

“鸡尾酒会”环境中的知觉线索的去掩蔽作用*

徐李娟 黄莹 吴玺宏 吴艳红 李量

(北京大学心理学系、智能科学系、言语和听觉研究中心; 机器感知与智能重点实验室(教育部), 北京 100871)

摘要 在有多人同时说话的嘈杂环境中,为什么具有正常听力的人能在一定的程度上听懂目标语句?研究这个著名的“鸡尾酒会”问题的一个新进展是将干扰言语的作用区分出了能量掩蔽和信息掩蔽两种成分。与发生在外周系统的能量掩蔽不同,信息掩蔽发生在心理层次并受到认知过程的调节。因此,主观空间分离、与目标语句节奏相关的视觉信号以及对目标语句某些特征的熟悉程度等知觉线索都具有去掩蔽作用。考察可减少信息掩蔽的知觉线索的交互作用及其高级认知调节是今后重要的研究内容。

关键词 “鸡尾酒会”问题; 知觉线索; 选择性注意; 能量掩蔽; 信息掩蔽

分类号 B842.2; B849

1 “鸡尾酒会”问题

在嘈杂的室内环境中,比如在鸡尾酒会中,同时存在着许多不同的声源:多个人同时说话的声音、餐具的碰撞声、音乐声以及这些声音经墙壁和室内的物体反射所产生的反射声等。在声波的传递过程中,不同声源所发出的声波之间(不同人说话的声音以及其他物体振动发出的声音)以及直达声和反射声之间会在传播介质(通常是空气)中相叠加而形成复杂的混合声波。因此,在到达听者外耳道的混合声波中已经不存在独立的与各个声源相对应的声波了。然而,在这种声学环境下,听者却能够在相当的程度上听懂所注意的目标语句。听者是如何从所接收到的混合声波中分离出不同说话人的言语信号进而听懂目标语句的呢?这就是Cherry在1953年所提出的著名的“鸡尾酒会”问题(Cherry, 1953)。

自Cherry提出“鸡尾酒会”问题半个多世纪以来,大量的科学家试图去解决这个问题,甚至试图制造一个计算机言语识别的智能系统使其具有在嘈杂环境中识别目标语句的功能。但到目前为止,“鸡尾酒会”问题还没有得到满意的解答。为了更好地理解“鸡尾酒会”问题,我们首先来看看干扰性言语对目标言语能产生哪些掩蔽作用,这对深入认识这个问题的本质有重要的意义。

2 能量掩蔽和信息掩蔽

当掩蔽声音和目标声音同时出现时,尤其当两者在频谱上相互重叠时,干扰声音与目标声音一起激活了听觉外周神经系统中的相同部分,如激活了基底膜的同一部位和同一群听神经细胞。干扰声音的激活作用导致听神经细胞对目标声音反应的动态范围变小,进而导致听觉外周对目标声音的编码失真,降低听觉系统对目标声音的觉察和识别。这就是能量掩蔽的主要机制。能量掩蔽使得进入到中枢阶段的目标信息有实质性的缺失,而这种缺失是高级中枢的加工所难以补偿的。在实验室的研究中,通常使用平稳的语音噪音作为对目标语句产生能量掩蔽的声音刺激(Arbogast, Mason & Kidd, 2005; Freyman, Balakrishnan & Helfer, 2001; Freyman, Helfer, McCall & Clifton, 1999; Li, Daneman, Qi & Schneider, 2004; Wu et al., 2005)。

其他说话人的干扰性语音信号在产生能量掩蔽的同时还产生另一种值得详细研究的掩蔽作用,即作用于认知水平的信息掩蔽(Li et al., 2004; Oxenham, Fligor, Mason & Kidd, 2003; Shinn-Cunningham, Ihlefeld, Satyavarta & Larson, 2005; Summers & Molis, 2004; Wu et al., 2005)。当掩蔽声音和目标声音在某些信息维度上具有一定的相似性时,例如当目标声音与掩蔽声音都是由同性别嗓音所读出的同语种的言语时,目标言语和掩蔽言语在高级认知层次上会争夺有限的心理资源并出现言语编码的混淆,进而使对目标信号的加工受到干扰。

收稿日期: 2008-09-28

* 国家自然科学基金资助项目(基金号: 30711120563; 30670704; 60535030)。

通讯作者: 李量, E-mail: liangli@pku.edu.cn

这种发生在高级加工层次上的干扰作用被称为信息掩蔽 (Brungart, 2001; Brungart, Simpson, Ericson & Scott, 2001; Freyman et al., 1999, 2001; Kidd et al., 2002; Li et al., 2004)。在一个典型的信息掩蔽过程中, 由于干扰言语的能量在时间上有明显的波动, 听觉外周系统可以利用高信噪比的时间段对目标信号进行有效的接收和神经编码, 即听者能够觉察到目标语句的出现。但由于信息掩蔽的作用, 听者仍然不能有效和准确地追随目标语句, 进而导致对目标语句的识别率下降。与能量掩蔽不同, 由于信息掩蔽发生在认知加工阶段, 一些认知操作可以减少或消除由信息掩蔽所引起的对目标语句识别率的影响。

因此我们认为, 认识“鸡尾酒会”问题之本质就是要认识信息掩蔽的本质, 而听者利用不同的知觉线索来减少和消除信息掩蔽作用的认知操作又是一个最为核心的问题。下面我们将对几种重要的相关的知觉线索作一个总结。

3 主观空间分离与信息掩蔽

Cherry (1953) 在他所提出“鸡尾酒会”问题的文章中, 同时也提出了几种可能被用来减少和消除言语掩蔽的线索, 其中一个重要的线索是干扰声音与目标声音之间的空间分离。

根据Zurek (1993) 所提出的理论, 目标声源和干扰声源之间的空间分离主要通过两种效应来减少干扰声音对目标声音的掩蔽。一是头影效应或优势耳效应。由于人的头颅对声音中高频成分的传播具有遮挡作用, 目标声源和干扰声源之间的空间分离可以提高某只耳朵上的信噪比, 从而提高对目标信号的识别。这是一种声学物理的效应。二是目标声源与掩蔽声源之间的空间分离可以造成双耳线索上的差异, 尤其是目标信号与掩蔽信号中的低频成分的双耳时间差的不同产生了双耳神经生理加工的线索, 进而降低了对目标信号的觉察阈限。那么, 目标声源和干扰声源之间的空间分离除了产生物理线索和神经生理线索外, 是否还能产生在心理层次上的线索以减少掩蔽作用?

已有充分的研究结果表明, 在有回声的环境中, 空间分离所带来的头影作用和双耳神经加工的作用都会受到削弱 (Freyman et al., 1999; Kidd, Arbogast, Mason & Gallun, 2005; Zurek, Freyman & Balakrishnan, 2004)。然而, 当掩蔽刺激是言语时, 目标言语与掩蔽言语之间的空间分离仍然能提高听

者对目标言语的识别 (Kidd et al., 2005)。这表明目标言语与掩蔽言语之间的空间分离不仅提供了头影效应和双耳加工这两类线索, 还提供了其他线索, 而且这种额外的线索所引起的去掩蔽作用是Zurek的理论所不能解释的。在认识这种线索之前, 需要介绍一下听觉优先效应的概念。

当两个相关 (correlated) 的声源发自不同的空间位置, 并且它们之间有充分短的延迟 (如1~10ms) 时, 落后声音中的知觉特征 (attributes) 会在知觉层次上被领先声音所“捕获” (Li, Qi, He, Alain & Schneider, 2005), 使得听者只知觉到一个融合的声音, 并且该声音的知觉位置在领先声音处或在其附近。这种现象被称为听觉优先效应 (the “precedence” effect) (近期综述见Li & Yue, 2002)。在有回声的环境下, 优先效应可以使人们似乎感觉不到来自不同方向的反射声, 进而在知觉层次上减少回声的干扰作用, 以帮助听者准确地判断出声源的空间位置。

在模拟回声环境的实验室条件下, 利用优先效应可以使不同的声音信号在知觉上产生空间分离。例如, 当左右两个扬声器都发出A、B两个不同声音时, 对A声音来讲, 左扬声器提前3ms, 而对B声音来讲, 右扬声器提前3ms, 由于听觉优先效应的作用, 两个扬声器发出的A声音就会有主观融合, 其声像来自左扬声器或其附近; 同样, 两个扬声器发出的B声音也会有主观融合, 其声像来自右扬声器或其附近。这样, A、B两个声像就会被知觉为来自不同的空间位置, 从而出现知觉上的空间分离。但实际上, A、B两个声源在每个扬声器都有呈现, 因而没有客观上的分离。Freyman等人 (1999) 首先在他们的实验中运用听觉优先效应实验范式来研究在噪声掩蔽或言语掩蔽的条件下目标语句和掩蔽刺激之间的主观空间分离所带来掩蔽作用的变化。他们的研究表明, 在噪声掩蔽条件下, 主观空间分离只给目标语句带来很小的掩蔽释放; 而在言语掩蔽条件下, 主观空间分离可以给目标语句带来4~9dB的掩蔽释放。这说明主观空间分离可以有效地减少信息掩蔽, 但是对能量掩蔽的作用非常有限。

随后的一系列研究结果也进一步证实, 当干扰声音是言语时 (主要产生信息掩蔽), 利用优先效应所形成的主观空间分离不改变信号的频谱与强度, 但可以帮助听者将注意分配在目标言语上, 进而提高对目标言语的识别。但当干扰声音是语谱噪声时 (主要产生能量掩蔽), 主观空间分离的作用

却很小 (Arbogast et al., 2005; Freyman et al., 1999, 2001; Li et al., 2004; Wu et al., 2005)。因而, 利用优先效应这种范式所产生的主观空间分离可以将信息掩蔽的效果和能量掩蔽的效果进行分离, 使得对信息掩蔽的深入研究成为可能。值得一提的是, 比较汉语和英语之间的主观空间分离对信息掩蔽影响差别的研究发现, 汉语的抵抗信息掩蔽的特性要强于英语的 (Li et al., 2004; Wu et al., 2005)。这可能是因为汉语的丰富而具有语意的语调变化有助于将不同说话人的语句流进行主观分离。有关信息掩蔽的跨语言 (文化) 的研究为探索语言 (文化) 之间的差异提供了一个新的研究方向。

如果目标语句与掩蔽声音之间没有实际的空间分离, 去掩蔽的头影效应和双耳时间差效应就不存在, 但人的中枢系统却能够利用优先效应所造成的目标与掩蔽信号之间的主观空间分离来显著地减少信息掩蔽量。这反映了人脑的一个重要的高级知觉功能。

不同言语信号之间的主观空间分离在实际运用中也有非常重要的意义。例如, 现代空军作战技术的发展对飞行员所用的通讯技术和其本人的心理通讯能力提出了更高的要求。空军飞行员在执行任务时往往会同时接受多个来源的语句信号, 如来自地面指挥部、机群指挥员以及机群中其他担任不同任务的飞行员的言语信号。在某一时刻, 某些来源的语句信号的重要性要大于其他来源的语句信号, 这就对发展更加先进的听觉通讯系统提出了迫切的要求。利用主观空间分离去信息掩蔽的原理以及头传递函数的原理, 可以将飞行员从耳机所接收到的不同来源的听觉信号分布在不同的主观空间上, 并根据该时刻下的优先性来对某一个来源的信号进行方位的确定 (如头的正上方代表地面指挥部的信号位置), 飞行员就会从众多的信号源中更有效地选择和分析不同来源的语句信号。

此外, 近年来听觉的老年问题受到了越来越多的关注。为什么在嘈杂和有混响的环境中, 老年人较年轻人在识别目标语句方面面临更大的困难呢? 最新的研究表明, 目标语句的直达波和反射波之间的知觉整合可以显著地减轻信息掩蔽, 而老年人与年轻人相比, 这种知觉整合的能力要显著的低, 特别是在较长的反射波延迟条件下 (Huang et al., 2008)。这为探索高级听觉认知加工的老年化开辟一条新的路子。

4 视觉线索与信息掩蔽

Cherry在1953年也提出与言语有关的唇读和肢体活动等视觉线索可能会帮助听者对目标语句进行觉察和识别。已有研究表明, 在有干扰情况下, 与目标信号相关的视觉线索 (如唇读) 可以提高听者对目标信号的识别 (Rudmann, McCarley & Kramer, 2003; Wang et al., 2008), 但是这种视觉线索作用的大小与掩蔽声音的类型有关 (Helfer & Freyman, 2005)。Helfer和Freyman (2005) 的研究发现, 当掩蔽声音为语谱噪声时, 提供阅读目标语句的唇读线索可以将被试对目标语音的识别成绩提高3dB。而当掩蔽声音为言语时, 该视觉线索会使被试的识别成绩提高9dB。这就表明视觉线索主要减弱了信息掩蔽。视觉线索可以帮助个体来觉察目标信号的出现时刻, 以助于区分不同的语音流, 从而把注意力集中到目标流上以提高对目标的识别 (Helfer et al., 2005; Rudmann et al., 2003)。听觉通路和视觉通路在外周部分并没有交互作用, 但是视觉线索却能在相当程度上减少听觉通路上的掩蔽作用, 这种对信息掩蔽的减弱作用应当发生在知觉层次甚至更高级的认知加工层次上。这说明在嘈杂环境中目标信号的提取和加工过程既存在自下而上的加工过程, 还受到自上而下的调节。

近期的研究发现, 对汉语来讲, 只有当目标语句的语速不可预测或目标语句本身受到瞬时的强干扰而出现“断裂”时, 与目标语句节奏同步的闪光信号减少信息掩蔽的作用才显著。因此, 这种跨感觉道的知觉线索在口语交流中, 尤其是在非常嘈杂环境下的口语交流中起到了显著的去掩蔽作用 (Wang et al., 2008)。由于唇读的信息成分比较复杂, 而所用的闪光信号只在节奏这一维度上为听者提供与目标语句有关的信息, 表明了听者对目标语句的节奏的掌握也是一个重要的去信息掩蔽的线索。

5 对目标语句的空间位置、内容以及噪音的熟悉程度的作用

在Kidd等人 (2005) 的研究中, 若在实验前给被试提供目标声音的位置信息, 被试在掩蔽的条件下对目标语句识别的成绩有显著的提高。Brungart等人 (2001) 的实验也表明, 如果在刺激呈现之前让被试对目标言语的特征有所了解, 信息掩蔽能够在一定程度上得到降低。这些有关目标声音的位置以及声音特征等方面的先验知识能够引导被试将注

意转移并集中于目标言语上,因而有助于追随目标语句流并提高对目标语句的识别。

对目标语句内容的熟悉和了解可以在言语掩蔽的条件下有效地提高听者对目标语句的识别(Freyman et al., 2004; Yang et al., 2007, 2008)。具体来讲, Freyman 等人(2004)用三种不同的方式在目标语句出现之前呈现目标语句中的一部分内容作为启动刺激,即启动句的内容与随后呈现的目标句的内容完全一致,但是启动句的最后一个关键词用相同长度的白噪声所代替。这三种启动句呈现的方式分别为:(1)由朗读目标语句的嗓音读出,(2)由不同性别的嗓音读出,以及(3)用书面的方式呈现。通过对目标言语最后一个关键词(该关键词在启动句中不出现)的计分表明,无论何种呈现方式,有启动的条件下被试的成绩均显著好于无启动条件的成绩,并且这三个条件下的去信息掩蔽的程度基本一致。这些结果表明,对于目标语句内容的了解,一方面可以使注意能更容易地捕捉和追随目标语句,另一方面由于目标语句的大部分内容已经在启动时得到加工,在目标语句和掩蔽语句同时呈现时,中枢系统对已经加工过的部分无需重新加工或者只需要很少的资源进行再加工,进而节省了注意资源以集中加工随后的关键词并提高对这个关键词的识别成绩。实验中目标语句内容的启动作用不受声音特性及呈现方式的影响,说明这种去信息掩蔽的功能是通过高级语言加工的过程来实现的。

近期的研究除了证实了 Freyman 的发现可以推广到汉语外,还首次发现对汉语目标说话人嗓音的熟悉也能显著地减少对目标语句的信息掩蔽而不显著地影响能量掩蔽(Xu et al., 2008; Yang et al., 2007, 2008)。这表明,在知觉加工的层次上,对朗读汉语目标语句的嗓音的熟悉可以帮助听者将注意力有选择地放在这个嗓音特征上,进而能捕捉和跟随目标语句的“嗓音流”,强化了目标语句流和掩蔽语句流之间的知觉性分离。

综上所述,对目标言语的主观空间位置、语句的节奏特征、目标言语的内容以及目标说话人嗓音的先验知识都可以帮助听者把注意集中到目标语句上进而减少信息掩蔽以提高对目标信号的识别。

6 各种知觉线索之间的交互作用

在嘈杂的“鸡尾酒会”环境中,当目标本身的一些特性保持一定的稳定性时,听者可以利用这

些与目标语句相关的各种知觉线索来觉察、捕捉和追随目标语句流,以达到减少信息掩蔽的目的。这也表明,在“鸡尾酒会”这个重要的问题中,人类高级认知加工过程起了关键的作用。对各种知觉线索的利用是一个复杂的动态过程。在不同情况下,各个线索的显现程度会随时间发生波动,因而听者会在不同的条件下选择利用不同的线索组合来减少信息掩蔽。这样,各个线索之间一定会存在着复杂的交互作用。

Noble 和 Perrett (2002)的研究发现,当其他线索可以被用来帮助被试将目标与掩蔽信号进行分离时,空间的分离线索作用就会相应地减弱。Freyman 等人(2004)的实验也发现,当目标语句与掩蔽语句之间有主观空间分离的时候,有关目标语句内容的知识就不能进一步用于减少信息掩蔽。只有当主观空间分离的线索不存在时,被试才可以有效地利用内容线索来减少信息掩蔽。近期的研究也证实了目标语句的嗓音线索与其他线索之间的交互作用,即嗓音线索的去信息掩蔽作用受到声音的强度线索以及目标与掩蔽刺激之间的延迟线索的影响(Xu et al., 2008)。例如,当目标语句与掩蔽语句之间存在1秒的延迟时(这个时间延迟是一个较强的分离目标信号与掩蔽信号的线索),嗓音线索可以显著地减少信息掩蔽。而进一步加强对目标嗓音的熟悉程度(重复呈现嗓音启动语句或者在实验前对目标嗓音进行学习)不能继续增大去信息掩蔽的量。而当目标语句与掩蔽语句同时呈现时(延迟线索被除去),只有熟悉程度被加强后的目标嗓音才可以被用来有效地减少信息掩蔽。这表明,各种知觉线索去信息掩蔽的作用是可以相互补偿而不能线性叠加的。

目前,我们对于“鸡尾酒会”问题的研究还基本处于了解各个单一线索在去信息掩蔽中的作用的阶段。而要最终破解“鸡尾酒会”问题,需要对各种线索之间的交互作用进行系统的研究,以形成对线索作用的动态特征的认识。

7 总结及展望

人类在嘈杂的“鸡尾酒会”环境下对于目标言语的识别是一个异常复杂的过程,用当前信号加工工程学的理论还无法完全解释这个复杂的过程。干扰言语所产生的信息掩蔽是由于目标言语和掩蔽言语之间在高级加工层次上出现了竞争与混淆。要减少这种在知觉/认知层次上的竞争和混淆,除了增大目标刺激和掩蔽刺激之间声学特性上的差异外

(Drullman & Bronkhorst, 2004; Darwina, Brungart & Simpson, 2003; Freyman, Balakrishnan & Helfer, 2006; Wastson, 2005), 利用与目标语句相关的知觉线索也可以有效地减少信息掩蔽 (Freyman et al., 1999, 2001, 2004; Kidd et al., 2005; Li et al., 2004; Wu et al., 2005; Yang et al., 2007), 这也是解决“鸡尾酒会”问题的关键。

主观空间分离给个体提供目标言语位置的信息, 这些主观空间位置的信息可以促进听者对目标信号的选择性注意。视觉线索作用于跨感觉道的高级加工层次, 也可以将选择性注意集中于与视觉线索相匹配的目标语句的节奏特征上。对于目标语句的内容或朗读嗓音的熟悉有记忆加工的成份在内, 而这种有记忆参与的利用知觉线索所进行的自上而下的认知调节过程反映了人脑在嘈杂背景中提取和加工目标信号的主动的执行功能。

我们认为, 当前有关“鸡尾酒会”问题的研究已经进入了一个关键的历史时期: 大量的研究业已证明低层的物理和生理线索可以有效地去信息掩蔽, 这些作用是通过自下而上的加工实现的, 可以运用工程方法进行模拟; 而作为更重要的高层知觉线索, 各个单一线索的去信息掩蔽作用得到了证实, 但其作用机理依然是未解之谜。探究其加工机理, 总结出人类的中枢系统自上而下运用这些线索的规律是破解“鸡尾酒会”问题的关键。今后对这个重大的问题的研究应当着重探索以下几个问题:

(1) 仅仅集中于自下而上加工的过程无法实现对“鸡尾酒会”问题的深入理解。对人类系统化知觉机制的深入研究是解决这个问题的基础, 对其中各个知觉线索的高级整合机能的研究, 是对“鸡尾酒会”研究的最高层次。勾勒出各个线索之间相互关系的蓝图, 总结出中枢系统在不同情境下灵活运用不同线索的规律, “鸡尾酒会”问题会迎刃而解。因此, 尽管目前的研究状态还基本属于对各个可能的线索进行分析和实证, 但应当积极开展“跨线索”的实验研究。

(2) 听者在“鸡尾酒会”环境中对语句的识别需要多项知觉能力的集合, 知觉能力之间存在交互作用, 探测知觉能力之间的相互关系是研究方向之一。例如, 近期的研究也已经表明, 处于知觉加工低层的对直达声和反射声的知觉整合的能力与处于认知加工高层的去信息掩蔽的能力之间有显著的相关 (Li, Huang, Huang, Chen & Wu, 2008)。这种“跨

加工层次”的研究, 也是认识“鸡尾酒会”问题的关键。

(3) “鸡尾酒会”问题应当有相应的语言和文化色彩。研究者的眼光不能只局限于心理学和神经认知科学。研究不同语言和文化的背景在信息掩蔽中的作用可以极大地促进相关学科的交流 and 融合。

(4) 尽管目前还没有任何一个计算机言语识别系统具有人的利用知觉线索来消除信息掩蔽的能力, 但随着信息科学、计算机科学与心理学和神经科学的交叉, 新一代的语言识别系统会逐步发展出人的去信息掩蔽的能力。因此, 研究“鸡尾酒会”问题会带来计算机科学和信息科学发展的一个飞跃。

(5) 对“鸡尾酒会”环境的适应体现了人脑的最高级的功能。研究各种感觉系统的疾病以及精神疾病如何影响这个高级功能也是深入认识这些疾病的一个重要渠道。因此, 今后应当强化有关“鸡尾酒会”的基础研究与神经系统疾病研究的结合。如可以利用精神病人作为被试, 以探测“鸡尾酒会”问题为手段, 折射病人与常人的高级认知功能的差别所在。

(6) 最后, 有关人脑利用线索来消除信息掩蔽的神经机制的研究刚刚开始, 希望更多的科学家重视和参与相关的神经机制的研究。例如已有研究结果表明, 听觉皮层的信号加工也出现了背侧的空间信息通路与腹侧的内容信息通路的分离, 而快速的空间信息通路到达前额叶后又对内容信息通路进行自上而下的调节 (Wang, Wu & Li, 2008)。这种自上而下的调节作用是否在主观空间分离的去信息掩蔽中起关键的作用?

参考文献

- Arbogast, T. L., Mason, C. R., & Kidd, G. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 117, 2169-2180.
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America*, 109, 1101-1109.
- 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. *Journal of the Acoustical Society of America*, 110, 2527-2538.
- Cherry, E. C. (1953). Some experiments on the recognition of speech with one and two ears. *Journal of the Acoustical Society of America*, 25, 975-979.

- Darwina, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *Journal of the Acoustical Society of America*, *114*, 2913–2922.
- Drullman, R., & Bronkhorst, A. W. (2004). Speech perception and talker segregation: Effects of level, pitch, and tactile support with multiple simultaneous talkers. *Journal of the Acoustical Society of America*, *116*, 3090–3098.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *Journal of the Acoustical Society of America*, *109*, 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. *Journal of the Acoustical Society of America*, *115*, 2246–2256.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2006). Variability and uncertainty in masking by competing speech. *Journal of the Acoustical Society of America*, *121*, 1040–1046.
- Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustical Society of America*, *106*, 3578–3588.
- Helfer, K. S., & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *Journal of the Acoustical Society of America*, *117*, 842–849.
- Huang, Y., Huang, Q., Chen, X., Qu, T. S., Wu, X. H., & Li, L. (2008). Perceptual integration between target speech and target-speech reflection reduces masking for target-speech recognition in younger adults and older adults. *Hearing Research*, *244*, 51–65.
- Kidd, G., Mason, C. R., & Arbogast, T. L. (2002). Similarity, uncertainty and masking in the identification of nonspeech auditory patterns. *Journal of the Acoustical Society of America*, *111*, 1367–1376.
- Kidd, G., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *Journal of the Acoustical Society of America*, *118*, 3804–3815.
- Li, L., Daneman, M., Qi, J. Q., & Schneider, B. A. (2004). Does the information content of an irrelevant source differentially affect speech recognition in younger and older adults? *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 1077–1091.
- Li, L., Huang, Y., Huang, Q., Chen, X., & Wu, X. H. (2008). The ability of temporally integrating acoustic waveforms is associated with release of speech from informational masking under reverberant conditions. *Journal of the Acoustical Society of America*, *123*, 3301.
- Li, L., Qi, J. G., He, Y., Alain, C., & Schneider, B. A. (2005). Attribute capture in the precedence effect for long-duration noise sounds. *Hearing Research*, *202*, 235–247.
- Li, L., & Yue, Q. (2002). Auditory gating processes and binaural inhibition in the inferior colliculus. *Hearing Research*, *168*, 113–124.
- Nathaniel, I. D., Christine, R. M., & Barbara, G. S. (2003). Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *Journal of the Acoustical Society of America*, *114*, 368–379.
- Noble, W., & Perrett, S. (2002). Hearing speech against spatially separate competing speech versus competing noise. *Perception and Psychophysics*, *64*, 1325–1336.
- Oxenham, A. J., Fligor, B. J., Mason, C. R., & Kidd, G. (2003). Informational masking and musical training. *Journal of the Acoustical Society of America*, *114*, 1543–1549.
- Rudmann, D. S., McCarley, J. S., & Kramer, A. F. (2003). Bimodal displays improve speech comprehension in environments with multiple speakers. *Human Factors*, *45*, 329–336.
- Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, & Larson, E. (2005). Top-down and bottom-up influences on spatial unmasking. *Acta Acustica united with Acustica*, *91*, 967–979.
- Summers, V., & Molis, M. R. (2004). Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level. *Hearing Research*, *47*, 245–256.
- Virginia, B., Erol, O., Frederick, J. G., Kamal, S., & Barbara, G. S. (2005). Spatial unmasking of birdsong in human listeners: energetic and informational factors. *Journal of the Acoustical Society of America*, *118*, 3766–3773.
- Wang, M. Y., Li, J. Y., Huang, Y., Wu, Y. H., Wu, X. H., & Li, L. (2008). Speech-synchronized visual cues release speech from informational masking. *Journal of the Acoustical Society of America*, *123*, 3302.
- Wang, W. J., Wu, X. H., & Li, L. (2008). The dual-pathway model of auditory signal processing. *Neuroscience Bulletin*, *24*, 173–182.
- Watson, C. (2005). Some comments on informational masking. *Acta Acustica united with Acustica*, *91*, 502–512.
- Wu, X. H., Wang, C., Chen, J., Qu, H. W., Li, W. R., Wu, Y. H., et al. (2005). The effect of perceived spatial separation on informational masking of Chinese speech. *Hearing Research*, *199*, 1–10.
- Xu, L. J., Li, J. Y., Wu, X. H., & Li, L. (2008). Modulation of the voice-cuing effect on releasing speech from informational masking. *Journal of the Acoustical Society of America*, *123*, 3864.
- Yang, Z. G., Chen, J., Huang, Q., Wu, X. H., Wu, Y. H., Schneider, B. A., et al. (2007). The effect of voice cuing on releasing Chinese speech from informational masking. *Speech Communication*, *49*, 292–304.
- Yang, Z. G., Chen, J., Huang, Q., Huang, Y., Wu, X. H., Wu, Y. H., et al. (2008). Both prime length and prime position affect the

- spatial release of nonsense speech from informational masking. *Journal of the Acoustical Society of America*, 123, 3295.
- Zurek, P. M. (1993). Binaural advantages and directional effects in speech intelligibility. *Acoustical Factors Affecting Hearing Aid Performance*, 255–276.
- Zurek, P. M., Freyman, R. L., & Balakrishnan, U. (2004). Auditory target detection in reverberation. *Journal of the Acoustical Society of America*, 115, 1609–1620.

Effects of Perceptual Cues on Releasing Speech From Informational Masking

XU Li-Juan; HUANG Ying; WU Xi-Hong; WU Yan-Hong; LI Liang

(Department of Psychology, Speech and Hearing Research Center, Peking University, Beijing 100871, China)

Abstract: In a noisy “cocktail-party” environment, target speech is masked by different sounds. The masking includes at least two different components: energetic masking and informational masking. Energetic masking occurs when peripheral neural activities elicited by the target signal are overwhelmed by those elicited by the mask signal, leading to a substantially degraded neural representation of the target signal. Informational masking occurs when the target and the masker are similar in some dimensions, such as when they are both speech signals. The similarity leads to a competition for the limited central cognitive processing resources between the target and the masker signal. Thus, the amount of informational masking is not only affected by the bottom-up processes but also modulated by the top-down processes. The effects, which occur at the perceptual level, of the perceived spatial separation, the visual cues such as lip reading and speech-synchronized light flash, and the familiarity with the characteristics of the target signal on releasing target speech from informational masking are summarized. Future research will focus on the interactive effects among these cues and the perceptual processing mechanisms of using different cues to release speech from informational masking.

Key words: “cocktail-party” problem; perceptual cues; selective attention; energetic masking; informational masking