所含的特征信息量不尽相同，而对于同一单词，其所含的特征信息也会有所差别。

对于概念特征的差异性，也可由Z区域神经元的学习年龄体现出来，如图3.12所示，图中Z区域神经元位置对应第i个（i=1,2,⋯,6）概念。因为对每个单词进行了1~18遍训练，从理论上来讲，Z区域神经元的年龄应一致。从图3.12中可以看出，Z区域各个神经元年龄相差不大，这也从另一个方面反映出该模型对英文单词的识别效果较好。

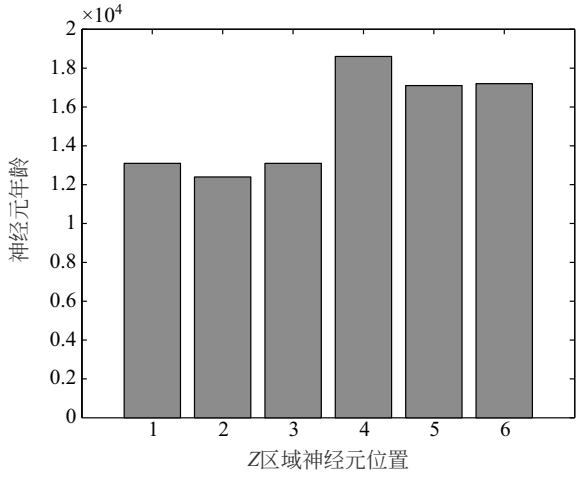


图 3.12　Z区域神经元年龄图

3.4.2　短语部分

以上分析了发育网络模型对于英文单词的识别性能，关于模型对短语的识别性能，表3.4给出了发育网络模型对短语识别的正确率。

表 3.4　发育网络模型对英文短语的识别率（一）

匹配阈值	0.80	0.85	0.90	0.95	0.98
识别率/%	77.68	85.12	85.17	86.31	82.44

发育网络模型对于短语的识别与英文单词具有类似的规律，只是模型对于短语的识别性能低于英文单词的识别性能，这是由于短语的信息比英文单词复杂。从表3.4同样可以看出，发育网络模型在匹配阈值为0.95时出现了最优值。同样，使阈值为0.95不变，Y区域神经元数目不断改变，实验结果如表3.5所示。

表 3.5 发育网络模型对英文短语的识别率（二）

Y 区域神经元数目	2000	3000	4000	5000	6000
识别率/%	81.04	82.14	86.31	86.31	86.31

从表3.5可以看出，当神经元达到4000时，识别率达到最高。神经元数目由低到高的过程中识别率则呈现出先增加后趋于平稳的状态。综上所述，当阈值为0.95、Y 区域神经元个数为4000时，发育网络模型对英文短语的识别率达到最高为86.31%。

与单词的识别类似，图3.13和图3.14分别给出了 Y_4 层神经元对英文短语的学习权重以及 Z 区域神经元的年龄。

从图3.13可以看出，不同概念在 Y_4 层神经元学习中出现重叠的神经元相对较多，这也就是短语单词的识别率低的原因。英文短语概念间的差异性可通过图3.14直观体现。

从图3.14可以看出，Z 区域神经元年龄参差不齐，这也暗示了该模型对英文短语的识别率不高。

3.4.3 对比实验

为了证明“MFCC ＋发育网络”模型在语音识别中的可靠性，与文献 [62] 中的方法进行比较，比较结果如表3.6所示。

表 3.6 不同方法对英文单词识别的正确率

模型	单词识别正确率/%
文献 [62] 的方法	96.3
MFCC ＋发育网络	99.21

由于文献 [62] 未对英文短语进行试验，本节利用作者录制的短语样本，实验结果如表3.7所示。

表 3.7 不同方法对英文短语识别的正确率

模型	短语识别正确率/%
文献 [62] 的方法	43.75
MFCC ＋发育网络	86.31

从表3.6和表3.7中可以清晰地看出：“MFCC ＋发育网络”模型在语音识别中有较好的应用效果。虽然文献 [62] 中的模型也是对人类听觉系统的模拟，但我们的模型更接近人类真实的听觉系统，因此其识别性能优于文献 [62]。

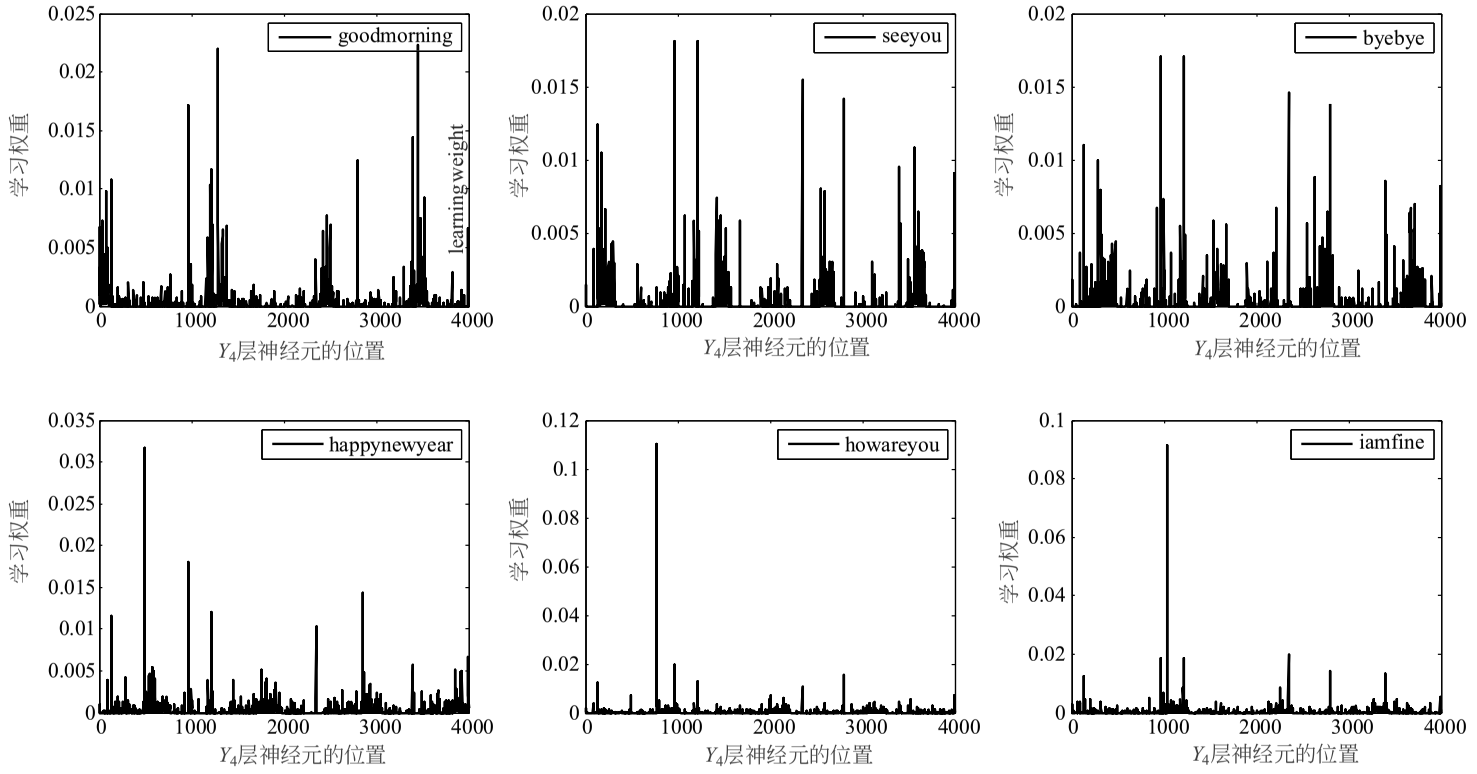


图 3.13　Y_4 层神经元对英文短语的学习权重

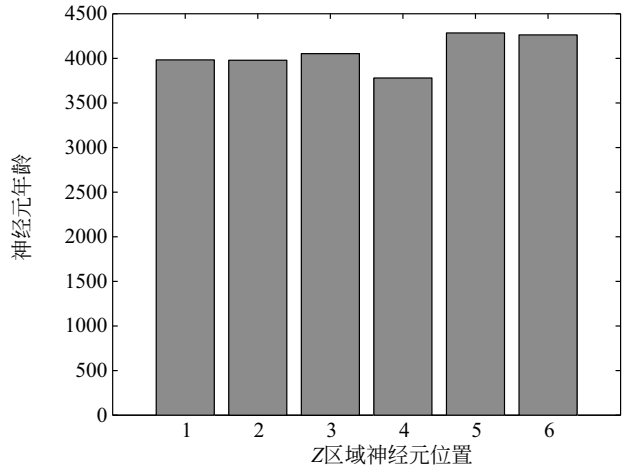


图 **3.14** Z 区域神经元年龄图