



现代噪声科学系列讲座之三

(上)——有关的心理声学

王士谦

(中国科学院声学研究所)

1988年10月15日收到

在前面我们曾说过，噪音链的另一端为噪声的接受与感知，即噪声的听觉方面。在这一个专题中，我们要介绍关于听觉的三个方面。一、一些与噪声有关的心理量和物理量的关系；二、一些重要听觉理论；三、介绍(主要是近些年来的一些)一些噪声感知实验的研究结果。因为国内对于言语声的感知研究已有较多介绍，而声乐的感知研究的文献却罕见；又因为本噪声讲座主要对象是艺术噪声的教育、医学和工程界，我们将以介绍声乐的感知为主。关于听觉系统的解剖生理，即人耳对声接受的通道方面这里从略。

一、一些与噪声有关的心理量和物理量的关系

心理声学的基本内容是研究声音客观物理量与听觉主观心理量之间的关系。在本讲座中，噪声感知的重要问题是：我们听到的是什么？听到后的主观心理量与客观物理量之间对应关系是怎样的？其对应关系的心理物理原理是什么？后一个问题的深层认识还是个“黑盒子”，前二个问题一般是靠实验的办法来逐渐认识的。

迄今的研究结果表明：我们听觉系统对噪声(言语与声乐)感知的基本特征要素是音高(*pitch*)、响度(*loudness*) [在音乐中称为音量(*volume*)]、音长(*duration*)和音色(*timbre*)。各主观量都依赖一个以上的可测客观物理量。

应用声学

其中，音高对应的主要客观物理量是基频 (*fundamental frequency*)；响度对应的主要客观物理量是声强 (*intensity*)；音色对应的主要客观物理量是声谱和包络 (*spectrum & envelope*)。

1. 人耳听觉的动态范围和锐度

人的听觉系统对声音物理量的感知有一定的动态范围。人耳的有效感知频率范围落在大约 20Hz—20000Hz 之间。有用的强度范围随频率变化而变化，而人耳敏感的频率范围约在 1000Hz—6000Hz 之间。在此范围内，可听的纯音强度范围大约有 120dB 之差。在 500Hz 以下，8000Hz 以上耳的灵敏度下降得很快，尤其在低强度时更显著。噪声的使用声级一般低于 70dB。在听阈的 1000Hz 处鼓膜仅偏移 10^{-3} cm，(小于一个氢分子的直径)；在痛阈处鼓膜核心也只偏移 10^{-3} cm，而基底膜位移还比耳膜偏移小 10 倍，对时间而言，人耳不仅可进行长时间的分析，而且可分析小到 2ms 的声学量。

2. 声音要素的主客观关系

(1) 响度与强度：(介绍纯音情况，其结果在不少情况下对复音是近似正确的)

响度为主观心理量，对应的主要客观量为强度，后者可用声压级 (SPL) 表示，它以 dB 为单位，基准声压一般取 2×10^{-5} (pa)。前者的主观评估比后者的客观测量要复杂得多，注意，只在中频范围内 (100Hz—2500Hz) 响度才由声强决定；在其余频带，频率对响度的影响

· 39 ·

就越来越大。

在过去的几十年中出现了几种响度标度，但迄今还没有一种可被满意地通用于任何情况。常见的几种响度标度有：

(a) 响度级 (*phon* 标度)：它是一种以响度级代替 *dB* 标度值的一种物理-心理混合响度。把可调音强的 1000Hz 的纯音作为标准音比较，当把它调到与欲测音等响时，标准音的声压级数即为该音的响度级数，且把响度级的单位取名为“方” (*phon*)。据 ISO 规定， $I_0=10^{-16}\text{W}/\text{cm}^2$ 的声强相当于 1000Hz 标准频率时的零方响度级的强度。

不同频率的纯音在和 1000Hz 纯音等响时它们的声压级数是不同的，这些不同声级作为频率函数所形成的曲线称为等响曲线。由它可知，低(频)音比中、高(频)音衰减的大。

响度级的方标度是心理声学中用以把响度强弱进行排序加以比较的一种渐强标度，但它不能给出用方所表示的量之间的倍数关系。有一种广泛使用的响度数量标度，即

(b) 宋 (*son*) 标度：它以“宋”为标度单位。它规定频率为 1000Hz 且声级为 40dB 的纯音的响度为一宋。当一个声音为 N 宋时，即表示其响度为宋标准的 N 倍响度。图 6 表示一个纯音的响度(宋)和其响度级(方)之间的关系曲线，由坐标轴刻度即可知，它们是非线性函数关系。

(c) 在音乐中，常以 1—3 个 *f*(*forte*)、*p*(*piano*) 和 *mf*、*mp* 来表示声音相对的强弱关系(音乐中的标准声级是六级)。一般噪声声强的动态声级变化在 30dB—80dB 之间。两个相邻的强弱符号之间的响度 (*phon*) 差大致为一整数倍。

此外，还有感觉级 (*sensation level*) 等其他办法，在此就不叙述了。

(2) 音高和基频：

音高主要由基频而决定，但纯音的音高还受音强影响，在一定范围内，可用强度差补偿频率差。音高和基频的关系在音乐与噪声的研究中比前面介绍的响度和音强关系更为重要，且

它们有更精确的数字量化表示。噪声频段约为 50Hz—1046Hz。

音高有下列几种标度：

(a) (纯音)美 (*mel*) 标度：定义在 40 方响度下的 1000Hz 纯音的音高为 1000 美。其它纯音的美数则以与它比较听感的倍数而计值。人耳对音高的评价美数和倍频在 500Hz 以下是线性关系；但在其以上，前者的值较后者增长为慢，因而在高音处，倍频还不够形成倍音高，例如，对应 1600Hz 纯音的美(音高)标度只是 2400 美，心理声学还经实验证明，对同一个音，各人的音高感知结果可能是不同的；音高与频率并无一对一的对应关系。这种心理声学美标度在音乐实践中并无多大用处。

(b) 音乐标度：在音乐中，以八度音程来表示音高标度，且八度内有全音程和半音程。自 C4 (即中央 C) 开始[现国际标准规定 A4 (即中央 A) 对应为 440Hz 还有其它标准]，每八度音程构成一组，对应键盘排成组序(国际上对此有不同的编序法)。音乐中，在 500Hz 以下，八度音程对应倍频，两者成线性关系；但在其以上，倍频对应的音程就逐渐地小于八度(虽在 4000Hz 以下仍颇近似) Feldkeller 和 Zwicker 曾把音乐的中央组命为 +1 组八度，其上各八度依次为相对应的正整数为序的八度；其下各八度依次为相对应的负整数为序八度。他们把这样的音乐八度(心理标度，用 H_m 表示)与谐波倍频(物理标度 H_b) 进行对比，(见图 7，以其上、右坐标轴为参照)。从中可看出，音乐八度对应的频差随频率倍数增加而不断增加。在 500Hz 以上(即大约在中央组以上)，八度对应的两个频率 f_1 和 f_2 就有 $(f_2/f_1) > 2$ ；在中央组以上的四个八度音程对应的是约七个倍频程。所以，许多乐器的中央组以上的音都必须靠调音师的音高感觉来校准，而不能单以倍频决定音准。注意，如上所述，(心理量的)八度和(物理量的)倍频不仅是概念不一，而且对应关系(除低音一小段外)并非线性，对于 4000Hz 以上，它们远离线性关系。不少文献把八度与倍频的概念和关系视为同一，

且使用同一个英文字 (*octave*), 这易形成混乱。在一些音乐或声学文献中用基频数来表示音高, 其实这种表示对中央组以上的音是不准确的。音乐中的音程关系是由训练而建立的一种条件化的音高感知 (*conditioned pitch perception*), 尤其是对于 4000Hz 以上的声音, 只有经过特殊训练的人才能判准这样的条件化音高。这种用旋律音程训练而成的条件化音高与用美单位对一般人测试的音高感知的结果也是不同的。

音乐上大多使用的是相对音高听觉, 还有一种绝对音高听觉, 它指的是在脑中形成的音乐音高与声音频率的绝对关系, 即不需要参照音而能识别音高的听觉。但它并非音乐家必须具有的素质。据有人统计, 仅占 0.01% 的人口能有绝对音高听觉。

八度内的全音程和半音程所对应的频率关系视律制不同而有所不同。通行的十二平均律可以 $f_2 = 2 \times f_1$ 来定其第 n 个半音音程的频率。音程也可用音分来表示:

1200 音分 = 12 半音 = 倍频。

关于复音音高, (依赖于声谱和时长) 有下列几种情况:

(a) 一般地说, 复音中频率最低的纯音分量为该复音的基音, 且其频率对应于该复音的音高。

(b) 对某个窄带内的若干纯音集合体来说, 其音高总是倾向对应于该窄带的频率中心。

(c) 复音由若干等差的不同频率分量组成时, 赫姆霍兹和一些其它人 (如 Sorge, Tartini, Young 等) 认为, 感觉音高通常等于以这个等差频率作为一个纯音所对应的音高。当基音分量不存在 (*missing fundamental*) 时, 所听到就是这个等差频率的音高。这是因为耳朵是非线性系统, 可听到差音的结果。按照他的观点, 人耳可感知到分量的和音及差音。例如, 两个分音 f_1, f_2 , 其和、差音 $|mf_1 \pm nf_2|$ (其中 m, n 为整数) 就对应着听觉感知中的非线性畸变的音高。赫姆霍兹的解释大体与实验结果相符, 但不完全正确。关于这一点, 我们将在介绍

Schouten 的残余理论时再予补充解释。此外, 音强的变化对复音的音高影响很小, 噪声和音乐声差不多都是复音, 所以音强的变化不致改变其音高。

(3) 音色与频谱:

音色的定义是, 听觉区别两个具有相同音高和响度的声音的特性。它主要由频谱决定, 但也与包络、波形, 尤其是瞬态特征 (音形) 有关。噪音音色包括语音音色和音乐音色。前者指的是区别各种语音音位的感知特征, 如元音的前、后、高、低等 (通常与低共振峰较多相关); 后者指的是区别各种乐音“音位”的感知特征, 如元音的明、暗、开、闭等 (通常与高共振峰较多相关)。

二、听觉理论

听觉理论是根据听觉器官与神经系统的解剖生理, 来解释声信号的感知过程和原理的理论。由各种设想和验证就形成了各种听觉理论。它们解释的范围包括音高、响度、音色、差阈、掩蔽、双耳效应等一切听觉现象, 尤其集中在音高、响度等重要经典课题上。音高在噪声及音乐中特别重要, 因此这里我们将着重介绍几个关于周期信号的音高感知的听觉理论, 并解释一些响度的确定等问题。过去心理声学对简单声 (纯音、脉冲噪声或短声) 的音高研究已有许多结果; 但涉及到复合的自然声, 特别是真实的或合成的言语声时, 对其听觉性质, 即便在分析的水平上也是较不清楚的。心理声学对音高的研究现正从以往一维纯音分析趋向多维复合声的分析, 实验目标也从听觉的感受性转向选择性和分辨力。传统上, 音高感知听觉理论是根据生理实验结果来解释声音的物理频率与心理音高的关系和原理的理论。它们包括位置理论 (即频率理论) 和周期理论 (即时间理论) 两大类; 前者主要有共鸣论和行波论, 它们立足于频域讨论; 后者主要有电话论、齐射论等, 它们立足于时域讨论。人耳似乎具有时、频两域分析计算并做判断的能力。迄今看来, 共鸣论和电话论存在较多的问题; 行波论、齐射论

各能合理解释许多关于音高感知的方面，它们虽也都还有解释不到之处，但尚没发现大的缺欠。

1. 位置理论 (*place theory*)

也称频率理论 (*frequency theory*)。它是一类以频率对应基底膜一定位置为线索解释音高的听觉理论。和位置理论有关的主要人物先后有 *Pythagorus*, *Galileo*, *Seebach*, *Fletcher*, 特别是 *Ohm* 和 *Helmholts* 等人。

(1) 共振论 (*resonance theory*):

十九世纪初，欧姆 (*K. Ohm*) 根据傅立叶 (*J. B. Fourier*) 分析提出欧姆声学定律 (也称欧姆第二定律) 的假设。他的基本论点是：当我们听一含有许多纯音成分的复合声时，听觉机构就把这种声音分析成它的各个频率分量，因此，我们就能分别感知每一个纯音。一般我们听声音时并不意识到这种情况，而经过训练的听觉，在一定程度上，可以把复合声分解成各个分音。他在批评 *Seebach* 的关于虚基频 (*missing fundamental*) 感知的实验解释 (即，是周期而不是基频决定音高感知) 时又认为，对应于一定频率的音高仅当声波在该频率处含有能量时才能被感知。这样，他就难以解释虚音高 (*vertual pitch*) 的现象；后来，他把虚基频的声音所被感知的 (虚) 音高视为声幻觉 (*acoustic illusion*)。

十九世纪后半叶，亥姆霍兹 (*H. V. Helmholtz*) 根据由显微镜所观察到的对内耳结构的认识支持欧姆的假设，他并认为：基底膜由大量 (2400 根) 神经纤维所组成，这种神经纤维，就象竖琴上调了音的弦 (*auditory string*) 一样，横向绷紧在耳蜗上。每一根神经纤维依其张力与质量而对某一特定频率发生调谐共振，且在基底膜上都有一定的固定位置。当镫骨踏板的运动导致耳蜗内的液体发生振动时，只有能与刺激中存在的频率发生共振的那些神经纤维才会振动起来。假设各条神经纤维分别把各个调谐成分传向大脑，那么我们感知的各纯音成分将同那些被激励的调谐神经纤维的共振频率相对应。

以上这些理论假说都属于听觉的共振论。后来实验证明亥姆霍兹接近正确，但不完全对。例如，基底膜上并不存在他所说的那种个体共振纤维，而是在基底膜上作为整体形成行波共振；另外，他用差音对虚音高的解释也与音高偏移实验的结果不相符合。

(2) 行波论 (*travelling wave theory*)

听觉行波理论是 *G. V. Bekesy* 根据对新尸体基底膜振动的直接实验观察和对基底膜机械特性的测量，在 1951 年所建立的听觉理论模型。他对内耳动力学的贡献于 1961 荣获诺贝尔奖金。听觉行波理论解释是：经卵圆窗而传进耳蜗的机械振动形成液压变化，从而在基底膜上产生行波。由于基底膜沿长度各处的纤维宽度及劲度的机械特性不同而形成行波的最大幅值，反映了声音频率与基底膜有关位置的关系。柯替氏器把基底膜所含的频率、幅值和时间信号，经盖膜与基底膜之间的剪切 (*shearing*) 运动及毛细管的弯曲形成神经兴奋，这样就把机械能转换成电化学脉冲，再经听觉神经传入大脑皮层颞叶的听觉神经中枢进行信号解码处理，从而形成声音的听觉。在对于镫骨踏板上来的纯音刺激，在基底膜上的振动不是局部的调谐，而是在整个基底膜上形成行波。它的振幅在从卵形窗向蜗孔移动的过程中不断变化。对于高频来说，最大振动靠近卵圆窗；对于低频来说，最大振幅靠近蜗孔。在频率低于 100Hz 时，最大振动幅总在基底膜的顶端。在一般情况下，行波在从卵圆窗到最大振幅处的移动过程中，振幅是逐渐增加的，在基底膜的一定部位振幅达其最大值，但一过了最大值，振幅就迅速地减小 (见图 8)，其最大振幅部位对应于一定的刺激频率。由于基底膜的运动，使得柯替氏器官内位于基底膜和盖膜之间的毛细血管由于剪切力而发生形状弯曲变化。这种形变就使有关神经细胞兴奋，从而把机械能转化成电能并在相应的神经纤维中产生电脉冲，传入中枢而引起音高感觉。*Bekesy* 提出，低频声音的感觉变化与神经冲动频率有关；而中、高频声音的感觉变化与行波引起基底膜最大振幅所在部

位有关。

任何一条神经纤维最容易作出反应的频率是柯替氏器官与这条神经纤维所联结部位上提供的最强反应的频率。对于其他频率的纯音来说，其强度必须更大才能使神经细胞放电。在听觉神经纤维中，频率响应的选择性并不是很大的。通常，一根神经纤维容易被低于其特性频率的纯音所激发，只要这个纯音的频率适当地高于阈限值就行了。但对于比它的特性频率稍高一些的纯音，就很难使神经纤维放电。在听觉通路较高层(如中央膝状体等)的那些纤维中，响应特性就不同。许多细胞对无论什么频率的纯音都不作出响应，但却对短声和噪声作出响应。在中央膝状体中或大脑皮层的听觉区域里，还有一些只对纯音作出响应的其它神经元。这些神经元似乎比听觉神经纤维和蜗神经核的细胞更具选择性。它们只对以其特性频率为中心的一条窄频带作出响应。目前的观点认为，纯音音高的感知在某种程度上依赖于在哪些神经纤维上脉冲传给大脑。这样看来，其过程应比这种简单的部位机理要更复杂。

另一方面，响度与每秒钟到达大脑听觉区域的脉冲总数有关。声音刺激越强，触发内耳并传向大脑的脉冲数量就越大。这个过程细节尚未搞清。不同的神经纤维有不同的阈值也许很重要。从大脑返回的反馈通路可以引起阈值的变化，这种情况取决于在一定的时间内存在的条件，以及一些更加复杂的因素。

近些年的实验也表明共振理论有许多局限。其中主要的一个是它不能解释人耳能对频率做精细分辨的功能；再一个是，它不能解释为何人耳能把复音感知为具有一个音高的整体音。研究已证明，耳蜗对频率的分辨率实际上远细于基底膜机械响应的结果。Evans 和 Wilson 等人假设有某种“第二滤波器”细化了基底膜的频率响应，从而使耳具有精细的频率分辨率。这种“第二滤波器”的生理本质尚未清楚，但它是位置理论的新进展。

2. 周期理论 (*periodicity theory*)

也称时间理论 (*time theory*)，涉及声感

应用声学

知过程中对声波进行的时间分析。它是一类主要从神经传导中时间分布特征为线索解释听觉感知过程的听觉理论。时间理论假定，听觉神经电脉冲的时间分布特征载有所听声波的时间分布信息，而对其信息解码是在中央神经系统中按自相关计算进行处理的。和时间理论有关的主要人物有 Huygens, Rutherford, 特别是 Schouten 以及 Wever 和 Bray。

(1) 电话论 (*telephone theory*):

这一理论是本世纪初由 W. Rutherford 提出，并在当时被越来越多的人接受。它认为神经系统中的传输在本质上是电传递；基底膜的工作机制与电话接收器的音膜相似。外来声波一振动，基底膜就全振动，把声波转换为频率、振幅和波形的神经冲动；神经如同是电话线，仅仅用来把形态不变的电信号传送到大脑。电话理论还认为音高取决于由声波频率所引起的神经冲动的频率；响度依赖于参与活动的神经纤维数量；音色差异则是以振动的不同模式为基础。这就是说，耳朵就象传声器将声波转换成电信号一样，只是简单地把声振动转化为电振动，而整个信息处理过程则都是在高级中枢神经系统中完成的。

电话理论看来有问题。神经元并不象电话线那样的方式传输信号；它也不能解释高频波的传导；再有，由这种理论体现出来的许多感知效果，与目前有效的心理声学测量并不相符。后来 E. G. Wever 和 C. W. Bray 在电话理论的基础上，把它修正成齐射论。

(2) 齐射论 (*volley theory*):

它是 E. G. Wever 和 C. W. Bray 1930 年在对猫听神经生物电进行研究后提出的一种听觉学说。由于神经传导有不应期，神经冲动频率有限，因而电话理论不能解释高频声波在神经纤维中的传导。齐射理论对它进行了修正。齐射理论假设神经纤维传导的相对不应期有长有短，它们可以根据相对不应期的长短分成“排”，各排可以轮流地接受刺激并发放神经冲动。例如分成 a、b、c、d、e、f 六排，a 排先反应，然后 b、c、d、e、f 各排相继反应。当 f 排反

应完毕, a 已过了不应期又可以接受刺激了. 因此单位时间内能传导的神经冲动数就可以大为增加. 而且图 9 又表明, 对刺激波如何产生脉冲齐射, 以及它如何同步于刺激纯音. 其一结果是, 每秒齐射数等于刺激纯音的频率. 这样齐射理论也就解决了高频声波的神经传导问题了.

Schouten 在他的虚基频的音高偏移 (*pitch shift of missing fundamental*) 实验中发现, 音高随载频的频率变化而偏移, 他并建立起残余理论 (*residue theory*). Schouten 对 1200 Hz 的载频音加以 200Hz 的调制, 这样得到它的两个边频为 1000Hz、1400Hz; 三者的复合的虚音高 (*virtual pitch*) 的确对应差音音频 200Hz. 可是当他改用 1240Hz 的音作为载频且再加以 200Hz 调制时, 虚音高却对应 207Hz, 而不是差音的音频 200Hz. 因此, 虚基频的复合音被感知的虚音高是由 $\{f_n/n\}$ 均值而定. Schouten 认为, 音高偏移现象是由于听觉神经的同步发射所致. 信号的低频分量可由基底膜

的局部激发而解决; 而对高频, 一些分量落在基底膜的给定域的共振带宽内, 这些未解决的分量形成一个保留了原包络周期的残余. 这个残余对神经的时间分布就提供了一个基本线索. 这样, Schouten 的残余理论不仅阐述了虚音高的感知原理, 还对亥姆霍兹的关于音高非线性畸变的假设提供了另一合理的补充解释. 近来的研究 (Goldstein) 认为, 周期信号的音高感知过程是: 首先由外围滤波机制对复合声进行分析并按一定的精度给出各分量的频率; 而其分析是以齐射的分析作为基础的. 然后听觉中枢把各谐波分量频率与其它成分频率进行均衡, 其最佳均衡量就是对应感知音高的频率.

从上看来, 音高感知中的频率分析和时间分析的线索都是存在的; 而且在不同的频带内两者的主次作用不同. 在低频带内, 以时间分析为主; 在高频带内, 以基底膜的频率分析为主. 关于两种线索的相对重要程度和各起主要作用的频带界限的问题, 虽已有些实验结果发表, 但现仍处在研究探索阶段.

(上接 47 页)

与局部声压和单位体积热的积累二者的乘积成比例. 这就产生了一个想法, 用适当的方式控制激光束使之随所产生的声脉冲移动, 使热的积累与声压的极大值合拍. 例如当声沿水面传播, 激光束就应以声速沿水面移动, 这样激光束产生的热量会使声波幅度不断变大. 当然声波幅度总是趋向一有限值, 因为声不断从水面向下衍射.

用稳定功率输出的激光束可以控制产生声波的频率. 方法是用光学手段把光束分成几束等间隔的平行光束, 当它们在水面运动时使每一点周期地加热, 不同点的加热周期相同, 因此产生的声波也有同样的周期.

光束移动的速度控制着向水中辐射的声束的方向, 正向超音速飞机的 Mach 数控制着入射声波的入射方向. 同时, 声束向下传播时的扩散可以用水面加热激光的分布尺寸来控制. 声波向下传播时声吸收时幅度减小的主要因素, 在被噪声淹没前, 信号是可检测的. 初步的计算表明, 应该可以用现有的激光技术建立这样的激光系统, 使它产生的主频为 30kHz 的声传播 5km—10km 到后仍能被探测到.

(张海澜 译自 *Physics Today* 1(1989), S-6.)