参赛队员姓名：李顺

中学：华南师范大学附属中学

省份：广东省

国家/地区：中国

指导教师姓名：唐杰、刘桦

论文题目：音高识别神经机理与仿生探究

论文题目：音高识别神经机理与仿生探究

摘要：

音高是听觉中枢神经核心感知要素之一。音高识别机理至今仍是生物物理和神经生理中的活跃研究课题。本文从物理和生理基础出发，综合分析了与音高识别机理相关的理论和实验数据，确认了在耳蜗声场中波前的驱动下，同源复合音各成分在耳蜗不同部位可以同时同步锁相触发耳蜗内毛细胞上的传入神经发放。这组信号通过一组具有和声倍频关系的神经通道传至皮层，若在这组通道中出现神经首发放信号，初级听皮层中的整合神经元将抑制减弱不同时出现的信号，易化加强同时出现的信号，传送至次级听皮层中的频选神经元提取基音作为这组信号的音高，不存在从单神经通道中的神经发放率变化规律中提取音高的机制。这一音高识别机理解释了为何电子耳蜗植入者能听清语音却无法听清音乐的原因，指出了人工耳蜗的改进方向是寻找更好的脑机接口。本研究项目依据这些原理，开发了一个识别率较高的“音乐转五线谱”仿生转换程序。

关键词：

波前，泛音列，锁相，频率调制曲线，抑制，易化，整合神经元，频选神经元，人工耳蜗，傅立叶变换，深度学习

Title: The Neural Mechanism of Pitch Recognition and a Bionic Explore

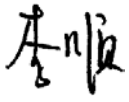
Abstract：

The pitch is one of the fundamental perceptional elements of auditory nerves. The mechanism of pitch recognition is still an active research topic in both biophysics and neurophysiology. Based on the physical and physiological basis, this article analyzes the related theory and experimental data about the mechanism of pitch recognition, and identifies that, driven by wave-fronts of sound field in the cochlea, all the components of complex sounds can simultaneously and synchronously phase-locked trigger the afferent nerve spikes of inner hair cells on the different positions of cochlea. These signals will be transmitted to the cortex by a set of nerves channel with harmonic relationship and temporal coincidence. When the first spike signals appear in these channels, the harmonic-selective neurons at primary auditory cortex will inhibit non-concurrent signals and facilitate the concurrent signals, while the pitch-selective neurons at secondary auditory cortex will extract their fundamental frequency as the pitch, thus completing the process of pitch recognition. According to this theory that we proposed, we can explain why people with electronic cochlear implant can't hear the music clearly while they can hear speech. The article also points out that the direction of advancement in cochlear implants, which should focus on realizing a better brain-computer interface. In addition, based on these elemental principles, we developed a bionic conversion program that transfers audio to music notes with a high recognition rate.

Key words:

Wave-front, Harmonic series, Phase-locking, Frequency turning curve, Suppression, Facilitation, Harmonic-selective neurons, Pitch-selective neurons, Cochlear implants, Fourier Transform, Deep learning

本参赛团队声明所提交的论文是在指导老师下进行的研究工作和取得的研究成果。尽本团队所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人或本团队已经发表或撰写过的研究成果。若有不实之处，本人愿意承担一切相关责任。

参赛团队签名： 

日 期： 2016 年 9 月 1 日

目 录

音高识别神经机理与仿生探究

1. 引言

自然界充斥着各种声音，声音是音源的周期性振动能量变化状态在媒介中的传播，故又称声波。声波是重要的信息载体，通过波幅、频率以及这两者随时间的动态变化，可以携带几乎无穷的信息。其中就包括对我们日常生活非常重要的语音信息和音乐信息。

既然声波的能量表达形式单一，在工程上我们很容易将其数字化。不管多么复杂，所有的声音信息都可以包含在标明声波时间、幅度和频率的数字中。这一技术已经在我们的生活中得到广泛应用。然而一个有趣的现象是，尽管有些声音信息在数字表征中几乎完全不同，但是在人的听觉感受中却表达出完全相同的信息。例如：在诵读一个相同词语时，朗诵者发音的频率、声音强度以及语速可以完全不同，但是听众却能识别是同一个词；同一段音乐旋律，由不同的乐器以不同音调、强度以及速度来演奏，听众也能听出是同一首曲。这说明，我们的听觉系统，并不是被动的接收和识别所有的声音参数，而是对异常复杂的声音信息进行主动的处理，提取出最能表征出声波所包含信息的那些参数，从而提高我们对外界信息反应的敏感度和准确性。

通过信息技术手段仿真模拟听觉系统的相关功能，是当前人工智能研究的热点，特别是语音识别、音乐识别和语义识别，已经成为对未来信息技术发展有重大影响的三大课题。近几年随着在云计算、大数据支持下的深度学习算法和人工神经网络的发展，语音识别已经取得了重大突破，个别技术指标甚至超过了生物听觉系统。不过，深度学习算法仅是在计算形式上“自上而下”地模拟了生物听觉系统的工作流程[1]，并不研究大脑本身的运作机理，缺少令人信服的在生物神经元中关于机器学习规则的解释[2]，对声音信息的识别还不能完全尽如人意，如在音高、音色这类基础声音要素的识别上并无优势。神经科学家们通过各种检测分析手段，特别是神经生理和神经影像学等技术，“自下而上”地分析动物和人听觉系统的工作机制，发现这些基础的声音要素的识别均在相同神经层级及相似的神经功能核团中实现，且具有密切关联。

但是，在音乐识别以及与其它音高相关的应用中，我们并没获得与语音识别相似的成功。例如：同样以深度学习算法应用于音乐识别软件，其识别效果却远低于语音识别软件；电子耳蜗植入者的语音识别能力已经接近 90%，但音乐识别率却仅 50%左右。这是为什么呢？难道语音和音乐的识别机制并不相同？

为此，我们还需要更深入地了解各种声音要素识别的神经机制，这不仅有利于我们认识自身，更重要的是还具有广泛的应用前景。对于听觉障碍的病人，人工耳蜗植入往往是最后的治疗手段。提高人工耳蜗装置的各种声音识别能力，将大大改善患者术后的效果。此外，日益应用广泛的人工智能需要在人机对话中更准确的识别相关声音信号中的有用信息，包括语音、语调、语义和情绪等，以使得这一技术更为可靠。研究人和动物听觉神经系统对声音信号，尤其是音高信息的识别机制，将为我们提供解决以上问题的重要线索。根据这些机制研发的产品和技术也将能“无缝”地投入到应用中，直接服务人类社会。

本研究课题希望通过借鉴生物听觉系统的音高识别机理，找到一种能快速进行音源分离和音高识别的计算机算法。我们注意到对于构成音乐基础的同源复合音，“声场波前振动的物理相位”与“传入神经发放的锁相机制”有密切联系。通过分析生物听觉系统的反

应特征，我们认为：同源复合音的各频率成分同时触发神经信号。同源必同时，是人工智能的基本假定，也是生物听觉系统的基本特征。基于这一发现，我们通过推断音高识别的神经机理，进而提出了电子耳蜗音乐识别率低的可能原因，并据此改进了一种仿生的计算机音高识别算法，开发了一个将音乐直接转换为五线谱的实用软件，取得了比较满意的识别效果。

2. 内耳音高信息的转换

音高反映听觉神经对声场周期性特征的感知，是一种特定的神经编码，而不是可测度的物理量。声音频率与音高密切关联，单频纯音音高基本就对应此纯音频率。不过，音高与频率之间并非简单的一一对应：音高随声强变化轻微变化，高、低频段的变化更显著；缺基频乐音尽管此时基频处无实际的能量分布，所识别的音高还是基频音；对用低频调制高频载波产生的调幅音，所感知的音高是低频调制频率，即为这组音频的等距频差。这些现象分别关联了音高识别机理的某些特征。

音高识别是一个从物理声音信号中提取并重组与音高表征相关的神经信号。耳蜗将声波振动转化为毛细胞上的细胞电位活动，并通过突触传递给神经细胞，因此是声场物理变量与大脑神经发放之间的桥梁。建立合理的耳蜗感音换能模型，追踪声波特性与耳蜗细胞反应之间的内在联系，是分析音高识别机理的基础。

与仅研究短纯音的音高识别不同，本文着重研究的是同源复合音，它是由一组关联的倍频频率成分复合而成，绝大多数的自然音有这一特点。对于同源复合音，分量之间的时相同步性是问题焦点。同源必同时同相，否则将不能合成为同一个音，进行正常的音高识别。在计算模型中，同源复合音成分间的时相同步均为缺省假设。在生理模型中，是否确实存在这一时相同步性呢？这是我们分析内耳音高信息转换功能的主要线索。

2.1. 音高物理信号的表征

为什么绝大多数的自然音是同源复合音呢？这是音源振动发音的特点决定的。当单一发音体触发振动后，起始振动在自身及共振腔的受限空间中传播并相互干涉形成了音源的受迫振动，它具有暂稳态的驻波振动能量分布特点。这种振动状态通过周边媒介以纵波(疏密波)方式向外传播，就形成了同源复合音。同源复合音具有如下特征[3]：

一、频谱分布具有等距倍频特征。仅驻波波长与音源振动体总长成整数比的振动得到加强。这组振动频率的最小频差恰好是基频(即这组振动的最低振动频率)。频差相等是同源音频信号的普遍特征，在周期性脉冲音和低频调制音中，也有类似频谱分布结构，只是这一频差不对应特定基频。

二、时相分布具有同源同相特征。同源复合音的各振动分量均源于同一发声体的受迫振动，故所有这些同源振动分量的相位(即特定时刻的初始振动相位)始终相同。当解除受迫振动或振动向外自由传播时，不同频率分量在不同时刻的振动相位将各不相同，难以简洁地给出相关数学描述，但唯一不会改变的是：在特定时刻前，各个振动分量均处于相同的初始相位。

三、能量分布具有分量递减特征。随着驻波波长变短，音源振动体被细分，驻波波节增加，驻波波腹振幅递减，故振动能量主要分布于较低频分振动，且基频振幅最大。显然，振动体细分能量分配是有限度的，不应无限细分，故有限的低频分量能量分布已经能有效表示振动体的能量总和。

耦合波动是声音特征变量的载体。振动状态在媒介中的传播，本质上就是媒介中振动质点间的能量耦合。这些质点一旦被激励，其振动状态将同时受到两种因素共同影响：新波前振动的耦合激励和已激发振动的时变规律。当不断有波前到达时，质点相当于受迫振动，主要影响因素为前者；否则，主要运动规律受制于后者。在开放空间，忽略声场阻尼因素，根据波动的独立传播原理和惠更斯-菲涅耳原理，这些特征将保持在波前上的所有振动点：即任一新到波前所引起的振动，均能够完整地包含音源的特定时刻前的原始振动状态。在封闭空间，声场与边界的耦合状况将决定声场中各质点的最终振动状态，通常称这种声场为驻波场。耦合状况大致分正阻尼、零阻尼和负阻尼三种情况，一般物理声场边界均介于正阻尼和零阻尼之间，出现负阻尼的情况不多。负阻尼对声场振动有增强作用，它能增加耦合程度，缩短耦合时间。耳蜗的主动机制就恰好属于这种情况。

尽管音源振动规律明显，但考虑时变因素后进行简单描述并不容易。例如：对于同源复合音，初始振动的能量分布中总是基频分量最大，但这组分量因频率不同而相位各异，故总可能在特定时点出现谐波的瞬时能量大于基频的情况。另外，由于它们互相之间的非线性耦合关系非常复杂，很难找到一种简捷的算法快速回朔它们是否来自同一音源[4]。

不过，既然声场中波前上的质点振动始终呈现出非常简洁的相互关系：相位相同，能量分布相对固定。因此，只要能准确检验出波前的组合振动状态，就能直接、准确和完整地还原音源的原始振动状态组合。这是同源复合音的音高感知的关键物理基础。

2.2. 传入神经信号的特征

声音物理信号为时域信号，听觉神经信号为频域信号。耳蜗作为傅里叶变换工具，其响应时延在毫秒级以下，可实时、且无失真地将声场波前的时域信号，转换为对应的频域神经信号。耳蜗可视为具有柔性边界的封闭声学腔体。在外界声场的激励下，腔体内柔性的基底膜上将形成一个瞬变的整体驻波场。每一时刻声场中各个音源的波前会瞬时到达基底膜所有激励部位，各部位的基底膜也会实时响应激励。即瞬时声场波前决定基底膜驻波场的振动分布，而这一振动触发相应的神经信号发放。

这一功能实现机理归结于两个关键的生理机制的发现：

一是行波理论。Békésy 根据强音刺激的实验现象[5]：传入振动以基底膜行波方式向蜗顶推移，基底膜的物理特性使行波振动幅度渐大，达到极值后迅速衰减并消失，就像抖动绸带上传播的行波，故称为行波模型。行波模型准确揭示了基底膜的声场响应的物理特征，找到了基底膜上“频率-部位”的对应关系的源头。对于简单纯音信号，其对应的频率调制曲线基本拟合。

二是主动机制。行波理论本身是有局限的，首先，在耦合程度上，神经频率调制曲线与基底膜频率调制曲线的拟合度并不够高。有人认为要引入耳蜗之后的第 2 傅里叶变换神经机制，直到后来实验发现：耳蜗外毛细胞有与电致伸缩效应相关的主动机制，它能反向调节基底膜的振动响应。临床实验数据表明，当外毛细胞受损后，增加音量使内毛细胞重振，这一音量的临界指标为 40dB，即正常外毛细胞所带来的主动强度增益达 4 个数量级。加入主动机制修正后，神经频率调制曲线就可以与基底膜的频率调制曲线精确拟合。其次，在耦合时间上，本文注意到行波理论的另一个局限，就是关于同源音成分的同时性问题：按行波模型分析的数据，基底膜的特定驻波振动激励需要一定的时滞，实验测得不同频率的稳态信号，其转换为神经信号的过程有约 5~6ms 的“行波时延”[6]。以此推论，同时进入的同源音成分因频率不同而不能同时实现转换，这将出现同源音的同时性悖论。

对此，主动机制应是保证同源复合音的同时性的前提之一。通过触发耳蜗主动机制，基底膜的振动能迅速与当前驻波场适配，从而大大缩短了基底膜的实际响应时间。研究表明，有不到 10%内毛细胞传入神经沿底回穿越 Corti 隧道并投射到外毛细胞(称为隧道底部纤维)，与多达 10 个外毛细胞建立突触联系[7] 。内、外毛细胞联动的主动机制有可能主要借助这一神经环路运作。内外毛联动的主动机制是一关键的负阻尼效应，其存在解决了同源复合波同时性问题，确保了同源音各频率部位的感音内毛细胞是同时开始接受激励的。

基底膜、盖膜以及两膜中间的内、外毛细胞共同构成为 Corti 器，其中内毛细胞是感音换能过程的实施主体。内毛细胞底部排列着许多细长的、表面附着或游离有许多囊泡的丝带状结构，每条传入神经纤维都会与这种结构共同构成一个带状突触。这种特殊的突触结构使传入神经信号具有以下独特的特点：

首先，特定传入神经放发率会因声强而变，特定声强下特定部位的传入神经有相对优势的发放。随着声强等级的增加以及该部位基底膜振动响应程度的增加，该部分不同自发率的传入神经依次进入发放、稳定和饱和，由低到高响应约 4 个数量的不同声强变化，各种声音参量的感知也会随之有一定的差异。因此，随着声强的变化，特定频率神经的频率调制曲线将会有一定的差异，并导致如音高等相关感官变量或多或少地随声强变化而相应改变。

其次，存在实时瞬态强度响应和精准的时相同步关系。内毛细胞内精细的发放动力学过程，通过神经发放中的超极化过程有效清除了不合理或不确切的兴奋性触突后电位，更能反映精准反映声场的相关时相特征，进一步促进传入神经的锁相精度，确保了传入神经发放在时相上的精准特征[8]。从另一角度，由于每一最终导致域上神经发放的动作电位均来自同部位、同时刻的波前激励，可推论：同一时刻到达耳蜗各处的同源音波前有同样的瞬时声强，导致在响应范围内的传入神经发放应有相同或极为接近的动作电位潜伏期和锁相反应。即同源音泛音列对应的不同特征频率的神经动作电位，无论在起始的触发时间和周期内的锁相同步，应该都是精确准确的。

最后，传入神经发放有两类差异很大但又密切关联的表征：首发放动作电位，发放率时变规律。前者属“瞬态原始基准参量”，强调变化量，对应于波前的瞬态阈下变量的积分累积过程，不易受神经调制信号的影响；后者属“时变调整比对参数”，强调累积量，对应于进入同一通道的各时变规律间的线性叠加，易受外来神经调制信号的影响。

瞬变原始基准参量主要响应原始声场每一时刻波前受迫振动的结果。它的其中一个关键特征是：原始声场的密切相关的、极易互变的物理变量，将按时空特征分别量子化地切分并转换到不同神经中进行传送。由于神经个体之间在很大程度上是完全分立，这使得离散在不同神经中的关联成分间联系不再方便、更不易变，相对也更容易保持。这些信号在适当的神经功能区经过一定的众数整合处理，可提取如音高等相关瞬变基础属性。

时变调整比对参数主要来源于同一神经通道中的发放率时变情况。其原始信号不仅有波前振动信息，也有所有未消失振动的时变影响结果。更重要的是，这种表征的数字信号处理方式与原始的物理变量处理之间没太多区别，有可能高效地从发放率时变规律提取大量的关联信息。时间-频率(简称“时间说”)就是认为：音高识别是将神经元发放率的周期性间隔频率转化到特定神经，投射到频率筑构中特定位置，不过，至今为止，并未发现时间-频率转换所需要的相关的自适应函数神经实现机制。

3. 听觉皮层的整合频选模型

听觉皮层是音高识别的关键部位。经过耳蜗的傅里叶变换，声音的物理信号转换为量子化的神经信号，被分时分配到一系列相互隔离的神经通道中，经听觉神经通路传入皮层中的对应区域，进行进一步的音高分析。这些神经通道打破了原物理量间紧密且易变的关联，形成了分时、分频、分量的一系列神经发放组合信号。

在神经通道内，发放率是唯一变量，信号处理仍可基于时域。起始时它是特定频率的物理振动能量的映射，瞬态特征反映了新波前引致的瞬变规律，总发放量可对应各种瞬态信号发放量线性叠加。随着上行通道中神经间的相互影响，不同通道内的信号会按需变换出新的时变规律，提取声音中各种可能有意义的附加携带信息，如语音共振峰、特殊调幅调频信号、特殊含义的时程变化等。如在上橄榄外侧核(LSO)神经元中，用发放数表示的最佳频率(BF)只在 46.2%神经元中与特征频率(CF)一致[9]。这说明神经通道内发放率的变化频率是随后续神经处理需求会不断发生变化，不再保持初始瞬变频率，无法重现原始信号音的频率关系，故它已不适合用以提取同源复合音的音高。

针对这些特点，本文提出了一种融部位说与时间说为一体的部位-时间(Spatial-Temporal)音高识别模型。它不再寻求直接通过神经发放率提取决定音高的频率，而是根据发放率时变规律的约束，通过音源分离中心的整合神经元在特定时刻下从众多神经信号中找出相关联的一组神经发放，经音高推断中心的频选神经元的处理，完成统一的音高识别。因此，音源分离中心的整合神经元和音高推断中心的频选神经元，将在音高识别中扮演重要的角色。

3.1. 音源分离与整合神经元

除纯音外，音高的识别都必须依赖于一组关联神经发放。当神经通道传来信号，听皮层的神经首先要根据音源线索，即这组信号的频率关联关系、时相同步关系和能量分布关系，提取同源信号组合，再传到频选神经元提取音高。其中，音源分离就是从同时发放的众多神经信号中找出同源神经信号组合的过程。就音高识别而言，“倍频组合”和“时相组合”是音源分离的两大核心判据。按听觉心理感受，同音高的非同源音，多会听到两个独立复合音；而当两者初相很接近时，也可感知为一个新的独立复合音；但一组符合泛音列频率组合规则的独立纯音，若各个频率初相各异，则很难被感知为一个新的独立复合音。由于声音物理相位信息已由耳蜗转为神经发放同时性表征，故对于同时刻出现的，特征频率有倍频关系的神经发放，整合神经元将其视为同组神经发放予以整合，即以神经发放的时相同时性作为高效音源分离的线索。

从“倍频组合”判据来看，整合神经元因涉及不同频率神经，可表现为多峰型频率调制曲线，这提示这些整合神经元接收多个不同频率的神经信息。从清醒猴 A1 区发现，大量神经具有的频率调制曲线的峰型，随声强的不同，会出现单峰型和多峰型频率调制曲线的互换，如图 1 所示。这种多峰神经元的比例高达 20%[10]。而在麻醉动物的频率调制曲线

中，多峰型约占 10%[11]。这一方面说明麻醉状态对生理数据的影响；另一方面也表明，某些麻醉状态下的单峰型神经元，其实就是清醒状态下的多峰型神经元。

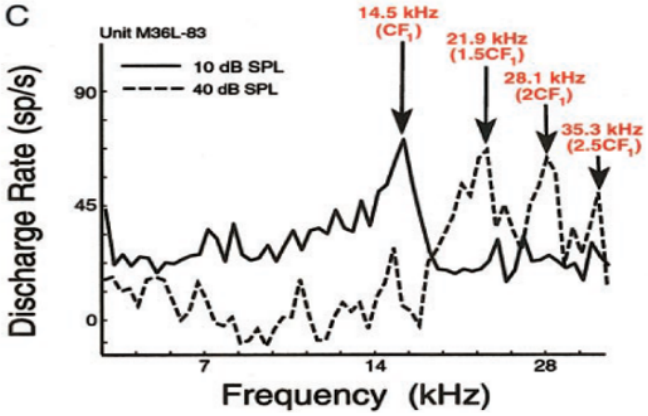


图 1、同一神经元不同声强下的频率调制曲线图 摘自(Kadia and Wang, 2003) [10]

同一神经元，在麻醉和清醒情况下，或在不同声强下，可分别表现单峰型和多峰型两种形式，故不能简单以调制曲线的峰型分类神经元。另外，具有单峰型频率调制曲线的神经元在双音激励下，却表现为多峰型，且多峰响应频点多表现出明显的远端抑制[10]。综上结果，至少有兴奋型和抑制型两类整合神经元，分别为音高识别提供不同的整合功能。

值得注意是，实验数据中的多峰响应频点，除部分抑制型整合神经元外，均按特定的倍频间隔规律离散分布，无论表现为特征峰、易化峰和抑制谷，其相对位置不变。如上图中，14.5kHz 随强度变化或峰、或谷、或强、或弱，变化明显。若以曲线中的最低峰值频率为基准，数据呈倍数和半倍数关系，如：0.5，1，1.5，2，2.5 等。由于同源复合音的泛音列倍频组合特点是：相邻泛音间的频差恰是同源复合音的基频，因此，整倍频的基准频率就对应基频，半倍频的基准频率则对应第一泛音频率，即基频的第一倍频。由此也可判断相关活跃频点是否同属于同源音。例如：当曲线中仅出现 0.5 活跃频点，而其它均是整数倍频关系，则这组整数倍频的频点属同一复合音，而这个单独的 0.5 活跃频点应属比此复合音至少低八度的另一个复合音的频率成分，应予以抑制减弱；而当曲线中同时出现有 0.5、1.5 等整倍频和半倍频活跃频点，则可认为此 0.5 频率点恰是这组频率组合的基频，故应予以易化加强，加强的程度取决于这组频率组合的时相关系。

倍频间隔规律是听觉中枢频率筑构的反映[12]，它对应于整合神经元突触的等距性，通过初级听皮层内神经的远距水平兴奋型联系，以及下丘内膝状体与初级听皮质神经环路的抑制型输入，形成了音高的整合神经功能区，以此分离、整合同源复合音。

从“时相组合”判据来看，兴奋型整合神经元最引人注目的重要特征是[11]：多数情况下，所观察到的易化响应强度取决于双音的重叠发放时间，当双音接近同步发放时，所获得的易化效应最强。有理由相信：在同源泛音列组合同步刺激下，相关的非线性易化效应会更大。这组频率的音互为易化本底信号，且并不需要整个完整的频率组合。只要存在此频组合中的另一个频率成分，即使其低于发放阈值，也可引起相关易化效应。这意味着仅有极少泛音列成分，就可出现稳定可靠的分组整合信号。由于同源复合音各成分的对应的传入神经是同时相发放，因此，这种易化效应将导致同源复合音相关成分的非线性增强。要注意同源复合音整合处理着重于迸发双音的瞬时同步关系，曾有实验给出相反的结论[13]，这主要是因为双音不同时，实为前音的时变结果与后音的相互作用，故不属同源复合

音识别所考虑的情况。

3.2. 音高推断与频选神经元

整合神经元的生理组构是“多对多”或“一对多”，并不能从中选出特定的单一频率。因此，听觉中枢必存在一类“多对一”神经元来进行频率筛选，以触发后续表征音高的神经元。这类神经元可统称为频选神经元。含有大量这类频选神经元的区域已经由脑神经影像学研究初步定位[14]，同时神经生理学实验也直接找出并验证了此区域位于与初级听皮层相邻的皮层（侧前区）[15]。神经通道的信号经整合神经元进行音源分离后，频选神经元将同源音的各频率成分重新“聚合”成一组，并以 “筛选”出这组信号的频差值作为基频，实现对声音音高的提取。因此，频选神经元是具有多峰型频率调制曲线的神经元。

这种“多对一”的频选神经元的传入神经突触数目是不可能无限增多，与整合神经元一样，频选神经元也应该是由有限个等距突触作为传入通道。实验表明[16]，不论基频如何，如果一组谐波只含有 20 次以上的成分，它就没有清楚的音调，这意味着一个整合神经元及频选神经元都应少于 20 个等距突触。进一步的实验还表明[17]，音高识别仅依赖于几个低阶的谐波成分，特别是 3 到 4 个最靠近 600Hz 的谐波成分起主导作用。只有当基频超过 600Hz 时，它才构成对音调有最重要的影响；当基频低于 600Hz 时，是一个更高的谐波成分对音高影响最重要；当基频为 100Hz 和 200Hz 时，其成分本身对音高几乎无影响，即便将它拿掉，也不会改变音高。这说明来源于耳蜗易感区的神经发放更能影响同源复合音的整合。在易感区内的神经发放更多地体现瞬态首发放动作电位的影响，整合频选也更多地基于信号的线性叠加，这种生理处理组构简洁合理，且在含大量同源复合音环境刺激下，因神经可塑性不断强化和固化，成为同源复合音整合频选过程的核心基础。

从时间-频率统一的音高识别模型来看，频选模板着重于提供可关联的有限个数的神经通道，神经通道内的所有信号（包括噪音）均会对结果有影响。至于神经通道内的不同频率发放信号混合结果，则主要取决于信号间的时变关系，时程因素也将影响最终复合信号发放的稳定性和显著度。在低频神经通道内的信号，由于其中的发放率变化周期长于皮层对瞬时信号的判断时长，则此通道中的信号将会被分成若干个更短瞬时信号，分别与其它相关神经通道内的信号整合。与简单的瞬时叠加不同，叠加信号将表现出更强的周期性特点，并会影响到音高识别的显著性。故对低频音高的提取，信号的周期性对音高识别的影响将会增大。

通过音高辨识行为实验，发现狨猴识别音高分别利用了两种频率线索：用包络线频率线索提取由高阶泛音组成的低音，用频谱线索提取由低阶泛音组成的高音。这与早期的心理物理实验提出的结论一致[18]。这意味着影响皮层音高识别的主要神经通道来自于耳蜗的瞬态敏感区。这里，如何提取包络线频率机制是一个较有争议的问题点。有观点认为，因耳蜗无法准确响应高频信号，故包络线频率应另从已传入的神经信号中，按“时间说”的机制提取；而我们认为：包络线频率的提取是由耳蜗实现的，不需要引入“时间说”。理由是：声场波前驱动下的耳蜗驻波场是一个连续统，其时域上和频域上的描述完全等价的。例如：用一个 100Hz 的低频音调制白噪音，此驻波场的频谱分布中必会存在这个 100Hz 的低频分量，相应位置的传入神经也必有信号。而经过耳蜗傅里叶变换后的信号，各自独立神经通道中的信号不再是复合信号，若不经通道间的关联处理，将不再具有提取包络线频率的前提条件。因此，包络线频率的提取不是引入“时间说”的证据之一。

4. 机器听觉仿生探究

机器听觉是对大脑听觉神经机理的仿生。仿生研究不但可以检测或实证相关功能区的工作机制，有助于逐步揭开大脑神经的奥秘；而且可以尝试替代听觉中枢某些功能。这将带来巨大的经济效益和社会效益，也正是未来信息技术发展的热点。

最引人注目的机器听觉当属是语音识别与音乐识别。语音识别主要依靠各神经通道内发放率变化特征提取信息，即着重提取包含在信号包络线中的语音信息。尽管语音识别起步不算早，详细机理研究也不多，但其相关应用却很成功。特别是 30 年前隐马尔可夫模型和近几年来的深度学习算法的引入，带来了两次重大的技术应用突破。音高识别是音乐识别的核心，按本文的音高识别机理，它是关联倍频神经通道内同时性信号整合的结果，虽然音高识别起步较早，详细机理研究相对多，但在应用技术上却始终没能实现的突破。

语音识别与音高识别在相同神经层级及相似的神经功能核团中实现，有密切关联。能否通过研究比对相关两者的异同，帮助提高音乐识别相关的仿生技术，这是无论从理论和技术上来看，都是一个值得探讨的方向。从实际需求出发，我们将讨论音高识别的两个应用问题：人工耳蜗技术改进和计算机听觉仿生实现。

4.1. 电子耳蜗音乐识别探究

过去 30 年，数十万重度耳聋患者借人工耳蜗技术重新获得语言交流能力，安静环境下对无语调语言的语音识别率可接近 90%。可是，相同环境下，其音乐识别率却仅 50%左右，仅略优于随机选择的比率。超低识别率，且时常夹杂着莫名其妙的高音，使电子耳蜗基本失去了音乐识别的功能。

人工耳蜗技术目前是采用电子耳蜗，是目前为数不多能直接影响细胞动作电位的侵入式脑机接口。它通过植入微电极，以电刺激方式将电子接收器转换的分频信号传入听神经。参照耳蜗的频率-位置对应，基底膜上的电极并非均匀分布：蜗顶植入电极少，在主要感音区电极间隔约 1mm 左右。不过，这 1mm 的长度范围内，已经存在有过百个毛细胞，并对应着数千个传入神经。也就是说，一个电极信号其实至少传入到数以千计的神经通道中。同时，因电流向神经组织周围非定向扩散，扩散范围随着刺激电流强度变化，从而触发非目标作用区域的听神经产生神经冲动，电极信号的影响范围实际上将更加扩大。这客观上限制了增加电极的数目的可行性。即使按现有电极间隔，当相邻电极同时发出刺激时，仍会导致严重互相干扰，为此不得不通过分时触发技术，同时触发关联信号也可能被人为分割为先后不同时间的触发信号。因此，电子耳蜗暂无法准确仿真耳蜗有的频率-位置对应关系，即特定频率信号并不与特定神经通道有严格对应关系。同时，它也无法保证关联神经信号的同时性，导致同源信号被分时触发，原来的一个音变成了多个音或其他音高的音，使音高识别无法正常进行。

人工耳蜗的这种信号特点，对于语音识别却影响不大。虽然人工触发的神经信号缺乏与特定频率关联的神经通道对应，但其中发放率时变信息仍然相当丰富，语音识别仅从检测神经通道内发放率的变化提取语音信息。但是对于音高识别，按本文的观点，并不存在以单神经通道信号变化来识别音高机制，故没有通过算法改进提高音高识别的准确率的可能。通过声音嵌合体的心理感知实验研究结果表明：语音识别能力主要取决于传入神经的信号的共振峰的时变规律包络线(Temporal Envelope，ENV)，而音高则取决于其时变规律的精细结构(Temporal Fine Structure，TFS)。对于正常人群，加强信号的精细结构能提高音

高识别能力。但是，在新型电子耳蜗 MED-EL 加入了精细结构编码策略后，电子耳蜗植入者的音高识别能力并没得到所预期的提高[19]。我们认为其可能原因是：加强信号的精细结构，对正常人群它提供了更多与特定音高识别相关的神经通道信号，而对电子耳蜗植入者它仅能增强了非特定神经通道内信号，却无法增加音高识别所需要的基础神经通道。可见，加强精细结构也无法改善人工耳蜗植入者的音高识别能力。

因此我们提出解决电子耳蜗植入者音高识别障碍的核心在于：确保同组倍频关联频率信号，同时通过与特征频率相符的神经通道传至听觉皮层，激活整合频选神经元。因此，改进脑机接口是提高耳蜗植入者音乐识别率的关键。在各种改进方案中，激光人工耳蜗方案尽管刚刚起步，或因其准确度高、保真度好、多余刺激少等优点，可能会有更大的发展空间，成为更成熟的耳蜗脑机接口[20]。

4.2. 计算机的听觉仿生探究

对于计算机听觉仿生，不需要受到电子耳蜗的种种限制，可独立实现大脑听皮质中音高分析的相关分析机理。各种听觉仿真应用不但可帮助探究或实证神经生理机制，其本身有广阔的应用前景，对未来的工程技术发展将有很大影响。

在音高识别中，傅里叶变换是信号处理的基础。经过长久的生物进化，耳蜗具有高效、低耗和实时的傅里叶变换功能。目前，唯一可用傅里叶变换功能仿真是通过计算机数字矩阵运算间接实现[21]。综合考虑数模转换及计算模型的相关限制，计算机仅能做到有失真的短时傅里叶变换，而不能像耳蜗一样实现无失真的实时傅里叶变换。引入适当的窗函数可一定程度上减少因截断效应所致的频率侧漏失真，不过，过短的变换时长不但加大失真，还会增加计算量。因此，实际应用中，根据不同的实际需求，要针对性对计算机仿真算法进行适当的优化处理 [22]。本文仅实现一个单音源音乐识别程序，以实证在本文提出的音高识别机理中关于同源复合音频选策略的仿生技术。

单音源音乐识别实际就是对按时间序列排布的同源复合单音的音高识别问题。这本质上就是一个对同源复合音相关成分频选策略问题，多数程序采用加权谐波峰值法提取优选基频作为音高。但由于对音高识别机理的理解常不到位，导致算法的效果并不好。如图 2 所示，图中是一段在安静环境下录制的小提琴独奏音频（贝多芬 F 大调小提琴浪漫曲），

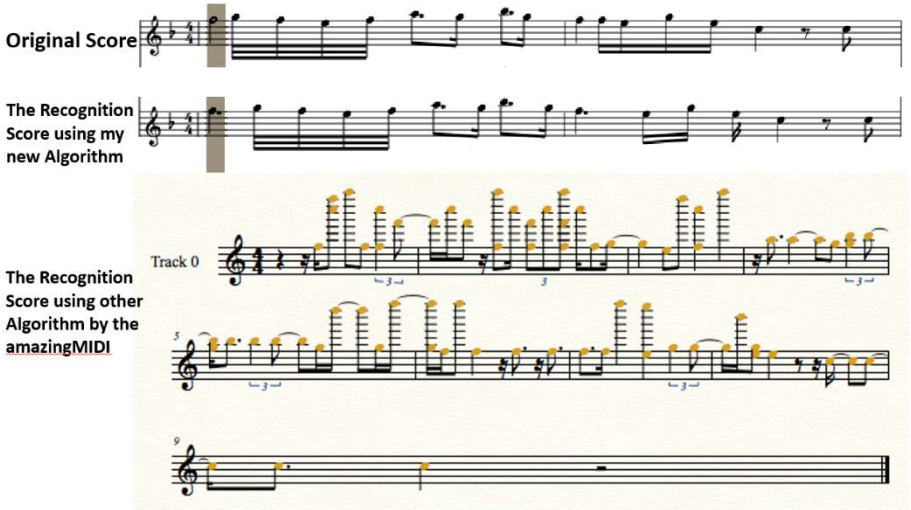


图 2、自编转谱软件与商业转谱软件（amazingMIDI）识别效果比较图

使用了市面上识别效果较好的音乐识别软件(amazingMIDI)。然而，该软件的识别准确率非常低，若直接查看所有这些用五线谱标记出音符的识别转换结果，现有算法识别出的结果

明显与原谱相去甚远，很难正常识别它所表示的旋律。此软件由于增加许多误判的高音谐波，导致完全无法形成可正常识别的音乐标记，难有准确率可言。另外，从图中也不难看出，听觉中枢不存在处理这类人为失误的机制，类似的对高阶泛音成分的误判，也应会导致电子耳蜗植入者听到的那些令人不快、机械的、难于理解的声音[23]。

同时，相对于原乐谱，识别的结果会额外识别出许多音高不定、时长很短的高频音。我们通过实验鉴别发现：所增加的高音恰好都是基频的倍频音，即程序将高阶泛音误判成了基音。与生物识别不同，计算机前处理并无考虑分辩声场波前振动，故采样结果反映是声场的时变规律。这样，泛音列频率分量受各自的相位因素影响，其瞬时能量分布随时间变化，并不等于音源的初始能量分布或平均能量分布。因此，由于采样频率高于正常音频的频率 1 至 2 个数量级，故当谐波与基频初始能量相差不大（这恰是小提琴的音色特征）时，常会在连续多个分析段出现“某谐波振动处于波峰，而基频振动尚在波谷”的情况，从而引起计算机的误判：将同一音符识别了一串由高阶泛音组成的乐段。

这种误判对单音音乐不难纠正：既然泛音列各频率成分都是乐音本身的不可分割组成部分,泛音列所有能量的总和才是此乐音的总能量。只是高阶泛音成分相对贡献很小，即使在生物音高识别中也被忽略[16]。因此，对候选频率，汇总其几个低阶倍频能量作为加权判据，就能简单快速地识别音高，且具有很高的准确率和很强的鲁棒性。另外，在过往的许多相关项目对泛音列的定义上有些疏漏：由于在乐理中八度音才是同一个音，故他们将八度以外的其它泛音排除，仅考虑八度泛音成分，也会额外导致误差。针对这几项分析，此程序根据本文提出的音高识别机理，并针对计算机的傅里叶变换与耳蜗傅里叶变换功能的差异特点，重新调整了置信度权重，增加了针对性的修正处理，形成加权谐波峰值法，使调整后加权谐波峰值法的音高识别准确大幅提高。

作为对比，我们按上述原理对程序算法进行了相关修正，并用 Objective C 在苹果电脑 OS X 系统上编写了一个“音乐转五线谱”软件的音高识别实证程序。本程序将单音音频以 44.1kHz 的采样频率数模转换到计算机，并以此频率倒数 0.02 毫秒作为加汉明窗的短时傅里叶变换(FFTW) 的转换时长，将单音音频转化到频域，再用上述加权谐波峰值法取此时段谐波能量分布的峰值频率作为待识别的音高。如图 3 所示，用改进后的加权谐波峰值法对

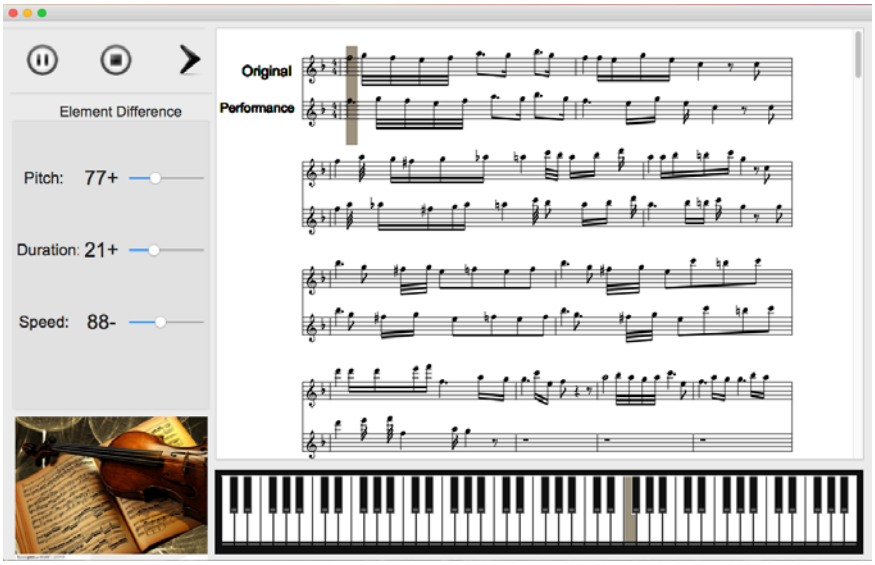


图 3、自编“音乐转五线谱”软件（浪漫曲，FFTW， Using Objective C on OS X）

上述同一段音频进行了识别，如果从五线谱标识的准确率来看，本程序已经能以可理解的

音乐标记忠实地还原了实际奏的音频，准确率接近 100%，程序已能转换出完全符合音乐规范的基本正确的乐谱。

综上分析，本文改进算法在音高识别上的显著效果也在一定程度上间接实证了：音高识别确实源于同源单音的低阶泛音列信号的整合，但信号相位会对结果有很大影响。

5. 研究总结

本文从机器听觉仿生技术改进的实际需求出发，综合了听觉神经生理研究和计算机听觉仿真研究的成果，对听觉神经生理和计算机仿真技术都做了相关探究：在听觉神经理论上，系统地分析了与音高识别相关的神经生理细节，并提出了音高识别神经机理的新观点；在听觉仿生实践上，依据这一音高识别神经机理，具体地分析了与机器听觉相关的仿生应用问题，并实现了优化计算机音高识别的新算法。

作为跨学科项目的研究，分析对比不同学科对同一环节的研究异同，更易找到解决核心问题的恰当切入点。本文正是通过对比物理声学、神经生理和计算机模拟技术三者的关联，从较少听觉神经生理研究对象的同源复合音入手，找到研究切入点：同源复合音神经信号的同时性；并陆续发现了耳蜗驻波声场、波前相位、生物锁相、外毛主动机制、内毛感音细节、初级听皮层音源分离整合和次级听皮层的音高频选提取等环节的内在相互联系，初步理清了音高识别机理。

作为实验科学技术的探究，灵活应用各种逻辑推断、技术手段和实验数据，更能挖掘数据的内在价值，把握相关技术改进的方向。本文正是利用学术论文中与音高识别相关理论和实验数据，逻辑推断出合理的技术解决方案，特别是以计算机为桥梁，将生物智能与人工智能联合一起研究，从而不但间接支持了部分听觉生理的逻辑推论，同时也对机器听觉仿真实践提出了一些方向性的建议。

本文的具体工作及主要贡献如下：

一、通过对耳蜗结构和功能的相关分析，本文明确了声场的波前振动、神经的首发放电位以及神经发放率时变规律三者之间的内在关联关系，提出它们在神经信号表征中的不同作用。本文还提出：同源复合音各成分在耳蜗不同部位所触发的传入神经冲动是同时同步锁相发放的，即同源复合音的相关神经信号具有严格精准的同时性。

二、通过总结相关实验结果，本文提出融合“时间说”与“部位说”的部位-时间(Spatial-Temporal)音高识别模型：在发放率时变规律的约束下，通过音源分离中心的整合神经元在特定时刻下从众多神经信号中找出相关联的一组神经发放，经音高推断中心的频选神经元的处理，完成统一的音高识别。我们认为，在听觉中枢中重新分离出可整合为同源复合声的神经信号，不但必须符合倍频关系的频率线索，还必须满足精准的同时关系的时间线索。

三、根据本文提出的统一音高识别模型，本文分析了现有电子耳蜗脑机接口以及计算机听觉仿真技术的限制，提出了相应的改进思路。受生物听觉音高识别机理的启发，本文还实现了一个计算机听觉仿真算法的改进，完整地编写了一个将音乐转换为五线谱的实用程序，达到了较满意的转换速度和识别准确率，也在一定程度上间接实证了本文提出的音高识别机理。

参考文献:

1. **Bengio Y., Lee DH., Bornschein J. and Lin Z.** Towards Biologically Plausible Deep Learning. Computer Science. *arXiv preprint arXiv:1502.04156, Mar 10, 2015*.
2. **Adam HM. Greg W., Konrad PK.** Towards an integration of deep learning and neuroscience *bioRxiv preprint first posted online Jun. 13, 2016; doi: http://dx.doi.org/10.1101/058545*.
3. **Frank S., Crawford Jr.** Berkeley Physics Course, Vol. 3 Waves *ISBN: 0070048606 1968 McGraw-Hill Book Company*
4. **Julyan HE. Cartwright, Diego LG., and Oreste P.** Nonlinear Dynamics, the Missing Fundamental, and Harmony *MCM 2007, CCIS 37, pp. 168–188, 2009.*
5. **Manley GA., Narins PM., Fay RR.** Experiments in comparative hearing: Georg von Békésy and beyond. *Hearing Research 2012, doi:10.1016/j.heares.2012.04.013*
6. **John MS., Picton TW.** Human auditory steady-state responses to amplitude-modulated tones: phase and latency measurements. *Hearing Research 141(2000, 57~79)*
7. **王坚，蒋涛，曾凡钢，等**《听觉科学概论》 *北京：中国科学技术出版社，2005.*
8. **Moser T, Beutner D.** Kinetics of exocytosis and endocytosis at the cochlear inner hair cell afferent synapse of the mouse J. *Proc Natl Acad Sci, 2000, 97, 883*.
9. **唐杰** 小鼠听觉神经元反应潜伏期对声信息的表征 *中国科学院生物物理研究所, 2006*
10. **Kadia SC, Wang X.** Spectral integration in A1 of awake primates: Neurons with single- and multipeaked tuning characteristics. 2003 *J. Neurophysiol. 89, 1603–1622.doi:10.1152/jn.00271.2001*
11. **Sutter ML.** Shapes and level tolerances of frequency tuning curves in primary auditory cortex quantitative measures and population codes. *J Neurophysiol* 84: 1012–25, 2000.
12. **Wang X.** The harmonic organization of auditory cortex. *Frontiers in Systems Neuroscience 7:114. 2013 CrossRef Medline*
13. **Brosch M, Schulz A, Scheich H.** Processing of sound sequences in macaque auditory cortex： response enhancement. *J Neurophysiol 82: 1542–1559, 1999.*
14. **Plack CJ, Barker D, Hall DA.** Pitch coding and pitch processing in the human brain. *Hearing Research 307 (2014) 53-64*
15. **Bendor D, Wang X.** The neuronal representation of pitch in primate auditory cortex. *Nature 436, 1161–1165.doi:10.1038/nature03867 2005*
16. **Ristma RJ**. Existence region of the tonal residue. *J Acoust Soc Am*, 1962. 34: p. 1224 - 1229.
17. **Dai H.** On the relative influence of individual harmonics on pitch judgment. *J Acoust Soc Am*, 2000 107(2): p. 953 – 959.
18. **Bendor D., Osmanski MS., Wang X.** Dual-Pitch Processing Mechanisms in Primate Auditory Cortex *The Journal of Neuroscience, November 14, 2012 • 32(46):16149 –16161 • 16149*
19. **Moon IJ., Hong SH.** What Is Temporal Fine Structure and Why Is It Important? *pISSN 2092-9862 / eISSN 2093-3797 http://dx.doi.org/10.7874/kja.2014.18.1.1*
20. **耿阳，叶青** 光学人工耳蜗的基础研究及展望*《国际耳鼻喉头颈外科杂志》2012 P.341*
21. **Bracewell, R.N.** The Fourier transform and its applications *(1986). (Vol. 31999). New York: McGraw-Hill*
22. **Camila A.S., Gael R., Benoit F.** Multipitch Estimation Using a PLCA-Based Model：Impact of Partial User Annotation. *IEEE International Conference on Acoustics, 2015*
23. **冯海泓，原猛，陈友元** 人工耳蜗植入者音乐感知研究 *声学技术 Vol.31, No.1, Feb., 2012. P53~P60*

简　历

团队成员：

李顺简历：

2011.07~2014.06：广东实验中学(初中部)　　1 班
2014.07~现在：　华南师范大学附属中学(高中部)　　首届大学先修实验班

指导老师：

唐杰简历：

1996.07~2000.06：华中师范大学生科院 生物化学专业 理学学士
2000.07~2003.06：华中师范大学生科院 动物学 硕士
2003.09~2006.05：中国科学院生物物理研究所　　生物物理学 博士
2006.05-2009.11：Research Associate, 美国华盛顿大学（ Washington University in St. Louis）生物系，美国科学院院士 Dr. Nobuo Suga 实验室
2009.12-2011.09：Research Fellow, 美国 Creighton 大学医学院，David He 毛细胞生物物理实验室
2011.09-2011.12：Research assistant professor， 美国 Creighton 大学医学院，生物医学科学系
2012.01-现在：　南方医科大学，基础医学院生理教研室， 教授，博士生导师

刘桦简历：

1996.07~2000.06：　华中师范大学生物化学专业　　理学学士
1997.09~1999.06：　华中师范大学计算机软件专业　理学双学士
2000.07~现在：　华南师范大学附属中学　　高中生物高级教师
2008.03~2008.08：　香港科技大学生物学部　　访问学者