

主观空间分离下的汉语信息掩蔽效应*

吴艳红^{1,4} 李文瑞¹ 陈 婧² 王 纯¹
屈宏伟² 吴玺宏² 李 量^{1,2,3}

(1 北京大学心理学系 北京 100871)

(2 北京大学言语与听觉实验室 北京 100871)

(3 多伦多大学心理学系 加拿大多伦多 L5L1C6)

(4 首都师范大学学习与认知实验室 北京 100037)

2004年6月30日收到

2004年11月24日定稿

摘要 基于听觉优先效应中的融合现象,探讨了主观空间分离下的汉语信息掩蔽效应。实验用左右两个扬声器来播放目标言语信号和掩蔽声音,并通过改变两个扬声器之间的延迟来操作掩蔽声音的主观空间位置。结果显示,尽管言语信号和掩蔽声音都由同样的扬声器播放而没有实际的空间分离,这种利用优先效应所产生的主观空间分离却可以提高言语识别的正确率。在信息掩蔽条件下由主观空间分离所造成的言语识别的改善显著地高于在能量掩蔽条件下的改善。这些结果为如何分离对言语信号的能量掩蔽与信息掩蔽,以及为相关的建筑声学 and 通讯技术的研究与应用提供了听觉心理学的参考。

PACS 数: 43.66

Informational masking of Chinese speech under perceived spatial separation

WU Yanhong^{1,4} LI Wenrui¹ CHEN Jing² WANG Chun¹QU Hongwei² WU Xihong² LI Liang^{1,2,3}

(1 Department of Psychology, Peking University Beijing 100871)

(2 Speech and Hearing Research Center, Peking University Beijing 100871)

(3 Department of Psychology, University of Toronto at Mississauga, Toronto L5L1C6)

(4 Learning & Cognition Lab, Capital Normal University Beijing 100037)

Received Jun. 30, 2004

Revised Nov. 24, 2004

Abstract This study investigated the effects of perceived spatial separation induced by precedence effect on Chinese speech recognition with informational masking. In the experiment, two loudspeakers which located in left and right respectively were used to play the target and masking, and the delay between the two loudspeakers was changed to produce different perceived spatial location of the masking. The results suggest that although the target and masking are sent from the same loudspeakers, the perceived spatial separation can improve the speech recognition. Under informational masking, the improvements are significant greater than under energetic masking. These results are useful in the separation of energetic and informational masking of speech.

引言

在人们日常生活和工作环境中,对言语信号的认知往往要受到各种干扰声音的掩蔽作用。这些干扰声音影响了对言语信号源特征的分析,从而降低了对言

语信号内容的识别。从掩蔽的机制上来看,干扰声音的掩蔽作用可以大致分为两类:能量掩蔽(energetic masking)和信息掩蔽(informational masking)^[1-7]。能量掩蔽的产生主要发生在听觉系统的外周部分。当干扰声音和目标声音同时出现时,两个声音不仅

* 国家自然科学基金(30370476, 30470568)、国家科技部重大基础研究前期研究专项项目(2002CCA01000)和首都师范大学学习与认知实验室资助项目

在外耳处要有声学物理上的交互作用,而且都要经过内耳的滤波分析。特别是当干扰声音和目标声音在频率上发生重叠时,听觉神经系统同时进行对目标声音频率的分析和对干扰声音的分析,因而听觉神经系统对目标声音的动态反应发生下降,一个很明显的表现就是,在干扰声音的背景下,对目标声音觉察和辨认需要更高的信噪比。清晰度理论(Articulation Theory)认为,对言语信号的能量掩蔽的大小完全依赖于目标声音和干扰声音之间声谱重叠的情况,对目标言语识别的正确率会随着信噪比的下降而线性降低^[4]。

除了能量掩蔽外,研究者还发现在声音识别中存在着另外一种掩蔽作用,即信息掩蔽。当干扰声音和目标声音信号的信息模式有一定程度的相似时,例如两者都是言语,它们之间就会出现心理加工层次上的混淆,对信号进行加工的一部分神经和心理资源就会被用于加工干扰声音,也就是说目标信号受到了掩蔽作用。信息掩蔽是一种在较高的神经和心理加工水平上所产生的掩蔽作用,即使同一时刻目标声音和干扰声音在频率上并没有重合,这种掩蔽作用也能出现^[4]。信息掩蔽概念的提出有助于解释为什么有时候目标刺激激活了外周神经系统但仍然不能被觉察^[8]。

干扰声音对目标声音信号的干扰是显而易见的,但Hirsh等人早就发现(1950),当干扰声音和目标声音来自不同的空间位置时,人们对目标声音的正确识别率会有很大的改善^[9]。在心理学领域中一个反映言语信号和干扰声音的空间分离对言语信号识别影响的著名例子就是所谓的“鸡尾酒会效应”(cocktail party effect)^[10]。这种空间分离的作用已被近期的心理物理学研究所确证^[1.3,11-13]。例如,Freyman等人(1999)的研究表明当干扰声音为语谱噪音时,空间分离所造成的对目标语句言语识别的阈限可以降低8.2 dB;而当干扰声音为言语时,识别阈限的降低可达到12 dB^[1]。

Zurek(1993)的双耳言语识别模型(Zurek's model of binaural speech recognition)试图对由于目标声音信号和干扰声音的空间分离所造成的言语识别的改善进行理论解释^[14]。这个模型认为,言语识别的改善是因为目标信号和干扰声音的分离降低了掩蔽水平。对于高频声音,由于头影效应(head shadow)的作用,离干扰声音较远的那只耳朵的信噪比会提高,从而降低了掩蔽水平和提高了言语识别的成绩;而对于低频声音,由于双耳的交互作用(binaural interaction),不同位置的声音会产生不同的双耳间时间延

迟,因此双耳听觉系统利用目标信号与干扰声音在双耳之间的时间差的不同而使掩蔽的效果变弱,从而提高言语识别的正确率^[1]。显而易见,Zurek的模型针对空间分离对能量掩蔽的影响进行了解释。但是,因为回音的存在会同时破坏头影效应和双耳交互作用的效果,该模型无法解释在有回音的声学环境中目标言语信号与干扰声音的空间分离为什么能改善对目标言语信号的识别。

在有回音的声学环境中,人耳不仅接收直接来自声源的声波,还接收多个来自不同方向的与该声源相关的反射波。如果直接来自声源的声波与反射波到达人耳的时间延迟足够小(1至10 ms或更长,取决于声音的性质),人们所感受到的只是一个融合而成的来自声源位置的声音,而不是来自不同方向的独立的听觉事件。这一现象被称为优先效应(the precedence effect)^[15-19]。优先效应的形成有助于减小反射波对声源空间信息加工的干扰。在实验室中,优先效应一般用两个在空间上分离的扬声器来模拟,用其中的一个领先扬声器发出的声音来模拟直接来自声源的声波,而用另一个落后扬声器发出的延迟的相关声音来模拟声源的反射波,进而可以对两个声音之间的相关性以及延迟对优先效应的影响进行系统的研究。尤其重要的是,利用对优先效应的模拟可以产生不同的声音信号之间的主观空间分离。例如,当左右两个扬声器都发出A、B两个声音时,对A声音来讲,左扬声器提前3 ms,而对B声音来讲,右扬声器提前3 ms。根据优先效应,两个扬声器发出的A声音会有融合,其声音影像来自左扬声器的方向。同样,两个扬声器发出的B声音也会有融合,其声音影像来自右扬声器的方向。这样A、B两个声音就会出现主观空间分离。但客观上,A、B两个声音的声源空间位置并没有分离,都来自同样的扬声器。

Freyman等人(1999,2001)指出,当干扰声音是言语时,利用优先效应所产生的主观空间分离可以显著地提高对目标言语的识别;但是当干扰声音是语谱噪音时,这种主观空间分离的作用却很不明显^[1,2]。因而Freyman等人认为,由优先效应所产生的主观空间分离是一种将能量掩蔽和信息掩蔽分开的方法。利用这种方法,能量掩蔽和信息掩蔽的性质可以得到详细的研究。

本研究将从几个方面扩展Freyman等人的研究。(1)尽管汉语是世界上使用者最多的语言,但有关汉语的能量掩蔽与信息掩蔽的差异的研究还很少,特别是还没有正式的关于主观空间分离对这两种掩蔽的影响的报告。汉语音节中的元音与英语和

其它的欧洲语言音节中的元音不同,其音调是有词义的。另外,汉语的清辅音出现的频率较英语的高。有研究表明在实验室安静环境下,不经过说话者自适应,汉语音素的识别率为 51.4%,而英语单词的识别率却可达到 90% 左右^[20]。因此,汉语的信息或能量掩蔽的特征也可能与英语的信息或能量掩蔽的特征有所不同。本研究采用汉语语句作为目标语句和掩蔽语句。(2) 在 Freyman 等人的研究中,其主观空间分离的程度只有一种,即在右侧与正前方之间的半感受野(semifield)内部进行分离,仅仅在“信号中间-干扰中间”和“信号中间-干扰右侧”之间进行比较。本实验采用半感受野内部和跨两个半感受野两种程度的主观空间分离,在“信号右侧-干扰右侧”、“信号右侧-干扰中间”和“信号右侧-干扰左侧”之间进行比较。

1 方法

1.1 被试

被试为 12 名(男 6 名,女 6 名)母语为汉语普通话并且听力正常的北京大学学生。年龄在 19 岁到 25 岁之间。把 12 名被试随机分成 A、B 两组,每组中男女各 3 名。实验中,A 组被试第一日进行语谱噪音干扰条件的实验,第二日进行语句干扰条件的实验;B 组被试反之。

1.2 声音材料

1.2.1 目标信号

目标信号为 312 个无意义句子。这些句子都是“修饰语+主语+修饰语+谓语+修饰语+宾语”的形式。句子符合语法规则,但不具有意义(例:你的融合正在归属他的强调)。在每个句子中,主语、谓语和宾语三个词语为关键词(句中画线的词)。所有目标句子被随机分为 24 组,每组含 13 个句子。女大学生甲用普通话以正常语速朗读所有句子,通过 KENYAK K-276 传声器并利用 cooledit 软件录入计算机中形成 .wav 文件。然后使用 Matlab 软件对句子进行处理,加入相应的干扰刺激,生成满足实验条件的双声道声音文件(.wav 格式)。

1.2.2 掩蔽声音

语谱噪声是从“863”语音库中随机选取 10 位年轻女声的语句文件(每位 500 个句子)并进行叠加平均,继而得到稳定的语谱噪音,作为含有能量掩蔽成分的噪音干扰刺激。

干扰语句类似目标语句,为语法正确的无意义

句子,但由两个女大学生(乙和丙)用普通话以正常语速朗读。其中乙和丙两人从不同时朗诵同一个句子。录音方法同目标刺激。干扰语句作为含有信息掩蔽成分的干扰刺激。

1.3 仪器和环境

实验在一个 $192 \times 181 \times 194 \text{ cm}^3$ 的屏蔽室(EMI Shielded Audiometric Examination Acoustic Suite, Texas, USA)中进行,被试坐在屏蔽室一角的椅子正中,前方左右两侧 45° 的地方各有一个扬声器(创通牌)。两扬声器中心的高度为 90 cm,与被试双肩在同一水平面;两扬声器距离被试身体中心的水平距离都是 135 cm。

一名主试坐在被试的正对面,使用北京大学自制的反应键盘按键启动实验,同时记录被试的反应;另一名主试在屏蔽室隔壁房间用计算机(Pentium IV, CREATIVE 24 bit/96 kHz/100 dB SNR 声卡)操纵实验程序。

实验时单扬声器的目标语句的音量(声级)保持在 54 dB,由一个 B&K 声级计(Type 2230)所监测。

1.4 实验设计

本实验采用 $2 \times 3 \times 4$ 组内设计,自变量为掩蔽类型(噪音、语句)、掩蔽位置(左、中、右)和信噪比(-12, -8, -4, 0 dB)。

实验时两个扬声器都播放目标语句,但右侧扬声器总是比左侧扬声器提早 3 ms。根据优先效应,被试总感觉到目标语句从右侧扬声器传出。掩蔽声音也是这两个扬声器都播放。(1) 在“左侧掩蔽”条件下,左侧扬声器比右侧扬声器早播放 3 ms,此时被试感到掩蔽声音来自左侧扬声器,(2) 在“中间掩蔽”条件下,两个扬声器同时播放,此时被试感到掩蔽声音来自正前方,(3) 在“右侧掩蔽”条件下,左侧扬声器比右侧扬声器晚播放 3 ms,此时被试感到掩蔽声音来自右侧扬声器。图 1 表明这三种掩蔽声音的主观空间位置。

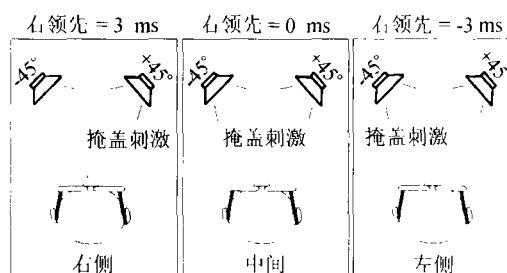


图 1 掩蔽刺激的呈现方式和主观位置

在所有实验条件下,目标语句的启动总是比掩蔽声音的启动晚 1000 ms,目标语句与掩蔽声音同时

停止。

因变量为言语识别的正确率。每个目标句子播放完之后,所有声音停止,被试复述出刚才听到的句子。用复述出的关键词占全部关键词的百分比表示识别的正确率。

2 结果

噪音掩蔽条件下(图2左),对掩蔽声音的主观位置(简称掩蔽位置)和信噪比做 3×4 组内方差分析。结果显示,掩蔽位置和信噪比之间交互作用显著, $F(6,66) = 11.95, P < 0.001$;信噪比和掩蔽位置的主效应都显著, $F_{SNR}(3,33) = 331.91, P_{SNR} < 0.001, F_{位置}(2,22) = 7.57, P_{位置} = 0.03$ 。

对不同信噪比时的正确率进行进一步分析,可以看到,四种信噪比条件下掩蔽位置对言语识别的正确率的影响也都显著。(1) -12 dB: $F(2,22) = 5.10, P = 0.015$, (2) -8 dB: $F(2,22) = 15.71, P < 0.001$, (3) -4 dB: $F(2,22) = 12.54, P < 0.001$, (4) 0 dB: $F(2,22) = 4.86, P = 0.018$ 。在信噪比为 -12 和 -4 时,掩蔽位置在右侧时言语识别的正确率显著低于干扰刺激在左侧或在中间;当信噪比为 -8 时,掩蔽位置在中间时的正确率显著高于其他两种条件时,而掩蔽位置在右侧时和在左侧的正确率没有显著差异;当信噪比为 0 时,掩蔽位置在右侧和中间的正确率没有显著差异,而且都显著高于干扰刺激在左侧时的正确率。

在语句干扰条件下(图2右),对掩蔽位置和信噪比做 3×4 组内方差分析。结果显示,掩蔽位置和信噪比之间的交互作用显著, $F(6,66) = 5.27, P < 0.001$;信噪比和干扰位置的主效应都显著, $F_{SNR}(3,33) = 263.72, P_{SNR} < 0.001, F_{位置}(2,22) = 21.46, P_{位置} < 0.001$ 。

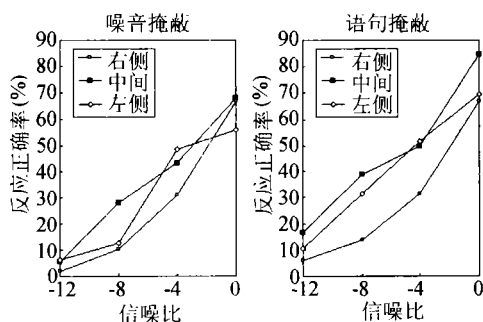


图2 不同掩蔽条件下言语识别的正确率

对不同信噪比时的正确率进行进一步分析,可以看到,四种信噪比的条件下,掩蔽位置对被试言语识别正确率的影响都显著:(1) -12 dB: $F(2,22) = 7.46,$

$P = 0.003$, (2) -8 dB: $F(2,22) = 21.41, P < 0.001$, (3) -4 dB: $F(2,22) = 17.17, P < 0.001$, (4) 0 dB: $F(2,22) = 7.25, P = 0.004$ 。根据事后检验的结果,在信噪比非 0 的3种条件下,掩蔽位置在右侧时言语识别的正确率都显著低于干扰刺激在左侧或在中间时的正确率。信噪比为 0 时,掩蔽位置在中间时的正确率显著高于其他两种条件;掩蔽位置在右侧和在左侧的正确率没有显著差异。

为比较主观空间分离对噪音掩蔽和语句掩蔽的影响,我们将图2中两个相应主观空间分离的“左侧”曲线和“中央”曲线分别与没有主观空间分离的“右侧”曲线之间的面积定义为主观空间分离所造成的语句辨认行为的增益,结果如图3所示。从图3可以看到,当掩蔽声音为语句时,主观空间分离所造成的对目标语句辨认行为的增益比当掩蔽声音为噪音时的显著地高。另外,当掩蔽声音的主观空间位置在正前方时,主观空间分离所造成的对目标语句辨认行为的增益要显著地比掩蔽声音在左侧时的高。

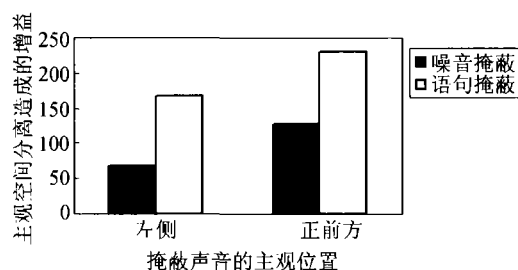


图3 主观空间分离造成的语句辨别增益

3 讨论

本实验的结果表明,当声音语料是汉语时,在各种实验条件下,掩蔽声音的掩蔽效果受信噪比影响,即随着信噪比的提高,掩蔽的效果减小。因此,能量掩蔽和信息掩蔽都能够影响对汉语目标语句的识别成绩,从而改变信噪比的阈限。

当掩蔽刺激由两个在空间对称的扬声器播放时,根据两个扬声器之间的三种不同的时间延迟,被试形成的掩蔽刺激的影像分别来自左边、中间或右边,这种声音的主观空间位置由优先效应所引起,并因此产生了目标语句与掩蔽刺激之间的主观空间汇聚或主观空间分离。然而,这种主观空间位置的变化并没有引入头影效应以及去掩蔽的双耳交互作用。

本实验重要的结果是,尽管掩蔽声音与目标语句没有实际的分离,当利用优先效应现象将掩蔽声音与目标语句信号进行主观空间分离时,被试的言语识别正确率有显著的提高。但主观空间分离所造

成的去信息掩蔽(语句掩蔽)的效果要比去能量掩蔽(噪音掩蔽)的效果高。这个结果基本符合 Freyman 等人(1999)用英文语料所得到的研究结果,说明由优先效应引起的主观空间分离所造成的去掩蔽作用具有跨语言性。

当掩蔽刺激是平稳的语谱噪音时,该掩蔽刺激没有产生信息掩蔽,主观空间分离只对被试提供了较小的去掩蔽的帮助。但当掩蔽刺激也是语句时,掩蔽语句与目标语句很相似,进而产生了语句对语句的信息掩蔽。主观空间分离提供了显著的线索来帮助被试在目标刺激与掩蔽刺激之间产生主观分离(segregation),从而显著地减少了信息掩蔽。

本实验的结果也表明,主观分离的效果并不是分离的程度越大去掩蔽的效果也越大。在本实验条件下,中间相对于右侧的主观分离的效果要比左侧相对于右侧的主观分离的效果大一些。为什么不同的半感受野的主观分离效果并不比相同半感受野的分离效果好,还需要今后进一步的实验研究来解释。

在 0 信噪比下所出现的结果的可能原因有以下几点:第一,在高信噪比的情况下,噪音或语句的掩蔽作用相对比较小,被试的正确反应率比较高,因此在右边的主观空间位置下的掩蔽效果与左边或中间的差异就有可能很小。第二,尽管在高信噪比的情况下掩蔽的作用很小,左边出现的噪音影像有可能会引导某些被试在这一部分实验中将注意力放到左边,因而造成了在左边的噪音掩蔽情况下的平均反应成绩比中间和右边低的结果。然而,我们目前还不清楚为什么在语句掩蔽的条件下,中间掩蔽情况下的成绩最好。

在本实验的控制条件下,目标语句与掩蔽噪声之间没有实际的空间分离,因此基本上不存在头影效应和去掩蔽的双耳交互作用。但人的听觉系统却能够通过优先效应中声音的心理融合现象所造成的目标语句与掩蔽噪声之间的主观空间分离,来达到显著的去掩蔽作用。这反映了一个重要的人脑对相关声音进行编组加工的功能。这种功能有助于在有回音的声学环境中提高对目标声音事件中的信息的获取。当前对这种编组加工的高级功能的认识还很不充分,有待于进一步的研究。毋庸置疑,这个方向的研究在心理声学(Psychoacoustics)中有重要的科学价值。

信号和噪声的主观空间分离所造成的去掩蔽现象在实际的应用中应当受到重视。例如飞行员从耳机所接受到的各种听觉信号如果能经过处理而造成主观空间分离的效果,飞行员就会从众多的信号源

中更有效地选择和分析最重要的信号。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它来源的语句信号,这就对发展更加先进的听觉通讯系统提出了迫切的要求。优先效应所造成的主观分离,可以将由同一对耳机发出的多个不同来源的语句信号安排到不同的主观“空间”位置上。本研究的结果说明这种主观的空间安排会有助于(1)飞行员对不同来源信号的主动选择,(2)目标语句选定后,其它信号来源的干扰作用会减少。

4 结论

(1) 由优先效应所引起的主观空间分离可以提高目标语句识别的正确率。

(2) 这种主观空间分离的去信息掩蔽的效果要高于去能量掩蔽的效果。

参 考 文 献

- 1 Freyman R L, Helfer K S, McCall D D, Clifton R K. The role of perceived spatial separation in the unmasking of speech. *J. Acoust. Soc. Am.*, 1999; **106**(6): 3578—3588
- 2 Freyman R L, Balakrishnan U, Helfer K. S. Spatial release from informational masking in speech recognition. *J. Acoust. Soc. Am.*, 2001; **109**(5): 2112—2122
- 3 Arbogast T L, Mason C R, Kidd Jr G. The effect of spatial separation on informational and energetic masking of speech. *J. Acoust. Soc. Am.*, 2002; **112**(5): 2086—2098
- 4 Brungart D S. Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.*, 2001; **109**(3): 1101—1109
- 5 Brungart D S, Simpson B D. The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. *J. Acoust. Soc. Am.*, 2002; **112**(2): 664—676
- 6 Kidd G, Mason C R, Deliwala P S, Woods W S, Colburn H S. Reducing informational masking by sound segregation. *J. Acoust. Soc. Am.*, 1994; **95**(6): 3475—3480
- 7 Kidd G, Jr, Mason C R, Rohtla T L, Deliwala P S. Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *J. Acoust. Soc. Am.*, 1998; **104**(1): 422—431
- 8 Richards V M, Tang Z, Kidd Jr G. Informational masking with small set sizes. *J. Acoust. Soc. Am.*, 2002; **111**(3): 1359—1366

- 9 Hirsh I J. The relation between localization and intelligibility. *J. Acoust. Soc. Am.*, 1950; **22**(2): 196—200
- 10 Cherry E C. Some experiments on the recognition of speech with one and two ears. *J. Acoust. Soc. Am.*, 1953; **25**(5): 975—979
- 11 Dubno J R, Ahlstrom J B, Horwitz A R. Spectral contributions to the benefit from spatial separation of speech and noise. *J. Sp. Lan. Hear. Res.*, 2002; **45**(6): 1297—1310
- 12 Duquesnoy A J. Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *J. Acoust. Soc. Am.*, 1983; **74**(3): 739—743
- 13 Gelfand S A, Ross L, Miller S. Sentence reception in noise from one versus two sources: Effects of aging and hearing loss. *J. Acoust. Soc. Am.*, 1988; **83**(1): 248—256
- 14 Zurek P M. Binaural advantages and directional effects in speech intelligibility. In: Studebaker, G. A., Hockberg. I. (Eds), *Acoustic factors affecting hearing aid performance*. Allyn and Bacon, Boston. 1993
- 15 Wallach H, Newman E B, Rosenzweig M R. The precedence effect in sound localization. *Am. J. Psychol.*, 1949; **62**(3): 315—336
- 16 Blauert J. *Spatial hearing*. MIT, Cambridge, Massachusetts, 1997
- 17 Li L, Yue Q. Auditory gating processes and binaural inhibition in the inferior colliculus. *Hear. Res.*, 2002; **168**(1-2): 98—109
- 18 Litovsky R Y, Colburn H S, Yost W A, Guzman S J. The precedence effect. *J. Acoust. Soc. Am.*, 1999; **106**(4): 1633—1654
- 19 Zurek P M. The precedence effect and its possible role in the avoidance of interaural ambiguities. *J. Acoust. Soc. Am.*, 1980; **67**(3): 952—964
- 20 蒋文建, 韦 岗. 基于掩蔽特性的噪音环境下语音识别新特征. *声学学报*, 2001; **26**(6): 516—520