31.
[15] D. R. Bacon, IEEE Transaction on Sonics and Ultrasonics, SU-29 (1982), 18—25.
[16] A. J. Lovinger, Science, 220—4602 (1983), 1115—1121.
[17] H. R. Gallantree, The Marconi Review, First Quarter, (1982), 49—64.
[18] T. R. Gururaja, et al, *IEEE Trans.* **SU-32**-4, (1985), 481.
[19] T. R. Gururaja, et al, *IEEE Trans,* **SU-32**-4, (1985), 499.
[20] 栾桂冬，应用声学**7**-4，(1988)，37—41.
[21] 奥岛基良，日本音響學會講演論文集 2-7-7 (昭和7 5年)，717.
[22] P. Walmslay, *PROC. INST. Acoustics* Pt3-6, (1984), 38—41.

* 本文得到国家自然科学基金赞助

听觉基础研究的若干问题展望(2)

方　至
（中国科学院心理研究所）

1990 年 6 月 8 日收到

三、心理声学

当前心理声学研究的主要趋向是从以往纯音听觉的单因素研究转向具有频谱时间表征的复合声的多维研究. 这个转变和生理声学的发展情况恰成有趣的对比. 如上节所述，生理声学的近期进展是对 Helmholtz 为代表的传统耳蜗观念的突破. 与此相反，心理声学的前进却不能不回到 Helmholtz 在其名著“乐音感觉：音乐理论的生理学基础”一书中提出的一个基本课题，即对音乐、言语这类复合声的知觉. 自然，这并不是说心理声学应该迳直去研究音乐和言语，去取代音乐声学和言语声学.在纯音心理声学和音乐声学、言语声学之间还有一大片被忽略了的领域，即具有复杂的频谱和时间模式的复合声. 这类声音在物理结构上和音乐、言语有许多共性. 它们不但都是复合声，而且都具有多声源和连续声流的性质. 对这类声音的知觉，也和音乐及言语的知觉有着不少共性.它们都是多维的，而且加工过程都不限于听系统外周，还有听觉中枢的参与. 因此，复合声心理声学的研究将一改以往心理声学和音乐、言语听觉脱节的现象，有希望真正成为理解两者的基础.

下面将扼要介绍 50 年代以来对复合声某些主观属性的实验结果和三项可能产生影响的近期发展.

1. 复合声的某些主观属性

(1) 分音 (*partrals*) 的分辨限度 Helmholtz 曾研究过人耳将周期性声波分解为它的类正弦成分的能力，即欧姆听觉定律，但没有给出这种能力的限度. 用包含有 12 个谐波的复合声的听觉实验表明，即使在最有利的条件下，人耳能分辨的谐波数不会超过前面的 5—7 个.这和临界带宽的概念是一致的.

(2) 合音 (*Combination Tone*) 当频率分别为 f_1, f_2 的两个纯音同时作用，主观上将产生一个频率相当于 $2f_1 - f_2$ 的差音. Helmholtz 认为这是欧姆听觉定律的又一限度. 以往，对原声 f_1, f_2 需要多强的感觉级才能引起差音听觉，几乎没有什么实验数据. Plomp 的实验说明，差音 $2f_1 - f_2$ 的觉察阈主要决定于频差 $f_2 - f_1$. 因此，它肯定和人耳的频率选择性有关，而且来源于内耳的非线性.

(3) 涩度 (*Roughness*) 和不协和度 (*Dissonance*)

调幅纯音产生拍 (*Beats*) 时，拍的变化将引起刺耳的粗涩感. 当 f_1, f_2 两个纯音的频差逐渐增加，最初听到的拍象响度在作缓慢的起

伏，随后听到的是一串间断的脉冲，最后是明显的粗涩感。心理声学实验指出，涩度的变化受声级的影响不大，而和调幅度有关。在调幅度为 60—80Hz 之间达到最大值。Helmholtz 也是从纯音频差来讨论乐音的涩度和不协和度的。实验证明，频差为 1/4 临界带宽时，和弦听来最不协和，当频差大于临界带宽，和弦就变得协和了。

(4) 复合声的音高 (*pitch*)

音高知觉本是听觉理论中的核心问题，而复合声的音高使问题变得更为突出。音乐和言语这类复合声都是周期性的信号。它们的音高常和波形的重复率相对应。而且，当它们的基频成分被去掉时，音高仍维持不变，即所谓"去基频"现象。最初，人们以为这种音高是在时间域内确定的，是听觉频率分辨力不完善的结果，因此得到了"周期性音高"，"余音音高"这样的别名。经过 70 年代的探索，现在基本弄清，余音音高仍是在频率域内由一种特征提取机制确定的。按 J.L. Goldstein 的理论，周期性音高知觉分为两个阶段。第一阶段，外周滤波机制将声信号的成分加以分离，并以一定的精度量出它们的频率。而且，假定这种分离工作是以听神经纤维发射的时间模式为基础，从而把耳蜗生理学和心理声学沟通起来。第二阶段，最可能成为基频的 f_0 是由"音高的中央处理器"试着确定的，使 f_0 的谐波和其它频率成分有最好的符合。f_0 就这样被确定为余音音高。

(5) 音色　在心理声学对声音主观属性的多方面研究中，音色是最薄弱的一环。原因可能有两个。首先，音色的内容相当模糊。两个复合声，它们的响度和音高都相同，如果听起来仍有差别，就归之于它们有不同的音色。可见，音色实际上囊括了除音高和响度之外的声音所有的其它主观属性。其次，在心理物理关系上，音色不同于响度和音高，缺少一个可以和它相对应的可定量的主要物理参量。针对上述特点，心理声学利用多元统计分析对音色的实验结果作了多维量表、主成分分析、因素分析和语义分辨等多种处理，这类研究结果之一说明，音色的 90% 可以只用 4 个维度来评估，其中能覆盖最大部分方差 (44%) 的因素是以"钝-利" (*dull-sharp*) 量表表征的，但表征其它维度的量表似乎不适宜作音色的一般描述，而且缺乏可用的词汇来表达。另一结果表明，和频谱形状比较，相位对复合声的音色影响很小，而且主成分分析发现，音色差异的主要维度和频谱差异的主要维度非常一致。

2. 近期发展——频谱时间模式的分析

近十年来，复合声知觉的研究有了更多方面的发展，特别在频率选择性、时间选择性、相位感受性和非线性效应等方面，限于篇幅，本文不拟介绍。下面列举的是对复合声的频谱时间模式的分析可能产生影响的三项研究。

(1) 谱形分析 (*Profile Analysis*)

D. M. Green 的实验发现，听系统对一个声信号的频谱形状的整体变化非常敏感。例如频谱中某一成分的强度有几个 dB 的增减，听者便能识别出来。由于实验的安排让信号的总声级在 50dB 上下作动态范围为 40 dB 的随机呈现，保证了听者不致用一般强度差别阈的标准来进行判断。即是说，听者在实验中不是对局部信息，而是把变化了的成分置于未改变的其它成分的背景下作为一个整体信息来进行比较分析的。实验表明，相位在这种分析中实质上不起什么作用，但是频率却有明显影响。最佳的分辨效果在 300—3000Hz 的范围内。谱形分析也超越了临界频带的限度：距中心频率 1kHz 以上 1.5 个倍频程的成分对识别也有显著影响。这进一步说明，听者比较的是整个频谱形状。更有趣的结果是听者的经验或训练对分析的效果很有影响。谱形分析研究上述有趣现象的意义在于它可能对听系统在言语波中怎样进行复杂的频域内的加工，对不同强度下何以能保持言语知觉的常性，提供一些启示。

(2) 协调制去掩蔽 (*Co-modulation masking release*) 效应

简称为 CMR 的这一效应是由 Hall, Haggard, Fernandes 等人首先提出的，可以看成是上述谱形分析的延续。他们发现，从窄带噪

声中觉察一个纯音的效果，在增加一个时间包络与噪声相似的另一窄带噪声后，可以得到大大的改进，所以称之为协调制去掩蔽。这一效应可能和临界频带间的相互作用有关。过去，Fletcher 在确定临界带宽时采用了阈值法，即围绕一中心频率，改变掩蔽噪声的带宽来观察被掩蔽纯音的阈值变化.当带宽达到某一值后，再增加带宽，纯音阈值维持不变，表明增加的成分对纯音不起掩蔽作用。临界带宽便由此确定。Hall 等人的实验与此类似。但用一个 0—50Hz 的低通噪声来调幅掩蔽噪声。实验发现，被调幅噪声带宽超过临界频带，纯音阈值继续下降.最大掩蔽效应在 1kHz 时可达 10dB 左右。这可能表明，在协调制条件下，相邻频带相互作用产生了解除掩蔽的效果。

(3) 知觉分组 (*Perceptual Grouping*)

这类研究虽然早在 70 年代便已开始，由于它涉及面广，和复合声知觉关系较大，在此一并提出该是适宜的。当人们听音乐时，并非对每个出现的单元作简单的处理，而且将某些单元归并，形成组的序列。这样的知觉分组一旦构成，进一步的趋向是某一组被置于注意的前景，而其它组退居背景的地位。这种图形-背景式的知觉结构的稳定性，取决于所听音乐的类型。例如带有伴唱的歌曲，歌声便有主次之分。与此相反，在对位性的音乐中，人们却想尽可能听到歌声的全部。也可以说，这时注意力徘徊于交互变动的图形-背景结构之间。

知觉分组所依据的原则，主要是知觉心理学中的近似律 (*Law of Proximity*)。对声信号的知觉，分组的基础可以是音高的范围（音域），也可以是音色和连续性。例如当独奏乐器演奏一个旋律及其伴奏时，两者的频率范围通常是不同的，就是以音域为基础的分组。Miller 和 Heise 曾研究过频率差别对快速纯音序列知觉的影响。他们观察到，当纯音频率相差小于 15%，听到的将是一单串有关系的纯音（即颤音）。若频率差别加大，纯音序列则被听为两个间断的毫无关系的纯音。在莫扎特，贝多芬等古典作家的音乐中，相邻短句常用不同的乐器演奏，就是利用音色分组来表现乐曲结构的例子。用连续性为基础的知觉分组，可以 Divenyi 和 Hirsh 的实验为例。他们发现，在识别三个纯音的时间秩序时，若序列的频率改变是朝一个方向，识别比较容易。Van Novrden 也有类似的发现。他所研究的是在不同呈现率的条件下能听到时间连贯性所必须的三个纯音序列的连续成分之间的最小间距。实验表明，当序列的三个纯音属同一方向时，能听到时间连贯性所需要的音高改变率比有两个纯音的序列所需要的大或者相等。但若序列的三个纯音属于两个方向，听到连贯性之前，频率改变率必需大大降低。

在上述三类研究中，谱形分析所用的声刺激是稳态的。其余两类则不同，声刺激不仅有频谱信息，而且有时间信息，这标志着复合声知觉的研究已进入一个新时代，即刺激是用频谱时间模式表征的，而且对刺激的知觉反应也由简单的觉察深化为多维的模式识别。今后心理声学的研究若能坚持这一方向，不脱离现实生活中的声音世界，其丰硕成果将是可以预期的。

四、言语知觉

言语知觉是听觉基础研究中比较活跃的领域。它受到邻近领域的冲击，也给它们以回击，但同时又受惠于它们，使之充满生机，有希望在近期取得突破。Pisoni 把当前言语知觉研究中存在的问题归结为：(1) 声学语音不变量的缺乏和切分，(2)言语声的内在表征，(3)言语知觉单元，(4)言语声规一化，(5)言语知识源的交互作用。这些问题反映出言语知觉研究所受到的主要冲击是来自工程和人工智能界。为了解决言语识别系统对自然连续语言的理解，急待改进机器前终端的性能。我们将着重讨论第一个问题，它似乎是这些问题的核心。

声学语音不变量和切分是 40 年代伴随着语图同时产生的问题。通过语图，人们恍然了解到语声原来如此多变和连绵不断。积数十年的努力仍然找不到可以和感知到的语音单位唯

一对应的声段和言语波的特征，也找不到严格的声学标准可以在连续语流中确定一个词的终点和另一词的起点．当问题由实验室挪到言语识别系统的设计室，其棘手的程度变得更加突出．但对人类听者说，却是一个轻而易举不成问题的问题．不管语音变异来自何方：语境也好，句法或语义也好，形形色色的说话人和不同的说话情趣也好，一直到传输失真，人类言语知觉都能以不变应万变，保持着"知觉常性"．

面对知觉有常而语声无常的困境，有过几种摆脱的选择．一种是相信存在着言语知觉的声学语音不变量，坚持继续寻找．从近年的发展看，这不一定是条死胡同，且光明似乎在望．Fant 关于相对不变量的讨论，Stevens，Blumstein 对塞音发音部位提出的起始处大频谱形状的不变特征，Kewley-Port 等人提出的动态不变量，都有一定的说服力．更让人寄以希望的是语图判读技术的发展，其识别准确率超过80%，可以和语音专家审音技术比美．它表明言语波中携带的声学语音线索有待进一步的发掘．另一种相近的选择是企图从语音的不同变异源和不同语境因素影响的分析中抽象出规律和一般原则的不变性．但这一工作包含的范围太大，需要的时间很长，不是常规工作条件所能设想的．可喜的是最近几年大数据库的推广使用，已为这项有希望的工作创造了条件．实际上，编集多个发音人的自然语言的大数据库已用来对言语中某些现象的出现率作出量的估计，并为机器的言语识别改进了算法和决策策略．

另一种不同的选择采用由顶往下（*Top-Down*）的办法，即模拟人的认识能力和利用高层次的言语知识来弥补声学语音信息的不足．由于历史原因，这正好是过去言语知觉研究的薄弱环节，直到近几年才有所发展．有关语词的识别，目前已提出多种模型．中心的问题涉及来自听外周的感觉信息的输入和高层次背景信息的交互作用．某些模型主张早期感觉信息是独立加工的，不受上层知识的影响，认知的作用发生在知觉后对决策标准的重新调整的过程中．另一些模型则强调不同知识源对早期感觉分析的影响．从实用角度看，目前一些言语识别系统采用的就是这一选择．以 Harpy 为例，它的音位识别的准确率为45%，表明这一途径的局限性．

自然，人们不会忘记另一种选择．一些认为从言语声波寻找声学语音不变量已经绝望的学者，把目光从声波转向声源，试图从发音器官的活动寻找不变量，并以此作为言语知觉的中介．作为一种言语知觉理论，应该肯定其中的合理因素．但几十年的实践效果表明，对解决言语知觉的有效线索来说，它似乎不是一条成功之路．

从上述讨论，我们是否可以说，呼之欲出的言语知觉的突破口．莫非就在这些选择的汇集点上，最后殊途同归．我们翘首以待．

参 考 文 献

[1] Berlin C. I., Recent Advances: Hearing Research, College-hill Press San Diego California, 1986.

[2] de Boer E. and Dreschler W. A., Ann. Rev. Psychol. 1987, 181—202.

[3] Evans E. F., in Noise Pollution ed Saenz A. L. and Stephens R. W. B., 1986, Scope. John Wiley & Son Ltd, 1 83—1 97.

[4] Flock A., Progress In Brain Research **74** 1988, 297—304.

[5] Green D. M. and Bernstein L. R., in The Psychophysics of speech Perception ed Schouton M. E. H., NATO ASI Series, 1986 314—327.

[6] Kemp D. T., Adv. Audiol. **5**, 1988, 27—45.

[7] Pickles J. O., An Introduction To The Physiology of Hearing, Academic Press, 1982.

[8] Pisoni D. W. J. Acoust. Soc. Am, **78**-1 (1985), 381—388.

[9] Plomp R., in Basic Issues in Hearing ed Duifhuis H. J. and de Wit, Academic Press, 1988, 2—13.

[10] Samuel A. G. and Tartter V. C., Am. Rev. Anthropol., 1986, 15, 247—273.

[11] Teas D. C., Ann. Rev. Psychol, **40**, 1989.

[12] Watson C. et al, J. Acoust. Soc. Am, **78**-1 (1985), 295—298.

[13] Zurek P. M., J. Acoust. Soc. Am, **78**-1 (1985), 340—344.

[14] Zwicker E., in Handbook of Sensory Physiology, Vol. 5 Part 2 ed Keidel W. and Neff W., Springer, Heidelberg, 1975, 401—448.