

文章编号: 0427-7104(2021)03-0279-09

# 人工耳蜗的音高感知编码机制和限制

孟庆林, 周华莉, 余光正

(华南理工大学 物理与光电学院 声学研究所, 广东 广州 510641)

**摘 要:** 本文综述了健听者和人工耳蜗植入者的外周听觉系统中的音高感知机制, 讨论了人工耳蜗音高感知研究的现状, 并介绍了改善人工耳蜗音高感知的研究方向和进展.

**关键词:** 音高感知; 人工耳蜗; 位置编码; 时域编码; 心理物理学

**中图分类号:** O428

**文献标志码:** A

DOI:10.15943/j.cnki.fdx-b-jns.2021.03.002

## 1 音高编码机制

音高是听觉感知的基本维度之一, 在音乐欣赏、声调和语调感知以及听觉场景分析中具有重要作用. 音高(Pitch, 或译为“音调”)可以被定义为“一种听觉属性, 依该属性, 可以将声音进行从高到低的排序”. 在音乐中, 音符之间的音高变化构成了音乐的旋律; 在语音中, 音节之间的音高变化会影响语调, 可以用来表达比如陈述或疑问的语气信息; 在声调语言中, 单个音节之内的音高随时间变化规律的不同, 形成不同的声调, 可以用来表达不同的含义; 在噪声场景中, 音高的不同可被用于分离不同的声源, 从而有助于噪声中的言语感知. 了解音高感知机制, 不仅有助于开发或改进各种与音高任务相关的音频音乐信号处理算法, 也有助于改进助听器、人工耳蜗等人工听觉设备对听力障碍者的助听效果.

### 1.1 健听者的音高编码机制

纯音(Pure tone)只有一个频率成分(即单独一个正弦信号), 频率越高, 产生的音高也越高. 纯音的频率可以通过耳蜗内基底膜上兴奋的峰值位置进行编码. 不同频率的纯音会引起基底膜上不同位置的兴奋, 可以认为纯音音高由基底膜上最兴奋的位置决定, 即“位置理论”. 但有研究者发现, 随着纯音声级的升高, 基底膜上最兴奋的位置会呈现一定的偏移现象, 而感受到的音高基本维持不变<sup>[1-2]</sup>. 另一个关于纯音感知的理论是“时间理论”, 与听神经的相位锁定有关. 在约 4~5 kHz 的上限频率以下, 听神经倾向于在刺激波形的某一特定相位进行放电, 因此, 听神经的放电时间间隔约为刺激信号周期的整数倍, 这种方式也提供了对纯音频率(或音高)的编码.

复合音(Complex tone)中包含多个正弦成分. 日常生活中常见的声音, 比如乐器音和语音, 包含丰富的谐波复合音(Harmonic complex tone). 谐波复合音中的正弦成分的频率都是某个频率  $F_0$  的整数倍, 其中  $F_0$  被称为基频值. 乐音和语音中的谐波复合音通常包含频率为  $F_0, 2F_0, 3F_0, 4F_0, \dots$  等连续若干个谐波成分. 谐波复合音虽然由多个谐波成分组成, 但是通常认为在每一瞬间它只会诱发一个音高感受, 而不是多个不同的音高感受, 其音高通常与以  $F_0$  为频率的纯音的音高非常接近. 即便如此, 复合音的音高并不是简单地由频率为  $F_0$  的基频成分(即第 1 次谐波)决定的. 当把第 1 次谐波从谐波复合音中去除后, 复合音的音高并没有变化, 这就是所谓的“基频缺失”(Missing fundamental)现象. 基频缺失在生活中也不少见, 例如如果一个小型扬声器不能播放出 150 Hz 以下的声音成分, 那么  $F_0$  为 100 Hz 的谐波复合音经

**收稿日期:** 2021-01-02

**基金项目:** 国家自然科学基金(11704129, 61771320); 广东省基础与应用基础研究基金(2020A1515010386); 广州市科技计划项目(202102020944)

**作者简介:** 孟庆林(1986—), 男, 博士, 副教授; 周华莉, 女, 硕士, 通信作者, E-mail: tuobamao@qq.com; 余光正, 男, 博士, 教授, 并列通信作者, E-mail: scgzyu@scut.edu.cn

由该扬声器播放后的声音就不包含基频成分,但这并不影响相应的音高感受.实际上,谐波复合音中并没有哪个谐波成分是对音高感受起决定性作用的,当噪声对谐波复合音产生污染时,残留的谐波成分(可能是在时间和频率维度上都是断续的)仍可以保证听者的音高感知不变.

听者对谐波复合音音高的感知机理可以根据耳蜗对声音的时频编码规律来分析.基底膜上不同位置对不同的频率成分产生最大响应,而这种频率-位置映射关系不是均匀地线性划分的,如果把每个位置对输入声音的响应过程看作是一个听觉滤波器,那么对更低频率(即更靠近蜗尖)产生响应的听觉滤波器的绝对频率带宽是更窄的.这种规律导致耳蜗对低频区域的频率分辨率更好.图 1(a)所示为 13 个谐波复合音组成的音字符串,每个复合音包含前 7 次谐波.对于谐波复合音来说,低次谐波可能被基底膜不同位置处的听觉滤波器分别进行处理,而高次谐波则会出现多次谐波进入同一个听觉滤波器的情况.例如,图 1(c)所示为中心频率在 50~8 000 Hz 之间的 80 个 Gammatone 听觉滤波器输出的波形,代表着相应位置处基底膜的振动情况.听神经对低次谐波的音高编码就同时包含位置编码(即不同谐波在基底膜上的不同位置产生最大响应)和时间编码(听神经放电的相位锁定),对高次谐波的编码主要为时间编码(因为多次谐波的叠加形成的波形会呈现出以  $1/F_0$  为周期的时域包络,听神经会对该包络进行相位锁定).其中低次谐波对音高感知的贡献较强,高次谐波的贡献较弱.由此可见,  $F_0$  的相关信息广泛分布于基底膜(或听神经)的不同位置,尤其是听神经对不同位置处的单个谐波或多个混合谐波的相位锁定会使得听神经的电脉冲间隔中大量分布着基频值的信息.正因为  $F_0$  的相关信息在听神经中不同位置的广泛分布,谐波复合音的音高感知是强烈地与以  $F_0$  为频率的纯音音高相同,且不容易受其他因素干扰.图 1(f),(g),(h)展示了  $F_0$  在正常耳蜗中的处理情况.

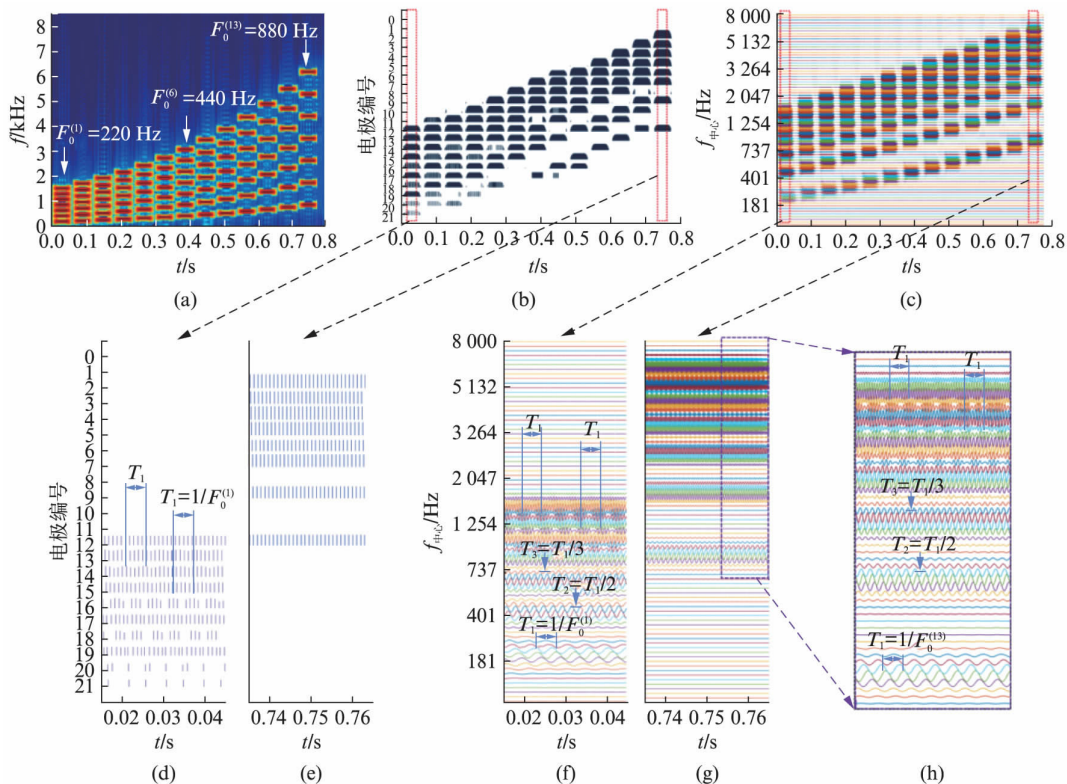


图 1 正常耳蜗和人工耳蜗对谐波复合音的编码示意图

Fig. 1 Coding of complex tones in a normal cochlear and a cochlear implant

(a) 13 个复合音组成的音字符串,相邻两个音之间的音程为两个半音,最低音和最高音的基频分别为 220 Hz 和 880 Hz(相差两个倍频程);(b) 某人工耳蜗对(a)中声音处理后的电极图;(c) 80 通路 Gammatone 滤波器组所代表的正常耳蜗的基底膜对(a)声音的响应图;(d)–(g) 分别为(b)和(c)中相应区域的局部放大显示;(h) 为(g)中相应区域的局部放大显示.

谐波复合音中谐波成分的强度比例(或谐波复合音的频谱包络形状)很大程度上决定了音色(Timbre)感知,不会对音高感知产生明显的影响.音色差异的典型例子是不同乐器的声音、不同元音、不同说话人的讲

话等. 因此, 可以认为频谱包络影响了音色感知, 频谱精细结构(即每个谐波的具体频率位置)影响了音高感知. 然而, 根据前述分析, 这里频谱的精细结构主要是通过低频可分辨谐波(每个谐波的周期性波动)和高频不可分辨谐波(局部多个谐波叠加形成的与  $F_0$  同步的时域周期性)被相应区域的听神经以相位锁定的形式进行编码的. 因此, 时域周期性(Periodicity)信息在音高编码中起到了决定性作用.

在心理声学实验中常用的纯音是人为设计出来的; 谐波复合音是在音乐、语音等声音中普遍自然存在的. 纯音可以看作是谐波复合音的特例. 纯音在声学研究和临床听力学中有大量应用, 而谐波复合音对于实际生活中的音高感知问题更具有现实意义.

## 1.2 人工耳蜗的音高编码机制

人工耳蜗是一种帮助重度以上感音神经性听力损失者恢复或获得听力的植入式听觉辅助设备, 其全球植入人数已经超过 80 万人. 人工耳蜗主要由 3 部分组成: 体外的言语处理器(通常包括一个或多个传声器, 一个将声音信号编码成适当的刺激信号的处理器和一个发射线圈), 植入的接收器/刺激器(接收处理器发送的信号并解码, 生成相应的电信号)和插入耳蜗内的电极阵列. 其基本原理为: 通过体外的传声器拾声信号, 然后在信号处理器中对声信息进行编码, 再通过植入耳蜗内的电极传递电流刺激听神经, 从而产生听觉. 人工耳蜗重建了耳蜗中毛细胞的声电转换功能, 将外部声环境与听觉神经通路重新连接.

人工耳蜗对声信号的编码方法通常称为信号处理策略, 目前常用的人工耳蜗信号处理策略, 比如连续间隔采样(Continuous Interleaved Sampling, CIS)<sup>[3]</sup>和高级混合编码(Advanced Combinational Encoder, ACE)<sup>[4]</sup>, 主要的信号处理流程如图 2 所示. 声音信号首先经过带通滤波器组分成分若干通道(通常在 24 以内, 依品牌而异, 每个通道对应一个电极), 分别提取各通道内的时域包络, 再进行非线性压缩以匹配植入者的动态范围, 然后用压缩过的包络对恒定速率的双相电脉冲串进行幅度调制, 用于各个电极的刺激. 总体来说, 人工耳蜗中的信号处理策略主要传递了有限通道内的时域包络信息.

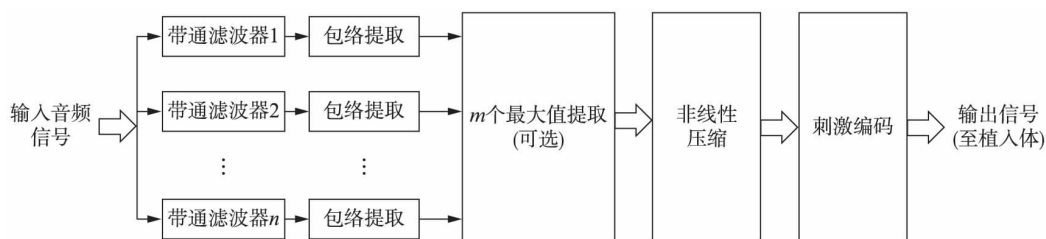


图 2 人工耳蜗的信号处理策略流程示意图

Fig. 2 Functional block diagram of the signal processing strategy in cochlear implants

在人工耳蜗中, 声音的频谱包络通过分布在耳蜗内的各个电极的刺激幅度及其相对大小来表达, 这是基于正常听觉系统的“位置编码”设计的. 然而目前的人工耳蜗电极数较少(12 到 24 个), 要将较宽的声音频率范围分配到这样少数的电极上, 能实现的频率分辨率是很低的. 由于电极位置距离听神经较远, 电极之间还会有电场干扰. 电极植入位置、频率分配、听神经特征频率之间的不匹配情况也会影响频域信息的精确表达. 另外, 如果采用的带通滤波器的幅频响应不是理想的矩形, 那么通道间就会存在频谱的泄露. 这些原因综合在一起, 导致人工耳蜗的位置编码是很粗糙的. 关于人工耳蜗中的核心信号处理技术可以参见文献[5].

纯音的音高(或频率)的变化, 在正常耳蜗中表现为基底膜最大响应位置的变化和听神经相位锁定信息的变化; 在人工耳蜗中则主要表现为放电电极位置的变化, 由于人工耳蜗信号处理策略是基于时域包络的, 那么纯音的时域周期性基本在人工耳蜗电刺激信号中得不到体现(除了在几百赫兹以下的纯音). 由于人工耳蜗位置编码的粗糙性, 纯音在人工耳蜗中的放电位置的调谐特性相比于正常耳蜗也粗糙许多, 甚至在中低频段通常一个纯音可以诱发两个或更多电极产生幅度较大的刺激(这与电听觉电流动态范围较窄有关).

谐波复合音的音高变化, 如前 1.1 节所述, 在正常耳蜗中通过高度的频率选择性和良好的时域跟踪能

力,以基频  $F_0$  信息编码的形式得到了丰富且一致的表达;在人工耳蜗中则只有粗糙的频率选择性和较弱的时域跟踪能力(详见 1.3 节讨论),导致  $F_0$  相关信息的表达被弱化.图 1(b)所示为(a)图对应的谐波复合音经过一个 ACE 策略后的电极图(横坐标为时间,纵坐标为电极号,短线高度代表脉冲幅度).人工耳蜗对谐波复合音的  $F_0$  编码主要表现为每个电极通道中的电刺激信号包络中包含的时域周期性信息.这种与基频值同步的电信号的时域周期性被听神经以相位锁定的方式进行编码,进而可能提供一定的音高感受,但是只局限于较低频率,如图 1(d)只对应于 220 Hz 的  $F_0$ ,而在高基频时编码很差;图 1(e)是  $F_0 = 880$  Hz 时对应的电极图,没有显示与  $F_0$  相应的时域周期性表达.其中仅存的少量与基频值同步的时域周期性是源自较宽的带通滤波器中进入的多个谐波叠加后的波形形式.人工耳蜗利用时域周期性进行音高感知的机理与健听者在高频区的音高感知机理是相似的.但是,相比于健听者的低频区音高编码,这种多谐波叠加后的周期性所传递的音高感知,无论对于健听者的高频区还是对于人工耳蜗的所有频段都是较弱的.仿真研究显示,人工耳蜗的有效通道数如果能增加 2 到 4 倍,音乐的音高变化才有可能得到较好地表达<sup>[6]</sup>.

总之,基于现有的人工耳蜗信号处理策略的水平,人工耳蜗的音高感知会明显弱于健听者.从对谐波复合音的分析来看,人工耳蜗在音乐、语音等声音的音高感知任务中都可能遇到显著的困难.

### 1.3 单电极的时域编码

在常用的人工耳蜗信号处理策略中,所有通道内的电脉冲刺激速率是恒定值(例如 900 Hz 或更高).因此,在人工耳蜗植入者日常聆听时只能利用电脉冲的幅度起伏来获取声音中的信息.在人工耳蜗的研究历史上,大量的研究采用了单电极的心理物理学实验,通过人为地精确操控单个电极上的脉冲发放时刻、脉冲幅度、脉冲波形等来开展电听觉编码规律的研究.尤其是,其中发放时刻和幅度的变化都可以表达时域听觉信息.有对动物的电生理研究发现单个听神经对频率高达至少 12 kHz 的电刺激依然存在相位锁定<sup>[7]</sup>,然而,心理物理研究发现多数人工耳蜗植入者对单个电极上电刺激的时域信息变化感知能力在约 300 Hz(少数人可以达到 800 Hz 左右)时会出现饱和,当电刺激的重复速率超过这个上限时,多数植入者不能辨别出时域信息的变化<sup>[8-10]</sup>.目前多数人工耳蜗信号处理策略中,各通道提取的包络也大致在 300 Hz 以内.电听觉的这个时域检测能力的上限(300~800 Hz)远低于正常耳蜗中听神经相位锁定的频率上限.人工耳蜗信号处理策略只保留时域包络,丢弃了时域精细结构信息,有一部分原因就是来自于电听觉的这个时域限制.语音的基频范围大约在 50~500 Hz 之间,其中一部分仍然在电听觉时域限制范围以内,因此时域周期性中表达的,约 300 Hz 以下的基频  $F_0$  相关信息,仍可能被植入者利用来对音高进行感知.

根据上述分析可知,人工耳蜗植入者利用有限的时域编码能力,可以对约 300 Hz 以下的  $F_0$  对应的音高进行编码,但这种编码能力是远弱于健听者的音高编码能力.另外,单电极上的这种时域音高编码还会受到刺激位置的影响,在靠近蜗尖的电极上进行时域编码可能比靠近蜗底的电极上更好<sup>[11]</sup>.

### 1.4 位置编码的作用:是音高,还是音色?

值得注意的是,虽然在文献中,“位置音高”(Place pitch)的说法很常见.比如有一些实验在不改变刺激信号的时域信息的条件下变化刺激的位置,引导受试者比较两个或多个声音的音高高低,即比较“位置音高”.另外,还有研究发现,通过同时刺激两个电极位置可以诱发出介于两个“位置音高”之间的音高感受,这项技术被用于某些人工耳蜗系统中,寄希望于能提供更精细的音高(或频率)分辨率.但是“位置音高”的研究通常不能排除音高以外的听觉属性的影响,比如音色.

对于谐波复合音来说,如前所述,音色受到频谱包络的显著影响,而人工耳蜗中电刺激位置相关的电能量分布表达了粗略的频谱包络信息,因此刺激位置的分布变化在很多情况下直接影响的是音色感知变化,而不是音高感知的变化<sup>[12-13]</sup>.例如,Nelson 等<sup>[14]</sup>在 1995 年的一项关于“位置音高”和言语感知的研究中,发现受试者报告在变化刺激电极时,声音会变“尖”,提出与改变刺激电极相关的感知维度可能是音色而不是音高.Marimuthu<sup>[15]</sup>在植入者的“位置音高”感知任务和健听者对声音的“明亮度”感知任务中观察到了类似的趋势,提出可能植入者在“位置音高”感知任务中感知到的实际是声音的“明亮度”,而不是音高.有研究表明,通过感知实验结合多维尺度分析方法,发现刺激位置和刺激速率呈现出类似于音色和音高之间的独立关系.当两者协同变化时,通常有助于音高感知任务的完成;当两者产生冲突时,音高感知

任务也会受到明显的影响.这两个感知属性呈现正交关系<sup>[16]</sup>.Reiss 课题组近年来开展了一系列关于双耳音高融合(Binaural pitch fusion)的研究,发现人工耳蜗的植入者非常容易把两侧耳听到的具有不同音高的两个声音,融合听为一个具有平均音高的声音<sup>[17-18]</sup>.这种现象也许可以用音色感知来解释.

在语音信号中,频谱中的共振峰分布差异可以用于区分不同的元音,这种能力也可以被认为是听者对音色的区分和范畴化感知.共振峰与基频是没有必然联系的,共振峰由声道的共振特性决定,而基频由声门的开合速率决定.例如,有研究者采用双元音进行音高感知研究<sup>[19]</sup>,其背后的原因就是双元音中包含复杂的共振峰滑动变化,在双元音中音色和音高的升降趋势可能出现冲突,这种复杂的变化会给人工耳蜗植入者的音高感知带来挑战.

因此,虽然在大量的基础心理物理实验中,“位置音高”和“时间音高”被广泛用于解释人工耳蜗的音高编码机制,但是从实际生活中常见的语音和音乐等音高感知任务中可以观察到,“位置音高”说法的意义较小,还往往与音色感知相混淆,需要在研究中引起注意.音色与音高在人工耳蜗植入者的各项感知任务中的贡献权重也是近期的研究热点<sup>[20-22]</sup>.

## 2 人工耳蜗的音高感知能力

大多数人工耳蜗用户在安静环境中都能获得较好的言语识别能力,但是在音高感知相关任务中的表现与健听者有较大差距.最直接的证据来自音乐音高感知实验.另外,此种差异还体现在语音音高信息的提取方面(下面以汉语的声调感知为例进行介绍),更进一步会影响复杂声场景中的语音感知.

### 2.1 音乐音高感知

在音乐中,两个音符之间的最小音程为一个半音,对应的基频差别约为6%.然而,大部分人工耳蜗植入者的基频差别分辨阈限均大于这一值.Goldsworthy<sup>[23]</sup>采用自适应的方法测试了一组人工耳蜗植入者(9名)和一组性别年龄均与植入者组匹配的健听受试者(9名),对经过带通滤波的谐波复合音(基频为110,220和440 Hz)的基频差别阈限进行测量,发现植入者的基频差别阈限几何均值为12.5%,显著高于健听者组的1.4%.Marx等<sup>[24]</sup>对比了植入者(10名)和健听者(15名)对5个谐波组成的复合音(基频为110,220,400,500和750 Hz)的差别阈限,发现植入者的基频差别阈限均值为34.0%,远高于健听者的2.2%.人工耳蜗植入者的基频分辨差别阈限高于一个半音对应的约6%,提示大部分植入者可能都无法准确感知音乐中的音高变化,即旋律.在人工耳蜗植入者的旋律感知研究<sup>[25]</sup>中,的确发现植入者在准确感知音乐旋律上存在很大困难.

以上例子是关于谐波复合音的基频分辨能力的,而旋律感知所依赖的不仅仅是区别音高差别,还需要能准确地感知音高之间的音程关系(例如一个八度)<sup>[12]</sup>.由于电极位置的失配和音高编码的时域上限,这种音程关系也无法得到有效表达.因此,在很宽的音乐音高范围内,基频分辨和音程关系表达方面的缺陷导致人工耳蜗植入者对音乐音高的感知能力很差,不能有效地进行旋律感知.对于双侧植入者来说,如果两侧音高的感知不匹配,双耳聆听比单耳聆听在旋律感知中有表现更差的风险<sup>[26]</sup>.

### 2.2 声调感知

在声调语言中,音高在音节内随时间变化的轮廓可以用来传递不同的语义信息.汉语普通话的典型声调有4个,即阴平、阳平、上声、去声,其音高或基频轮廓分别为高平、升、降-升、降.汉语语音的音高除了传递声调信息,也可以像其他语言一样传递语调、情感、说话人的特质等.语音的基频大致在50到500 Hz之间,根据前述人工耳蜗的时域音高理论,利用时域周期性信息可以进行一定程度的语音音高编码.已有数据表明,与健听儿童相比,植入人工耳蜗的儿童声调识别要差一些<sup>[27]</sup>.人工耳蜗植入者的声调识别能力有较大的个体间差异,其中表现优秀的植入者可以获得非常好的识别率(例如大于90%)<sup>[28]</sup>.

从声调的声学特征的角度来看,音高或 $F_0$ 的相关信息是主导健听者进行声调感知的主要特征.已有研究表明,除了 $F_0$ 这个主要线索外,还有一些次要线索可能在某些特殊情况下辅助声调识别,例如响度轮廓(或称为振幅轮廓)、音节时长、音色轮廓等.对于健听者来说,有研究者采用人为去除部分 $F_0$ 相关信息来进行声调测试,观察到了次要线索的贡献<sup>[29-31]</sup>.而对于人工耳蜗植入者来说,他们的 $F_0$ 编码本来就比较弱,因此次要线索的影响和贡献就自然成为另一个重要的研究课题.各个声学特征对于人工耳蜗声

调感知的贡献大小仍需要进行详细的实验研究,并且对具有不同听觉经验的人工耳蜗的受试者来说,声学特征的贡献比例也可能存在明显不同。

### 2.3 听觉场景分析

健听者可以在嘈杂的声音场景中听懂讲者的讲话,在历史上这类现象被称为“鸡尾酒会现象”(Cocktail party phenomenon),在心理声学的研究体系中与“听觉场景分析”(Auditory scene analysis)密切相关。研究发现,音高差异有助于听者从两个说话人中提取其中某一说话人的语音信息,例如在一男一女同时说话的场景中听一个男声比在两个男性同时讲话时听其中一个男声的难度更低。如前所述,健听者的音高( $F_0$ )信息广泛分布于耳蜗基底膜上和听神经中,而语音信号在时间上是断续的,在频率上有时也是离散的线谱,因此两个语音同时发生时它们之间不容易产生很强的能量掩蔽。当说话人  $F_0$  范围差距较大时,听者可以根据倾听需求整合具有相似  $F_0$  信息的时频区域为目标语音,同时忽略其他  $F_0$  偏离较大的区域。更多相关信息可以参见文献[32]。

人工耳蜗的频率分辨率较低、时间检测能力也有限,这些因素导致  $F_0$  只能通过通道内的时域周期性得到较弱的表达。当两个说话人同时讲话时,两个语音流的谐波很容易在人工耳蜗通道内产生叠加,这既进一步破坏了谐波的可分辨性,也破坏了残存的时域周期性信息。因此,人工耳蜗植入者在噪声干扰条件下的语音识别能力会明显下降。例如 Chen 等<sup>[33]</sup>对比了健听者和人工耳蜗植入者在多人同时交谈的语音中的语音提取能力,发现说话人的性别差异(即目标语音和背景干扰为不同性别说话人的语音与两者为同性别说话人的语音进行对比)让健听者可以获得高达 12 dB 的优势,而人工耳蜗植入者获得的优势则仅为 2 dB。

## 3 改善人工耳蜗音高感知能力的研究方向和进展

由以上讨论可见,人工耳蜗的音高感知能力与健听者有较大差距。另外,植入者个体之间的差距也很大,通过特定的训练可以帮助改善部分人工耳蜗用户在一些心理物理任务中的音高感知<sup>[34-36]</sup>。从技术层面看,如何改进信号处理策略以改善人工耳蜗音高感知能力是目前的一个研究热点。具体的改进方向列举如下:

a) 幅度调制:在 CIS 策略中,对时域包络进行额外的幅度调制,其中幅度调制的频率与当前声音中的基频值相等,如 Laneau 等<sup>[37-38]</sup>提出的 F0mod 策略,Vandali 等<sup>[39-40]</sup>提出的 eTone 策略(后更名为 OPAL(Optimized pitch and language)策略)等。Fu 等<sup>[36]</sup>采用一种长电极(MED-EL 品牌,插入深度为 1.5~2.0 圈)探究了不同电极位置上用 100~1500 Hz 的速率进行刺激,发现越靠近蜗底的电极在低刺激速率时更容易引起不舒适的音质感受,因此作者认为在低频电极进行时域增强会更有利。

b) 脉冲发放时刻控制:在一些低频通道中,在过零点或峰值点发放电脉冲,用非均匀的脉冲速率来取代原来的固定脉冲速率,原始声音信号中的部分时域精细结构通过电脉冲的发放时间显式地表达。如奥地利 MED-EL 公司的 FSP(Fine structure processing)策略<sup>[41]</sup>,Van Hoesel 等<sup>[42]</sup>提出的 PDT(Peak derived timing)策略等。

c) 基于抖动的听觉重启:在固定高速率的电脉冲串中插入少量电脉冲(以基频为周期间隔),如 Lindenbeck 等<sup>[43]</sup>的研究,寄希望于通过少量不规则抖动时刻的脉冲来重启或加强听神经对时域信息的灵敏度。该方法源于对双侧人工耳蜗的双耳时间差的增强研究。研究者认为刺激速率分辨和双耳时间差分辨在生理机制上有相似性,的确,初步实验数据也支持了此假设。

d) 隐式的时域精细结构增强:通过移频的方法,将部分原本快变的时域精细结构隐式地转换为慢变的时域精细结构,如 Nie 和 Li 等<sup>[44-47]</sup>提出的 HSSE(Harmonic-single-side-band encoder)策略和 Meng 等<sup>[48-49]</sup>提出的 TLE(Temporal limits encoder)策略。

这些方向都是基于现有的电极水平进行的时域编码调整,从发表的文献来看,它们都在一定程度上提升了某方面的音高感知能力,需要结合更多的临床实验来进行信号处理算法的优化。

## 4 结 语

健听者的音高编码依赖于正常耳蜗中基底膜上或听神经中  $F_0$ 。相关信息的广泛分布,其中主要包括:

1) 对每个低频可分辨谐波的位置编码和相位锁定的时间编码;2) 对高频不可分辨谐波的听觉滤波器通道内与  $F_0$  同步的时域周期性的相位锁定的时间编码。

人工耳蜗是通过植入耳蜗内的少数电极刺激听神经来帮助听力损失者恢复听力,其中电极数目和电极刺激听觉时域能力的局限性导致  $F_0$  的相关信息不能得到充分编码,仅在通道的时域周期性信息上得到反映。电极数目和电极间电场干扰限制了对频率的分辨率,但是电极间的分辨是否代表音高差别是有争议的,其中还涉及到音色感知属性的问题。人工耳蜗植入者对于音高(与基频同步的周期性)和音色(与共振特性相关的共振峰)的感知和依赖程度与正常听力者,甚至与其他听力损失人群的区别是研究热点。

人工耳蜗植入者的音高感知能力较弱,这直接影响了他们在音乐音高感知、声调感知、干扰噪声中的语音感知等。这方面的算法改进也是人工耳蜗领域的研究热点,需要进一步结合临床应用开展更多的理论和实验研究工作。

#### 参考文献:

- [1] ZWISLOCKI J J. What is the cochlear place code for pitch? [J]. *Acta Otolaryngology*, 1991, **111**(2): 256-262.
- [2] ZWISLOCKI J J, NGUYEN M. Place code for pitch: A necessary revision [J]. *Acta Otolaryngology*, 1999, **119**(2): 140-145.
- [3] WILSON B S, FINLEY C C, LAWSON D T, et al. Better speech recognition with cochlear implants [J]. *Nature*, 1991, **352**(6332): 236.
- [4] VANDALI A E, WHITFORD L A, PLANT K L, et al. Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system [J]. *Ear and Hearing*, 2000, **21**(6): 608-624.
- [5] 孟庆林. 人工耳蜗中的声信号处理 [J]. 听力学及言语疾病杂志, 2018, **26**(4): 431-436.
- [6] MEHTA A H, OXENHAM A J. Vocoder simulations explain complex pitch perception limitations experienced by cochlear implant users [J]. *Journal of the Association for Research in Otolaryngology*, 2017, **18**(6): 789-802.
- [7] GLASS I. Phase-locked responses of cochlear nucleus units to electrical stimulation through a cochlear implant [J]. *Experimental Brain Research*, 1984, **55**(2): 386-390.
- [8] ZENG F G. Temporal pitch in electric hearing [J]. *Hearing Research*, 2002, **174**(1/2): 101-106.
- [9] MCKAY C M, MCDERMOTT H J, CLARK G M. Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees [J]. *The Journal of the Acoustical Society of America*, 1994, **96**(5): 2664-2673.
- [10] TOWNSHEND B, COTTER N, VAN COMPERNOLLE D, et al. Pitch perception by cochlear implant subjects [J]. *The Journal of the Acoustical Society of America*, 1987, **82**(1): 106-115.
- [11] STAHL P, MACHEREY O, MEUNIER S, et al. Rate discrimination at low pulse rates in normal-hearing and cochlear implant listeners: Influence of intracochlear stimulation site [J]. *Journal of the Acoustical Society of America*, 2016, **139**(4): 1578-1591.
- [12] TODD A E, MERTENS G, VAN DE HEYNING P, et al. Encoding a melody using only Temporal information for cochlear-implant and normal-hearing listeners [J]. *Trends in Hearing*, 2017, **21**: 1-15.
- [13] MAAREFVAND M, BLAMEY P J, MAROZEAU J. Pitch matching in bimodal cochlear implant patients: Effects of frequency, spectral envelope, and level [J]. *The Journal of the Acoustical Society of America*, 2017, **142**(5): 2854.
- [14] NELSON D A, VAN TASSELL D J, SCHRODER A C, et al. Electrode ranking of "place pitch" and speech recognition in electrical hearing [J]. *The Journal of the Acoustical Society of America*, 1995, **98**(4): 1987-1999.
- [15] MARIMUTHU V M R. Exploring place pitch & temporal pitch perception with cochlear implants [D]. Sydney, Australia: Macquarie University, 2013.
- [16] LANDSBERGER D M, MAROZEAU J, MERTENS G, et al. The relationship between time and place

- coding with cochlear implants with long electrode arrays [J]. *The Journal of the Acoustical Society of America*, 2018, **144**(6): EL509-EL514.
- [17] HARTLING C L, FOWLER J R, STARK G N, et al. Binaural pitch fusion in children with normal hearing, Hearing Aids, and Cochlear Implants [J]. *Ear and Hearing*, 2020, **41**(6): 1545-1559.
- [18] REISS L A J, FOWLER J R, HARTLING C L, et al. Binaural pitch fusion in bilateral cochlear implant users [J]. *Ear and Hearing*, 2018, **39**(2): 390-397.
- [19] GREEN T, FAULKNER A, ROSEN S. Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants [J]. *The Journal of the Acoustical Society of America*, 2004, **116**(4): 2298-2310.
- [20] WINN M B, MOORE A N. Perceptual weighting of acoustic cues for accommodating gender-related talker differences heard by listeners with normal hearing and with cochlear implants [J]. *The Journal of the Acoustical Society of America*, 2020, **148**(2): 496-510.
- [21] SKUK V G, KIRCHEN L, OBERHOFFNER T, et al. Parameter-specific morphing reveals contributions of timbre and fundamental frequency cues to the perception of voice gender and age in cochlear implant users [J]. *Journal of Speech, Language, and Hearing Research*, 2020, **63**(9): 3155-3175.
- [22] LUO X, SOSLOWSKY S, PULLING K R. Interaction between pitch and timbre perception in normal-hearing listeners and cochlear implant users [J]. *Journal of the Association for Research in Otolaryngology*, 2019, **20**(1): 57-72.
- [23] GOLDSWORTHY R L. Correlations between pitch and phoneme perception in cochlear implant users and their normal hearing peers [J]. *Journal of the Association for Research in Otolaryngology*, 2015, **16**(6): 797-809.
- [24] MARX M, JAMES C, FOXTON J, et al. Speech prosody perception in cochlear implant users with and without residual hearing [J]. *Ear and Hearing*, 2015, **36**(2): 239-248.
- [25] MCDERMOTT H J. Music perception with cochlear implants: A review [J]. *Trends in Hearing*, 2004, **8**(2): 49-82.
- [26] ARONOFF J M, KIRCHNER A, ABBS E, et al. When singing with cochlear implants, are two ears worse than one for perilingually/postlingually deaf individuals? [J]. *The Journal of the Acoustical Society of America*, 2018, **143**(6): EL503-EL508.
- [27] PENG S C, TOMBLIN J B, CHEUNG H, et al. Perception and production of mandarin tones in prelingually deaf children with cochlear implants [J]. *Ear and Hearing*, 2004, **25**(3): 251-264.
- [28] TAO D, DENG R, JIANG Y, et al. Melodic pitch perception and lexical tone perception in Mandarin-speaking cochlear implant users [J]. *Ear and Hearing*, 2015, **36**(1): 102-110.
- [29] LUO X, FU Q J. Enhancing Chinese tone recognition by manipulating amplitude envelope: Implications for cochlear implants [J]. *The Journal of the Acoustical Society of America*, 2004, **116**(6): 3659-3667.
- [30] WHALEN D H, XU Y. Information for Mandarin tones in the amplitude contour and in brief segments [J]. *Phonetica*, 1992, **49**(1): 25-47.
- [31] FU Q J, ZENG F G, SHANNON R V, et al. Importance of tonal envelope cues in Chinese speech recognition [J]. *The Journal of the Acoustical Society of America*, 1998, **104**(1): 505-510.
- [32] OXENHAM A J. Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants [J]. *Trends Amplif*, 2008, **12**(4): 316-331.
- [33] CHEN B, SHI Y, ZHANG L, et al. Masking effects in the perception of multiple simultaneous talkers in normal-hearing and cochlear implant listeners [J]. *Trends in Hearing*, 2020, **24**(1): 1-12.
- [34] VANDALI A, SLY D, COWAN R, et al. Training of cochlear implant users to improve pitch perception in the presence of competing place cues [J]. *Ear and Hearing*, 2015, **36**(2): e1-e13.
- [35] GOLDSWORTHY R L, SHANNON R V. Training improves cochlear implant rate discrimination on a psychophysical task [J]. *The Journal of the Acoustical Society of America*, 2014, **135**(1): 334-341.
- [36] FU Q J, GALVIN J J, WANG X, et al. Benefits of music training in mandarin-speaking pediatric

- cochlear implant users [J]. *Journal of Speech, Language, and Hearing Research*, 2015, **58**(1): 163-169.
- [37] LANEAU J, WOUTERS J, MOONEN M. Improved music perception with explicit pitch coding in cochlear implants [J]. *Audiology & Neuro-Otology*, 2006, **11**(1): 38-52.
- [38] MILCZYNSKI M, WOUTERS J, VAN WIERINGEN A. Improved fundamental frequency coding in cochlear implant signal processing [J]. *The Journal of the Acoustical Society of America*, 2009, **125**(4): 2260-2271.
- [39] VANDALI A E, VAN HOESEL R J. Enhancement of temporal cues to pitch in cochlear implants: Effects on pitch ranking [J]. *The Journal of the Acoustical Society of America*, 2012, **132**(1): 392-402.
- [40] VANDALI A E, DAWSON P W, ARORA K. Results using the OPAL strategy in Mandarin speaking cochlear implant recipients [J]. *International Journal of Audiology*, 2017, **56**(S2): S74-S85.
- [41] HOCHMAIR I, NOPP P, JOLLY C, et al. MED-EL cochlear implants: State of the art and a glimpse into the future [J]. *Trends in Hearing*, 2006, **10**(4): 201-219.
- [42] VAN HOESEL R J M, TYLER R S. Speech perception, localization, and lateralization with bilateral cochlear implants [J]. *The Journal of the Acoustical Society of America*, 2003, **113**(3): 1617-1630.
- [43] LINDENBECK M J, LABACK B, MAJDAK P, et al. Temporal-pitch sensitivity in electric hearing with amplitude modulation and inserted pulses with short inter-pulse intervals [J]. *The Journal of the Acoustical Society of America*, 2020, **147**(2): 777-793.
- [44] LI X, NIE K, IMENNOV N S, et al. Improved perception of music with a harmonic based algorithm for cochlear implants [J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 2013, **21**(4): 684-694.
- [45] LI X, NIE K, IMENNOV N S, et al. Improved perception of speech in noise and Mandarin tones with acoustic simulations of harmonic coding for cochlear implants [J]. *Journal of the Acoustical Society of America*, 2012, **132**(5): 3387-3398.
- [46] LI X, NIE K, ATLAS L, et al. Harmonic coherent demodulation for improving sound coding in cochlear implants [C]//Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. Dallas, USA: IEEE, 2010: 5462-5465.
- [47] NIE K, ATLAS L, RUBINSTEIN J. Single sideband encoder for music coding in cochlear implants [C]//Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing. Las Vegas, USA: IEEE, 2010: 4209-4212.
- [48] MENG Q, ZHENG N, LI X. Mandarin speech-in-noise and tone recognition using vocoder simulations of the temporal limits encoder for cochlear implants [J]. *The Journal of the Acoustical Society of America*, 2016, **139**(1): 301-310.
- [49] MENG Q, ZHENG N, LI X. A temporal limits encoder for cochlear implants [C]//Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing. South Brisbane, Australia: IEEE, 2015: 5863-5867.

(下转第 296 页)

the three-dimensional scatterer articulation trajectory and spatial displacement between normal subjects and patients with dysarthria of different disease degrees by using phrase text. It is found that the articulation movement of the tongue of patients with dysarthria is closer to the back, left and lower part of the mouth, and the more serious the disease is. Secondly, the centroid of articulation movement space is calculated by K-means algorithm, and significance analysis is conducted. It is found that the difference between different degrees of dysarthria is mainly in the up-down direction, and the average significance level in this direction is only 0.078. The experimental results show that the classification recognition performance of random forest classifier is more than 98%, and the classification accuracy of random forest is 6.45% higher than that of J48.

**Keywords:** dysarthria; spatial characteristics of articulation; disease classification

---

(上接第 287 页)

## Pitch Perception with Cochlear Implants: Coding Mechanisms and Limitations

MENG Qinglin, ZHOU Huali, YU Guangzheng

(Acoustics Laboratory, School of Physics and Optoelectronics, South China University of  
Technology, Guangzhou, Guangdong 510641, China)

**Abstract:** This paper reviews the peripheral pitch perception mechanism of normal hearing listeners and cochlear implant listeners, discusses the performance of pitch perception with cochlear implants, and reviews approaches to improve pitch perception of cochlear implants.

**Keywords:** pitch perception; cochlear implant; place coding; temporal coding; psychophysics