

• 研究构想(Conceptual Framework) •

听觉信息掩蔽的亚成分：基于行为和脑成像研究的证明^{*}

杨志刚¹ 张亭亭¹ 宋耀武¹ 李 量²

(¹ 河北大学教育学院, 保定 071002) (² 北京大学心理学系, 北京 100871)

摘 要 听觉掩蔽的机制是解决嘈杂声学环境下对特定目标声音进行加工, 即“鸡尾酒会”问题的关键。听觉掩蔽可以分为能量掩蔽和信息掩蔽两种类型。前者是因为目标和掩蔽声音在听觉外周在时间和频率上的重叠所造成的, 而后者被认为是掩蔽声音在听觉中枢和目标声音竞争加工资源所造成的。长久以来, 信息掩蔽一直被当作一种单一成分的现象, 这一概念框架已经成为制约对其机制进行深入研究的一个瓶颈。信息掩蔽中至少包含有知觉信息掩蔽和认知信息掩蔽两种亚成分, 它们源于不同的中枢机制。多个说话人声音掩蔽的条件下, 总体掩蔽量是能量掩蔽、知觉信息掩蔽和认知信息掩蔽等成分总和的结果。操纵掩蔽和目标声音的知觉空间分离、掩蔽声音的可懂度以及掩蔽和目标声音的知觉相似度, 可以实现对两种信息掩蔽亚成分的双重分离。应用功能性核磁共振成像技术可以发现两者有不同的神经机制。

关键词 信息掩蔽; 知觉信息掩蔽; 认知信息掩蔽; 双重分离; 脑机制

分类号 B842

1 引言

嘈杂声学条件下对特定目标声音的追踪和识别问题——即鸡尾酒会问题, 自上世纪 50 年代由英国科学家 Collin Cherry (Cherry, 1953; Cherry & Sayers, 1956) 提出以来, 一直是来自于心理学、神经科学、工程学、人工智能等领域中众多研究者所关注的焦点问题。尽管经过半个多世纪的研究, 人们已经积累了关于此问题的大量知识, 但时至今日, 仍有很多问题是悬而未决的。当今最先进的人工言语识别算法在噪声干扰条件下的言语加工同人类的听觉系统相比仍然有很大的差距(Kim & Stern, 2011), 这其中很重要的一个原因在于我们对人类本身在相似条件下的言语加工机制还不是很清楚。

听觉掩蔽现象以及人类听觉系统的抗掩蔽性能是鸡尾酒会问题的核心。自然条件下的言语交流往往会受到来自于外界噪声——这其中既包括环

境噪声, 也包括目标说话人以外的其他人的声音——的干扰。但是经验表明, 人类听觉系统在这种条件下的表现仍具有很强的鲁棒性(robustness), 表明它具有很强的抗掩蔽性。听觉系统是如何做到这一点的? 如何才能设计出同人类听觉系统一样鲁棒的言语识别系统, 并将其应用于计算机听觉以及人工耳蜗等领域, 这是摆在很多研究者和工程人员面前的一道难题。

然而, 考察当前国内外的研究现状, 我们发现对于听觉掩蔽机制的研究陷入了一种瓶颈状态。对于听觉掩蔽——特别是其中的信息掩蔽, 人们至今不清楚其发生机制是什么。比如, 干扰声音的哪些特性会导致信息掩蔽的发生? 信息掩蔽是在听觉系统中的哪一阶段发生的? 人脑的自上而下加工机制在听觉掩蔽的释放(release from informational masking)中所起的作用是什么? 信息掩蔽的神经生理机制是什么? 等等。所有这些问题研究者都缺乏比较明确的答案。我们认为, 所以出现这些困难, 是因为我们对信息掩蔽概念认识的一种误区, 即长久以来一直把它当成一种单一成分的现象去研究。近年来虽然有一些研究者尝试考察信息掩蔽的亚成分问题, 但相关理论

收稿日期: 2013-09-10

^{*} 国家自然科学基金青年基金项目(31200759)。

通讯作者: 杨志刚, E-mail: yangzg.psy@gmail.com

并不成熟, 行为和神经生理证据也很缺乏。

基于此, 我们提出一种新的听觉掩蔽概念框架, 在这一概念体系中, 信息掩蔽并非一种单一成分的现象, 而是包含知觉和认知两种不同的亚成分。我们将以一系列行为研究和脑成像研究证明这两种亚成分的存在和相对独立性。在行为层面上, 我们将证明, 知觉掩蔽和认知掩蔽可以实现双重分离(double dissociation), 在脑成像研究中, 我们将证明两种掩蔽成分有不同的脑激活模式。

1.1 能量掩蔽与信息掩蔽

在一个鸡尾酒会场景中, 听者需要利用各种线索从混合的声音流中区分出来自于不同声源的信息(segregating), 然后再将来自同一个声源的信息整合(grouping)成一个声音流进行加工(Bregman, 1990)。在此过程中, 干扰声音可以在声音加工的多个阶段对目标声音造成干扰, 导致听者对目标声音不能进行很好的追踪与识别, 这种现象即为听觉掩蔽(Moore, 2003; Watson, 2005; Durlach, Mason, Kidd, et al., 2003)。

听觉掩蔽研究对于鸡尾酒会问题的解决具有重要作用。早期关于听觉掩蔽的研究主要聚焦于干扰声在听觉外周阶段对目标声音的干扰。当来自于不同声源的信息进入到人的听觉系统时, 会在听觉外周(耳蜗和听神经)发生时间和频率上的重叠, 这种重叠会在一定程度上破坏目标声音的表征, 从而对目标声音的识别造成干扰, 这种掩蔽称为能量掩蔽(Watson, 1987)。但因为人的言语信号所具有的内在冗余性, 即使存在能量掩蔽, 个体还是可以通过干扰声音频谱的缝隙(glimpses)对目标声音进行一定的加工(Cooke, 2006)。

然而, 频谱的重叠不能用来解释所有的听觉掩蔽现象。很多研究表明, 既使在没有能量掩蔽的条件之下, 听觉掩蔽依然可以发生。比如, Kidd等人利用非言语的纯音模式(tone pattern)所进行的研究中, 即使目标信号和干扰信号间没有任何频谱重叠, 但如果干扰信号和目标信号之间具有相似的变化模式时, 依然会有大量的听觉掩蔽发生。并且相对于能量掩蔽而言, 这种形式的掩蔽更容易为目标和信号之间的空间分离所释放(Kidd, Mason, Rohtla, & Deliwala, 1998)。应用言语材料所进行的研究得到了类似的结果, Arbogast等人应用滤波技术使干扰声音和目标声音处于不同的频率带上。在相同的空间条件下, 不同频带

的言语信号所造成的掩蔽量要比不同频带的噪声所导致的掩蔽量高20 dB左右(Arbogast, Mason, & Kidd, 2002)。研究者们把这种不能用能量掩蔽进行解释的掩蔽现象称为信息掩蔽(Freyman, Balakrishnan, & Helfer, 2001, 2004; Agus, Akeroyd, Gatehouse, & Warden, 2009; Watson, 2005; Durlach, Mason, Kidd, et al., 2003; Shinn-Cunningham, 2013)。

同能量掩蔽相比, 信息掩蔽在很多方面都表现出明显的不同之处。这些差异主要表现为如下几个方面: 首先, 相对于能量掩蔽, 信息掩蔽对个体的影响存在更大的个体差异, 不同个体的识别阈限差异可达59 dB (Neff & Dethlefs, 1995), 相对于成人, 儿童更容易受到信息掩蔽的影响(Wightman, Kistler, & O'Bryan, 2010); 其次, 通过训练, 被试在信息掩蔽条件下的成绩可以得到一定程度的改善(Leek, Dorman, & Summerfield, 1987)。受过系统音乐训练的被试相对于没有受过训练的被试在信息掩蔽条件下可以取得更好的成绩(Oxenham, Fligor, Mason, & Kidd, 2003); 第三, 人脑的高级认知功能, 比如选择性注意以及个体关于目标或掩蔽的先行知识可以显著提高个体在信息掩蔽条件下的成绩, 但这些操作对于能量掩蔽的作用却很小(Wu et al., 2011; Agus et al., 2009)。

综合这些差异, 我们可以这样认为, 能量掩蔽因为受限于听觉外周自身的时间和频率处理能力, 在很多情况下, 除了在物理上提高信噪比之外, 没有更好的办法对其加以解决; 但在信息掩蔽条件下, 听者自身具有更强的主动性。我们可以通过适当的手段使信息掩蔽在很大程度上得以释放(release), 从而改善在言语交流中的表现。

正是基于这一原因, 在过去的10多年间, 信息掩蔽受到了来自于听觉领域中众多研究者的广泛关注。那么, 信息掩蔽的发生机制是什么? 到底是掩蔽声中的哪些特性导致了信息掩蔽的发生? 信息掩蔽是一种单一成分的现象, 还是可能存在若干亚成分? 对于这些问题, 目前研究者们还没有很好的回答。

1.2 信息掩蔽的机制

信息掩蔽这一术语在很长一段时间内都被人们当成一个行李箱词汇(suit-case word, Watson, 2005; Shinn-Cunningham, 2008, 2013), 用以解释

所有不能用目标和掩蔽声音间在时频上的重叠所造成的干扰现象。换言之,信息掩蔽一直被当成一种单一成分的现象加以看待。这一概念框架在一定程度上制约了对信息掩蔽机制进行深入的研究。已经有越来越多的研究者指出其局限性(Watson, 2005; Durlach, 2006; Mattys, Brooks, & Cooke, 2009; Mattys, Carroll, Li, & Chan, 2010)。但近来有两种不同的观点在探讨信息掩蔽的机制方面产生了一定的影响。这两种观点试图从不同的角度对信息掩蔽进行分类。一种观点认为信息掩蔽的发生是因为刺激的不确定性和目标与掩蔽信号间的相似性(Durlach, Mason, Kidd, et al., 2003; Durlach, Mason, Shinn-Cunningham et al., 2003; Lutfi, Chang, Stamas, & Gilbertson, 2012)。另一种观点认为,信息掩蔽产生于目标声音加工时掩蔽声音的呈现所带来的加工负荷,依据加工负荷类型的不同可以将信息掩蔽分为不同的类型(Cooke, Lecumberri, & Barker, 2008; Mattys et al., 2009; Mattys et al., 2010)。

1.2.1 基于刺激不确定性和目标—掩蔽相似性信息掩蔽分类

所谓的不确定性是指在同一次实验的不同试次间的刺激模式的随机性。研究者们很早就认识到刺激不确定性会导致信息掩蔽的发生(Watson, 1987)。这方面的研究大多是以复合纯音模式作刺激进行的。有很多种方式可以引入刺激的不确定性,比如说随机改变掩蔽声音的频率、时长或起始时间以及空间位置等。研究发现通过随机改变掩蔽声音的频谱,可以使一个特定目标纯音的检测阈限提高 20~40 dB (Kidd et al., 1998)。Fan 等人使用信号检测的方法考察听者对受到信息掩蔽的纯音的辨别力指数 d' , 发现对掩蔽声音的空间位置随机化也可以降低听者的感受性(Fan, Streeter, & Durlach, 2008)。

在一项近期的研究中, Allen 等以言语信号为材料所进行的研究中,发现掩蔽声位置的随机化会使听者对目标音节的识别阈限增加 2.6 dB 左右。对目标声音位置的随机化也会起到相似的效果(Allen, Alais, Shinn-Cunningham, & Carlile, 2011)。

另一种可能导致信息掩蔽发生的机制是目标和掩蔽声音间的相似性。目标和掩蔽声音之间在很多维度上都有可能彼此相似,比如强度、频率、

知觉特性、语义特性等等。这些维度上的相似性均会增加掩蔽声音对目标声音的掩蔽效果。比如,同性别的说话声之间所造成的掩蔽要大于不同性别的说话声(Brungart, Simpson, Ericson, & Scott, 2001),相同语言间的掩蔽效果要大于不同语言间的(Freyman et al., 2004)。

Watson 提出应该把总体的信息掩蔽区分为两种亚型,即基于刺激不确定性的信息掩蔽(IMu, Informational masking uncertainty)和基于目标-掩蔽相似性的信息掩蔽(IMS, Informational masking similarity) (Watson, 2005)。然而,这种对信息掩蔽的分类法存在着一些难以解释的问题。首先,不确定性和相似性的概念都难于进行精确的界定。关于不确定性的研究多数都是以非言语的纯音模式作为刺激进行的,在这种条件下,不确定性被定义为不同试次间刺激变化的随机性。但这一定义很难用于描述言语刺激材料,也许正是基于这一原因,以言语刺激为材料考察不确定性在掩蔽中作用的研究才很少见。

另一方面,对于目标和掩蔽之间的相似性,更难于给出一个客观化的定义。我们可以说相同性别的说话人声音之间的相似度高于不同性别的说话人声音,同种语言的相似度高于不同种语言,但这种判断都是基于主观经验做出的。要从操作的层面上对其给出明确的界定却很难。并且,目标和掩蔽间的相似可能体现在从知觉特性到语义特性等诸多方面,很难相信不同维度间的相似所导致的信息掩蔽的增加是同质的。

再次,不确定性和相似性间的关系很难明确区分。目标和掩蔽声音间相似性的增加必然会导致听者在对目标和掩蔽进行区分时不确定性的增加。这种关系也得到了经验数据的支持,比如, Durlach 等人所进行的一项研究中,同时对刺激不确定性和目标与掩蔽间的相似性进行操纵,发现两者间会相互影响,降低目标与掩蔽间的相似性可以在很大程度上克服刺激不确定性所带来的掩蔽效应(Durlach, Mason, Shinn-Cunningham, et al., 2003)。两者之间的相关性为从不确定性和目标掩蔽相似性的角度来区分信息掩蔽不同类型的观点带了很多挑战。

1.2.2 基于加工负荷的信息掩蔽分类

近期有一些研究者试图从不同的角度对信息掩蔽进行分类,这主要包括 Cooke 等人的观点和

Martys 等人的观点。

Cooke 等人将总体的掩蔽分成能量掩蔽和信息掩蔽两种成分, 而信息掩蔽中又被认为一共包含 4 种成分, 分别是刺激成分的错误分配、掩蔽声音所带来的注意竞争、掩蔽声音导致的认知负载的增加和“能识别的语言”所带来的干扰(Cooke et al., 2008)。

刺激成分的错误分配是指当目标声和掩蔽声同时呈现时, 听者在决定可听到的声音成分(audible elements)是来自于哪一具体声源时可能会犯错误。既可能将来自于掩蔽声源的成分当作是目标声源的, 也可能相反。在言语掩蔽言语的条件下, 被混淆的成分有可能是词汇, 但也有可能是一些更小的声音单位, 如音节(syllable)或音素(phoneme)等(Simpson & Cooke, 2005)。

掩蔽声音所带来的注意竞争是指掩蔽声音会和目标声音竞争有限的注意加工资源。目标和掩蔽之间的相似性会加大听者进行区分的困难性。

另一种信息掩蔽的潜在成分是掩蔽声音的语言内容所带来的。相对于听者不理解的言语, 对听者具有可懂度的言语会带来更多的掩蔽。包括 Cooke 等人的研究在内的很多研究者都证明过言语的可懂度会对掩蔽声音所造成的掩蔽现象产生很大的影响, 所用的材料包括对听者不具有可懂度的外语和时间逆转言语等(Cooke et al., 2008)。

Mattys 等近期提出一种类似的观点, 认为信息掩蔽中包含三种成分, 第一种为掩蔽声对注意资源的竞争, 或者说听者为忽视掩蔽声音的存在而消耗的注意资源; 第二种为当掩蔽声音本身为可识别的语言时给目标声音造成的干扰; 第三种为认知负载, 是指掩蔽声音作为一种附加任务对听者加工目标声音所造成的认知资源消耗, 这种附加任务并不一定是听觉的, 一种视觉的附加任务同样会增加个体的认知负载(Mattys et al., 2009; Mattys et al., 2010)。

然而, 不论是前述哪种观点, 都面临着一些共同挑战。首先, 他们所提出的几种不同信息掩蔽成分间的区分含糊不清。比如注意资源竞争和语义干扰被认为是两种不同的掩蔽类型, 但语义相关性也同样会带来注意竞争, 依据什么标准才将其区分开呢? 第二, 缺乏足够的实证证据能够证明两者间的分离, 至于两者间在神经生理机制上的分离, 目前更是没有见到相关的研究。

1.3 多个说话人掩蔽条件下的掩蔽成分研究

在同时有多人进行交流的场景中, 目标说话人的声音会受到来自多个声源信息的干扰, 这些干扰中都包含哪些掩蔽成分? 并且, 这些成分随着掩蔽声中说话人数目的变化会发生怎样的改变? 这些问题的解决对理解听觉掩蔽的机制有着重要的意义。

最早关于多个说话人声音对目标声识别效果影响的研究是由 Miller 完成的。他考察干扰说话声对 VCV 结构(vowel-consonant-vowel structure)识别的影响。结果表明, 随着掩蔽声中说话人数目的增加, 它对目标声音的掩蔽效果会发生一种非线性的变化。1 个掩蔽声音的掩蔽效果远低于 2 个掩蔽声音的, 两种情况下识别阈限相差约 8 dB。2 个声音的掩蔽效果又低于 4 个声音的(言识别阈限相差 3~4 dB)。但是当掩蔽声音增加到 6 个和 8 个时, 掩蔽效果基本没有变化(Miller, 1947)。其他的一些研究也都发现了这种非线性的变化模式, 例如 Duquesnoy 以及 Festen 和 Plomp 的研究都表明, 当掩蔽声音中包含 4~9 个说话人时, 掩蔽效果最大, 而当说话人声音数小于或远大于这个范围时, 其掩蔽效果都会变小(Duquesnoy, 1983; Festen & Plomp, 1990)。

近期的一些研究表明, 当掩蔽声中包含多个说话人的声音时, 掩蔽效果不仅会受到信噪比、掩蔽声音知觉特性(如掩蔽声说话人的性别)的影响(Brungart et al., 2001; Freyman et al., 2004), 而且会受到掩蔽声音可懂度的影响。比如, 当掩蔽声与目标声是同种语言时, 它对目标声音的掩蔽效果要大于可懂度较低或没有可懂度的声音, 如外语或时间逆转的言语(Freyman et al., 2001)。而且, Freyman 等人应用知觉空间分离(perceived spatial separation)的范式还发现, 随着掩蔽声中说话人数目的变化, 知觉空间分离所带来的去掩蔽量会发生改变, 两个声音掩蔽条件下的去掩蔽量最大, 当掩蔽声音数目进一步增加到 3 个、4 个、6 个和 10 个时, 空间去掩蔽量逐渐减小(Freyman et al., 2004)。由于知觉空间分离所导致的掩蔽释放主要是信息掩蔽成分, 这提示随着掩蔽声中说话人数目的增加, 信息掩蔽量也经历了一种非线性的变化模式。

受 Miller (1947)研究的影响, Simpson 和 Cooke 的一项研究考察了多个掩蔽声音对 VCV

(元音-辅音-元音)结构中辅音识别的掩蔽效果。他们所用的掩蔽声音除了自然的混合语音外,还有相应数目的杂语(babble)的波形包络(waveform envelop)所调制的语谱噪声和未经调制的语谱噪声。结果发现,当掩蔽声音为自然的杂语时,辅音识别随掩蔽声中说话人数目的增加呈一种非单调的变化趋势。掩蔽声中说话人数从1个增加到8个时,掩蔽量都随之增加,之后再继续增加说话人数目直到128个时,掩蔽量都没有明显的改变。当掩蔽声音数目继续增加到512个时,辅音识别有一个显著增加的趋势,也就是说,掩蔽量显著降低了。512个掩蔽声音的掩蔽量又高于稳态语谱噪声的。但当掩蔽声音为杂语包络所调制的语谱噪声时,识别率随着说话人数目的增加基本呈一种单调下降的趋势,并且在掩蔽声音数目从2个增加到512个这一区间内,它所造成的掩蔽量都显著低于相应数目的自然杂语的(Simpson & Cooke, 2005)。

这些结果都表明,在多个说话人掩蔽的条件下,总体的掩蔽量中包含不同类型的掩蔽成分。并且不同的掩蔽成分随着说话人数目的增加有不同的变化趋势。但除了能量掩蔽和信息掩蔽这两个大的分类之外,是否还存在一些更为具体的亚成分?前述研究结果暗示了这种可能,比如,当掩蔽声音可懂度不同时,对目标声音所造成的掩蔽量会有差异(Freyman et al., 2001)。并且不论是否具有可懂度,掩蔽声音和目标声音知觉位置的分离都有助于掩蔽量的释放,这暗示它们对目标声音所造成的主要都是信息掩蔽。

然而,综合来看,当前的研究结果并没有能够直接证明多个掩蔽声存在条件下不同的掩蔽成分,特别是不同的信息掩蔽亚成分的分离。此外,当掩蔽声数目变化时,不同的掩蔽类型也会发生变化,那么这些掩蔽亚成分的变化模式是什么?特定变化模式的内在机制又是什么?这些问题目前都还没有相关的研究可以解答。

2 研究构想

基于前述对当前听觉掩蔽研究现状和问题的认识,在本研究中,我们提出一种新的理论框架来解决当前信息掩蔽研究中所面临的困境,并通过系列相关实验来证明这种理论。具体地,我们提出在总体的信息掩蔽中,包含着两种机制不同

的亚成分,分别是知觉掩蔽和认知掩蔽。在此,我们将知觉掩蔽定义为进入听觉系统的不同声音流之间相互影响所造成的掩蔽;而将认知掩蔽定义为因干扰声音所具有的意义对目标声音所造成的掩蔽。前者和知觉声音流的数目以及不同声音流间的可区分度(在频率、空间位置、协同变化等方面)有关;而后者则只和干扰声音(不管其中所包含的声音流的数目)中所包含的有效信息量有关。所谓有效信息,是指可以对听者造成一定程度影响的信息,它取决于信息的可懂度以及信息的个体相关性等因素。比如,我们假定由不同语言所表达的相同意义的句子(如,“我有一个苹果”和“I have an apple”)中所包含的绝对信息量相等,但对听者而言,具有不同语言经验的个体对两种语言的理解能力可能不同,它们所包含的有效信息量就是不同的。或者,可以认为由同一种语言所说的两个词(一个普通名词和一个代表听者名字的词)所包含的绝对信息量是相同的,但对于听者而言其有效信息量是不同的。

我们假定知觉掩蔽和认知掩蔽是相对独立的,这种独立性体现在两个方面。第一,在行为上两种掩蔽可以实现双重分离。通过操纵特定的变量,既可以保持知觉掩蔽恒定,改变认知掩蔽量,也可以保持认知掩蔽恒定,改变知觉掩蔽量。第二,知觉掩蔽和认知掩蔽的解决具有不同的神经生理机制,通过功能成像和事件相关电位等手段,我们预期能够发现分别对应两种不同信息掩蔽亚成分的生理相关。

基于上述假设,在本研究中,我们拟通过两方面的工作来证明听觉信息掩蔽中包含有知觉掩蔽和认知掩蔽两种亚成分,并且它们是相互独立的。首先,在行为实验中,我们将用多个说话人掩蔽和知觉空间分离去掩蔽两种研究范式来考察不同的掩蔽成分在总体掩蔽量中的构成,从而证明知觉掩蔽和认知掩蔽的双重分离。其次,我们将应用功能性磁共振成像手段考察听觉信息掩蔽和认知信息掩蔽的脑机制。具体包括如下几个方面的研究:

(1)掩蔽声音数目和可懂度对无意义语句识别的影响

在多个说话人声音掩蔽的条件下,有许多因素会对总体的掩蔽效果造成影响。通过对各种变量的操纵,可以考察总体掩蔽中各种掩蔽成分各

自的变化趋势。在本研究中, 我们考察在多个说话人掩蔽的情境下, 掩蔽声音数目和可懂度对于汉语无意义语句识别的影响。具体的, 我们操纵信噪比、掩蔽声中说话人数目、掩蔽声和目标声的知觉空间分离、掩蔽声音的可懂度, 考察目标无意义语句的识别率的变化。通过操纵目标和掩蔽声的知觉空间分离, 我们可以分离能量掩蔽和信息掩蔽。通过操纵掩蔽声音的可懂度, 我们可以分离信息掩蔽中知觉成分和认知成分的作用。从而, 我们将证明不仅能量掩蔽和信息掩蔽是可以分离的, 信息掩蔽的知觉成分和认知成分也是可以分离的。

实验所用目标材料是句法正确的无意义语句, 每个句子都包含 12 个汉语音节。这些句子句法正常, 每个句子均包含 3 个关键成分, 主语、谓语和宾语。这 3 种成分也是 3 个关键的双字词。所有的目标语句都由同一个年轻女性的声音读出。

掩蔽刺激分别是包含不同说话人数目(1、2、4、6、8 和 16)的正序和时间逆转汉语无意义语句。这些无意义语句是由 4 个与目标说话人同性别的其他说话人说出的。在实验中根据实验条件的不同将 4 个原始声音进行随机等强度叠加, 得到包含不同说话人数目的混合说话声。在任何情况下, 所叠加的任何两个声音片段的内容都没有重叠。

实验中, 目标信号和掩蔽信号都是由两个扬声器播放。播放目标声音时, 右侧扬声器的信号在呈现时间上始终领先左侧扬声器 3 ms。这样, 基于优先效应原理, 目标声音的知觉位置就是右侧的扬声器。实验采用四因素混合设计。3 个被试内变量分别是掩蔽声音和目标声音的知觉位置关系、掩蔽声音中说话人数目和信噪比。被试间变量为掩蔽声音的可懂度, 分为两个水平, 分别是正序和时间上逆转的汉语无意义语句。

在本研究中我们预期能够得到如下两点主要结果: 1) 不论是对于正序还是时间逆转的无意义语句, 知觉的空间分离都能带来显著的去掩蔽效果, 但正序无意义语句掩蔽条件下的去掩蔽量更大; 2) 随着掩蔽声数目的增加, 总体掩蔽量、知觉空间分离导致的去掩蔽量以及正序和时间逆转无意义语句掩蔽条件下的空间去掩蔽量之差都将发生变化, 并且几者间变化模式不同。

(2) 不同信噪比操纵方式对听觉掩蔽成分的影响

在实验一中, 我们在改变目标声和掩蔽声声强水平对比时所采用的指标是信噪比(signal-to-noise ratio, SNR), 即是说, 我们将目标信号与掩蔽声音总体的能量水平进行比较。在这样的设置中, 由于掩蔽声音对目标声音的掩蔽效果会受到它们之间相互掩蔽的影响, 单个的掩蔽声音流和目标声音在总体的声音混合体中并不是处于一种对等的地位, 因而我们不清楚每一个个体掩蔽声音在对目标声的掩蔽中所起的作用。显然, 在现实的生活场景中, 干扰声音并不总是处在一种模糊的背景的地位的。它们对于目标声的识别和知觉会有何影响? 这种影响随着总体说话人数目的增加又会有何变化? 在本研究中, 我们操纵目标声音与个体掩蔽声音的声强比, 即目标-掩蔽比率(target-masker ratio, TMR), 并考察随着掩蔽声中说话人数目的增加, 掩蔽会发生怎样的变化。

实验二在实验设计上与实验一相似, 所用的掩蔽刺激材料也与实验一相同。但不同的是, 我们在本研究中不再以信噪比(SNR)作为度量目标掩蔽声音强度差的指标, 而是改用 TMR。在这种度量框架下, 我们可以考察掩蔽声中每一个体掩蔽声音对于总体掩蔽效果的影响。在实验一中, 随着掩蔽声数目的增加, 由于掩蔽声音间彼此的掩蔽, 使得掩蔽声音流的增量效应受到削减(随着掩蔽声音数目的增加, 等强度叠加的掩蔽声音倾向于融合成一个单独的声音流), 因而掩蔽效果会随着声音流数目的增加有一种非单调的变化趋势。但在实验二中, 目标-掩蔽声强关系是根据目标声与单个掩蔽声声强之比操纵的, 因而, 在掩蔽声数目增加时, 总体掩蔽效果并不会因为掩蔽声之间的相互掩蔽(掩蔽的掩蔽)而消减。因而, 我们预期总体的掩蔽量及知觉空间分离所带来的去掩蔽量不会随着掩蔽声中说话人数目的增加出现非单调变化趋势。

(3) 掩蔽声音知觉属性对于听觉掩蔽的影响

在前面的两个研究中, 我们主要通过操纵掩蔽声音的可懂度(信息量)来考察在知觉掩蔽成分不变的情况下, 认知掩蔽是如何变化的。然而要证明知觉掩蔽和认知掩蔽的分离, 只是单方面的分离是不够的。我们还需要能证明认知掩蔽不变的情况下, 知觉掩蔽可以独立发生变化, 从而证明知觉掩蔽和认知掩蔽之间可以发生双重分离。因此, 在本研究中, 我们将在平衡掩蔽声音中语

义信息作用的同时,改变其与目标声音在知觉属性上的关系,从而进一步考察两种因素在总体听觉掩蔽中的作用。我们预测,当掩蔽声音和目标声音在知觉属性上差异很大,从而使其很容易与目标声音分成不同的声音流时,掩蔽声音的影响将会降低。

因而在本研究中,我们将在平衡掩蔽声音可懂度的同时操纵其知觉属性。具体的,我们采用与实验一相似的实验设计。但在掩蔽声音材料上有所差别。掩蔽声音都是正序的汉语无意义语句,但是是由不同性别的说话人说出的。

(4)不同掩蔽亚成分在神经生理机制上的分离

前述几个实验旨在行为层面上证明知觉掩蔽成分和认知掩蔽成分间可以发生双重分离。但要证明它们确实是两种相互独立的信息掩蔽亚成分,还必须从神经生理机制上证明其可分离性。Scott 等人应用正电子发射断层扫描(PET)技术发现能量掩蔽和信息掩蔽条件下大脑皮层有不同的激活模式(Scott, Rosen, Wickham, & Wise, 2004; Scott, Rosen, Beaman, Davis, & Wise, 2009)。是否信息掩蔽的不同亚成分,即知觉掩蔽和认知掩蔽也是由大脑皮层的不同模块调节的?在本研究中,我们将进一步通过一项核磁共振成像(fMRI)实验证明两种不同的掩蔽成分是由不同的脑区负责调节的。

在本项目中,我们将通过一项功能性核磁共振成像(fMRI)研究来考察掩蔽声中声音流数目和可懂度内容分别变化时大脑不同区域的活动模式。实验中通过空气传导式耳塞给被试呈现刺激,分别考察知觉属性(掩蔽声音流数目、掩蔽声音和目标声音间的知觉相似性)改变和认知属性(如掩蔽声音可懂度,掩蔽声音语速)改变时大脑皮层的激活模式。

3 小结与展望

信息掩蔽概念的笼统和模糊已经成为制约人们对其机制进行深入研究的一个瓶颈。虽然有大量研究考察它的各种影响因素,但是对于它到底是在中枢系统的哪些阶段、以及以何种方式发生的,至今依然没有一个合理的描述。已经有越来越多的研究者认识到这一概念的局限性。本研究提出信息掩蔽并不是一个单一成分的概念,而是包含有知觉和认知两种子成分。其中知觉掩蔽子

成分的发生是因为进入听觉系统的不同声音流(或声音客体)相互竞争所造成的,而认知掩蔽子成分是因为掩蔽声中所包含的高级信息(包括语言学、句法、语义甚至情绪等信息)具有自动加工特性,从而竞争高级认知资源所造成的。知觉掩蔽受总体声音流数目和掩蔽声与目标声知觉相似度的影响,而认知掩蔽和掩蔽声中总体的有效信息量有关。知觉掩蔽和认知掩蔽是相对独立的,这种独立性可以通过行为上的双重分离和脑成像研究中的激活区分离来证明。

本研究所提出的信息掩蔽概念框架是解决当前信息掩蔽概念模糊性的一种理论尝试。这一理论框架能够给既有的大量信息掩蔽相关研究发现(如掩蔽声可懂度对其掩蔽效果的影响、掩蔽说话人性别对掩蔽效果的影响、实际和知觉空间分离的掩蔽释放效应,以及听觉启动效应等)给出一种相对简单的解释。同时,该理论也可以生出一些可以直接进行验证的假设(如本研究中所提出的行为上的双重分离和神经机制上的分离)。

未来的研究一方面可以继续为知觉和认知掩蔽亚成分的功能独立性寻求实证证据,另一方面可以对包含能量掩蔽、知觉(信息)掩蔽和认知(信息)掩蔽在内的整体掩蔽框架进行计算机建模。这些工作对于我们理解竞争情境下的言语加工机制、从而最终解决鸡尾酒会问题具有重要的意义。

参考文献

- Agus, T. R., Akeroyd, M. A., Gatehouse, S., & Warden, D. (2009). Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise. *The Journal of the Acoustical Society of America*, 126(4), 1926–1940.
- Allen, K., Alais, D., Shinn-Cunningham, B., & Carlile, S. (2011). Masker location uncertainty reveals evidence for suppression of maskers in two-talker contexts. *The Journal of the Acoustical Society of America*, 130(4), 2043–2053.
- Arbogast, T. L., Mason, C. R., & Kidd, J. G. (2002). The effect of spatial separation on informational and energetic masking of speech. *The Journal of the Acoustical Society of America*, 112(5), 2086–2098.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K.

- R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, 110(5), 2527–2538.
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25(5), 975–979.
- Cherry, E. C., & Sayers, B. (1956). Human “cross-correlation”—A technique for measuring certain parameters of speech perception. *The Journal of the Acoustical Society of America*, 28(5), 889–895.
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, 119(3), 1562–1573.
- Cooke, M., Lecumberri, M., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, 123(1), 414–427.
- Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *The Journal of the Acoustical Society of America*, 74(3), 739–743.
- Durlach, N. I. (2006). Auditory masking: Need for improved conceptual structure (L). *The Journal of the Acoustical Society of America*, 120(4), 1787–1790.
- Durlach, N. I., Mason, C. R., Kidd, J. G., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. (2003). Note on informational masking (L). *The Journal of the Acoustical Society of America*, 113(6), 2984–2987.
- Durlach, N. I., Mason, C. R., Shinn-Cunningham, B., Arbogast, T. L., Colburn, H. S., & Kidd, J. G. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *The Journal of the Acoustical Society of America*, 114(1), 368–379.
- Fan, W. L., Streeter, T. M., & Durlach, N. I. (2008). Effect of spatial uncertainty of masker on masked detection for nonspeech stimuli (L). *The Journal of the Acoustical Society of America*, 124(1), 36–39.
- Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, 88(4), 1725–1736.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 109(1), 2112–2122.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 115(5), 2246–2256.
- Kidd, G., Mason, C. R., Rohtla, T. L., & Deliwala, P. S. (1998). Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*, 104(1), 422–431.
- Kim, W., & Stern, R. M. (2011). Mask classification for missing-feature reconstruction for robust speech recognition in unknown background noise. *Speech Communication*, 53(1), 1–11.
- Leek, M. R., Dorman, M. F., & Summerfield, Q. (1987). Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 81(1), 148–154.
- Lutfi, R. A., Chang, A. C., Stamas, J., & Gilbertson, L. (2012). A detection-theoretic framework for modeling informational masking. *The Journal of the Acoustical Society of America*, 132(2), 109–113.
- Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, 59(3), 203–243.
- Mattys, S. L., Carroll, L. M., Li, C. K. W., & Chan, S. L. Y. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication*, 52(11), 887–899.
- Miller, G. A. (1947). The masking of speech. *Psychological Bulletin*, 44(2), 105–129.
- Moore, B. C. (2003). *An introduction to the psychology of hearing* (4th ed.). San Diego: Academic press.
- Neff, D. L., & Dethlefs, T. M. (1995). Individual differences in simultaneous masking with random frequency, multicomponent maskers. *The Journal of the Acoustical Society of America*, 98(1), 125–134.
- Oxenham, A. J., Fligor, B. J., Mason, C. R., & Kidd, G. Jr. (2003). Informational masking and musical training. *The Journal of the Acoustical Society of America*, 114(3), 1543–1549.
- Scott, S. K., Rosen, S., Beaman, C. P., Davis, J. P., & Wise, R. J. (2009). The neural processing of masked speech: Evidence for different mechanisms in the left and right temporal lobes. *The Journal of the Acoustical Society of America*, 125(3), 1737–1743.
- Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *The Journal of the Acoustical Society of*

- America*, 115(2), 813–821.
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, 12(5), 182–186.
- Shinn-Cunningham, B. G. (2013). Understanding informational masking from a neural perspective. *Proceedings of Meetings on Acoustics*, 19(1), 1–3.
- Simpson, S., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N (L). *The Journal of the Acoustical Society of America*, 118(5), 2775–2778.
- Watson, C. S. (1987). Uncertainty, informational masking, and the capacity of immediate auditory memory. In W. A. Yost, & C. S. Watson (Eds.), *Auditory processing of complex sounds* (pp. 267–277). NJ: Erlbaum Associates Hillsdale.
- Watson, C. S. (2005). Some comments on informational masking. *Acta Acustica United with Acustica*, 91(3), 502–512.
- Wightman, F. L., Kistler, D. J., & O'Bryan, A. (2010). Individual differences and age effects in a dichotic informational masking paradigm. *The Journal of the Acoustical Society of America*, 128(1), 270–279.
- Wu, X. H., Yang, Z.-G., Huang, Y., Chen, J., Li, L., Daneman, M., & Schneider, B. A. (2011). Cross-language differences in informational masking of speech by speech: English versus mandarin Chinese. *Journal of Speech, Language, and Hearing Research*, 54(6), 1506–1524.

The Subcomponents of Informational Masking: Evidence from Behavioral and Neural Imaging Studies

YANG Zhigang¹; SONG Yaowu¹; ZHANG Tingting¹; LI Liang²

(¹ College of Education, Hebei University, Baoding 071002 China)

(² Department of Psychology, Peking University, Beijing 100871 China)

Abstract: The processing of a target speech under noisy environment is influenced by the presence of concurrent sounds, this phenomenon is referred to as auditory masking. Previous research showed that there are two kinds of masking, energetic masking and informational masking. Energetic masking originated from auditory periphery, where the masking sound overlaps with the target sound in time and frequency, resulting in an incomplete representation of the latter. On the other hand, informational masking is believed to occur in the central auditory system. However, for a long time, the concept of informational masking has been used as a “suitcase” or “catchall” word. All the influencing effect that can not be explained by energetic masking is attributed to informational masking, though its underlying mechanism is still largely unknown. In this article, a new conceptual framework of informational masking is proposed. In this framework, informational masking is divided into two subcomponents, perceptual informational masking and cognitive informational masking. Perceptual masking is caused by the competition between the sound stream/object, while cognitive informational masking is caused by the higher cognitive information included in the masking sounds, which could be automatically processed by the auditory or linguistic system. Psychophysical and neural imaging studies are proposed to test the above hypothesis. In behavioral studies, the perceived spatial separation between the target and the masker, the intelligibility and the voice number in the masking sound, as well as the perceptual similarity between the target and masker streams are manipulated to demonstrate that perceptual masking and cognitive masking could be double-dissociated. In the neural imaging study, fMRI technique is used to show that perceptual and cognitive informational masking may have different neural mechanism.

Key words: informational masking; perceptual masking; cognitive masking; double-dissociation; neural mechanism