



虚拟分离声源空间去掩蔽的初步研究

谢志文*, 金晶

华南理工大学物理科学与技术学院, 广州 510640

* E-mail: phzhwxie@scut.edu.cn

收稿日期: 2007-08-17; 接受日期: 2008-03-20

国家自然科学基金资助项目(批准号: 10374031)

摘要 提出一种使用耳机虚拟声重发的实验方法, 研究了掩蔽和被掩蔽声源在水平面空间分离时的前向掩蔽效应, 并将此情况下的掩蔽值与声源在同一位置时的掩蔽值进行了比较. 结果表明, 虽然在 0° 和 $\pm 30^\circ$ 声像位置, 掩蔽阈值均随掩蔽声的声压级增加而增加, 但确实观察到空间去掩蔽现象. 最大的去掩蔽值高达 15 dB. 这种空间去掩蔽现象可主要归结于较优耳的贡献.

关键词

空间去掩蔽
虚拟声源
前向掩蔽

当掩蔽与被掩蔽声源在空间上分离时, 掩蔽阈值下降的现象称为空间去掩蔽. 有许多研究者利用不同的信号进行了空间去掩蔽研究, 所使用的信号包括纯音^[1]、语言^[2]、脉冲串^[3]和复合音等^[4]. 这些实验的结果并不完全一致, 空间去掩蔽的值有的只有 7~8 dB^[5], 而高的可达 24 dB^[6].

对掩蔽声为宽带噪声的情况, 空间去掩蔽主要来自于较优耳效应和双耳效应. 较优耳效应之所以会导致去掩蔽的原因, 是因为当掩蔽声和被掩蔽声在空间分离时, 某个耳朵处的信掩比(signal-to-masker ratio, SMR)与掩蔽声和被掩蔽声在同一位置时相比会增加. 而双耳效应会导致去掩蔽的原因, 是因为掩蔽声和被掩蔽声分离时, 会产生双耳时间差(interaural time difference, ITD)和双耳强度差(interaural level difference, ILD).

但是, 目前的大多数空间去掩蔽研究都是针对同时掩蔽, 而且声源主要是扬声器. 前向掩蔽与同时掩蔽具有一些不同的特性, 例如, 在同时掩蔽中, 起码对宽带掩蔽声来说, 阈值通常对应于一个恒定的信掩比, 而对前向掩蔽, 掩蔽声声压级的一个增量并不产生一个等量的阈值增量^[7]. 此外, 对前向掩蔽来

说, 掩蔽声与被掩蔽声之间的时间间隔很短, 从 0 到几毫秒之间, 掩蔽声和被掩蔽声本身的持续时间也很短^[8,9]. 这就导致了使用扬声器来研究前向掩蔽在某些时候并不是一个很好的选择, 因为扬声器本身的过渡过程通常较长. 但要注意的是, 使用耳机来研究前向掩蔽, 尽量减少和消除头中定位效应是很重要的^[10,11].

在本论文中, 我们研究了以纯音信号作为被掩蔽信号, 窄带噪声作为掩蔽声时前向掩蔽的空间去掩蔽现象. 提出了一种使用耳机的虚拟声重发方法, 应用了个性化的头传输函数(head-related transfer functions, HRTF)来虚拟声源. 这种方法既给倾听者呈现了真实的声学线索又可以仔细地分析实验时到达倾听者耳朵的信号. 掩蔽声与被掩蔽声之间的时间间隔为 0 ms. 通过使用这种虚拟声重发方法, 最后不仅达到了主观感觉上声像分离的良好效果, 而且也得到了一些不同于使用扬声器作为声源的实验结果.

1 声源空间分离对掩蔽阈值的影响

1.1 水平空间声像的虚拟

人类听觉系统对声源的定位, 主要是根据人的

头部、耳廓以及躯干等对入射声波的散射作用, 使得声音到达双耳时会形成双耳声级差和双耳时间差. 人的听觉系统和高层神经系统根据这些信息并和过去的经验比较, 从而决定声源的方向. 人头及耳廓等对声波的散射作用以及相应的 ITD 和 ILD 等可用头相关传输函数 HRTF 来表述和计算. 如果声源的水平角为 θ , 则水平面上的左和右头传输函数定义为

$$\begin{aligned} H_L(r, \theta, \omega) &= P_L(r, \theta, \omega)/P_0(r, \omega), \\ H_R(r, \theta, \omega) &= P_R(r, \theta, \omega)/P_0(r, \omega), \end{aligned} \quad (1)$$

其中 P_L 和 P_R 为左右耳处的复数声压, P_0 为倾听者不在时头中心处的自由场声压, r 为声源到头中心处的距离, ω 为声源信号的角频率.

由于 HRTF 包含了声音定位的信息, 因而利用其进行信号处理, 就可以实现空间声像的虚拟. 设有原始声频信号 $E_0(\omega)$, 经信号处理后, 左右耳处的信号分别为

$$E_L(\omega) = H_L(r, \theta, \omega)E_0(\omega), \quad (2)$$

$$E_R(\omega) = H_R(r, \theta, \omega)E_0(\omega). \quad (3)$$

如果再用耳机重发, 此时双耳处的声压将正比于(或等于) θ 角方向的真实声源产生的声压, 从而在主观感觉上虚拟出水平 θ 角方向的声像.

与以往大多数用于信号处理的统计性 HRTF 数据不同, 本实验中我们采用个性化的 HRTF 数据来进行信号处理(即对听音者采用其自身的 HRTF 数据来处理他将听到的声音信号), 这样可以一定程度上避免由于个人的生理差异造成的 HRTF 数据不适用现象, 实现更加逼真的水平空间声像虚拟. 使用个性化 HRTF 处理过后的信号为

$$E'_L(\omega) = H'_L(r, \theta, \omega)E_0(\omega), \quad (4)$$

$$E'_R(\omega) = H'_R(r, \theta, \omega)E_0(\omega). \quad (5)$$

其中个性化的 HRTF 是经过均衡处理的, 排除了扬声器-传声器频率响应的影响, 即:

$$H'(r, \theta, \omega) = \frac{E'_{\text{测}}(r, \theta, \omega)}{H_{\text{扬-传}}}, \quad (6)$$

$H_{\text{扬-传}}$ 表示扬声器-传声器的频率响应. 由于是虚拟自由场中扬声器重发下的双耳处的信号, 所以我们希望双耳处的信号为

$$E_{\text{耳}} = H'(r, \theta, \omega)E_0(\omega)H_{\text{扬}}, \quad (7)$$

其中 $H_{\text{扬}}$ 表示扬声器的频率响应. 再考虑到耳机重发时耳机-传声器的频率响应, 我们对原始信号 $E_0(\omega)$ 做如下处理:

$$E_{\text{in}} = \frac{H'(r, \theta, \omega)E_0(\omega)H_{\text{扬-传}}}{H_{\text{耳-传}}}, \quad (8)$$

其中 $H_{\text{耳-传}}$ 表示耳机-传声器频率响应. 将上述处理过的信号馈给耳机重发, 相当于在频域上再乘以 $H_{\text{耳}}$ (耳机频率响应), 所得信号正好为 $E_{\text{耳}}$.

1.2 信号制作

掩蔽声信号采用以临界带^[12]为带宽的矩形滤波器滤波后的粉红噪音信号, 共有 3 种频率范围: 200~300, 1080~1270 和 3700~4400 Hz, 每段信号有 5 ms 的上升沿、5 ms 的下降沿和 200 ms 的平稳状态时间. 被掩蔽声信号分别是频率为 250 Hz, 1170 Hz 和 4 kHz 的正弦信号. 250 Hz 的信号有 5 ms 的上升余弦平方过渡区、5 ms 的下降余弦平方过渡区, 信号总长度为 10 ms; 1170 Hz 和 4 kHz 的信号有 2 ms 的上升余弦平方、2 ms 的下降余弦平方过渡区和 6 ms 的平稳状态时间, 信号总长度为 10 ms. 实验信号分组如表 1 所示.

表 1 实验信号分组

	第一组	第二组	第三组
被掩蔽声频率/Hz	250	1170	4000
掩蔽声频率/Hz	200~300	1080~1270	3700~4400

信号中上升沿和下降沿的选取是为了减小人耳对由短信号突然发生和突然消失而产生的爆破声的觉察能力. 正弦被掩蔽声信号紧跟在窄带噪声之后(延迟时间为 0 ms). 掩蔽声声压级从 30 dB 变化到 70 dB(间隔为 10 dB).

将上述原始信号按(8)式做处理, 作为实验用信号. 当虚拟掩蔽声和被掩蔽声由同一方向, 即正前方 0° 方向发出情况时, 馈给耳机左右通道的信号分别为

$$E_{\text{左}} = \frac{H'_L(r, 0, \omega)E_0(\omega)H_{\text{扬-传}}}{H_{\text{耳-传}}}, \quad (9)$$

$$E_{\text{右}} = \frac{H'_R(r, 0, \omega)E_0(\omega)H_{\text{扬-传}}}{H_{\text{耳-传}}}. \quad (10)$$

当虚拟被掩蔽声由 30° 方向发出情况时, 馈给耳机左

右通道的信号分别为

$$E_{\text{左}} = \frac{H'_L(r, 30, \omega) E_0(\omega) H_{\text{扬-传}}}{H_{\text{耳-传}}}, \quad (11)$$

$$E_{\text{右}} = \frac{H'_R(r, 30, \omega) E_0(\omega) H_{\text{扬-传}}}{H_{\text{耳-传}}}. \quad (12)$$

当虚拟掩蔽声由 330° 方向发出时, 馈给耳机左右通道的信号分别为

$$E_{\text{左}} = \frac{H'_L(r, 330, \omega) E_0(\omega) H_{\text{扬-传}}}{H_{\text{耳-传}}}, \quad (13)$$

$$E_{\text{右}} = \frac{H'_R(r, 330, \omega) E_0(\omega) H_{\text{扬-传}}}{H_{\text{耳-传}}}. \quad (14)$$

2 实验方法

实验分为两个部分: 第一部分为耳机重发的虚拟掩蔽声和被掩蔽声都来自 0° 方向; 第二部分为耳机重发的虚拟掩蔽声和被掩蔽声分别来自 $+30^\circ$ 方向(右前方向)和 -30° 方向(左前方向). 两部分实验都按信号的频率分为 3 组(见表 1). 每组实验又按掩蔽声声压级不同划分为 5 小组(30~70 dB). 掩蔽阈值的测量是采用与一般扬声器重发实验相同的两段信号随机选择并结合三上一下的适应调整程序的方法^[13].

掩蔽实验开始之前先测量每个参加听音实验者对每个被掩蔽声信号的绝对闻阈, 并记录下来. 实验时, 在计算机中产生所需的信号, 经多通路声卡(u24)直接馈给 SENNHEISER HD 250 Linear II 耳机重发.

参加实验的倾听者共 3 人(2 男 1 女), 分别记为 S1~S3, 年龄 23~42 岁, 听力正常, 都曾有心理声学实验经验. 最后的实验结果取个体结果的平均值.

3 实验结果与分析

掩蔽阈值的实验结果如图 1 所示. 从图 1 中可以看出, 无论是在 0° 或 $\pm 30^\circ$ 声像位置情况下, 随着掩蔽声声压级的增大, 被掩蔽信号的阈值都随之增大. 这一点与以前关于时域掩蔽性质的研究结果基本一致.

对各个不同的频率(250, 1170 和 4000 Hz)和不同的掩蔽声声压级(30, 40, 50, 60 和 70 dB), 掩蔽声和被掩蔽声声像位置分开时的掩蔽阈值与掩蔽声和被掩蔽声声像位置相同时相比, 都有明显下降的趋势. 为了方便表述, 令

$$\Delta = M_1 - M_2, \quad (15)$$

其中 M_1 表示掩蔽声和被掩蔽声声像位置相同时的掩蔽阈值; M_2 表示掩蔽声和被掩蔽声声像位置分离时的掩蔽阈值. 由图 1 可知, 对不同的倾听者和频率范围, 去掩蔽量 Δ 均大于 0, 最大值达到了 15 dB. 因此尽管 3 位倾听者个人心理、生理状况上不同, 从而造成掩蔽阈值不同. 但是我们仍可以认为这一发现说明了在时域掩蔽中掩蔽声和被掩蔽声主观感觉水平空间位置上的分离确实产生了空间去掩蔽效应, 使得被掩蔽声的掩蔽阈值明显降低.

为了得到更直观的结果, 我们将图 1 转化为图 2

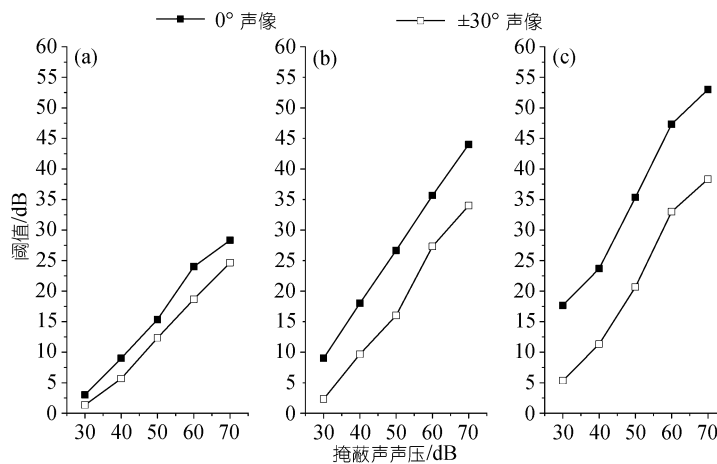


图 1 掩蔽阈值与掩蔽声声压的关系

(a) 掩蔽声为 200~300 Hz 的窄带噪声, 被掩蔽声为 250 Hz 纯音信号情况; (b) 掩蔽声为 1080~1270 Hz 的窄带噪声, 被掩蔽声为 1170 Hz 纯音信号情况; (c) 掩蔽声为 3700~4400 Hz 的窄带噪声, 被掩蔽声为 4000 Hz 纯音信号情况

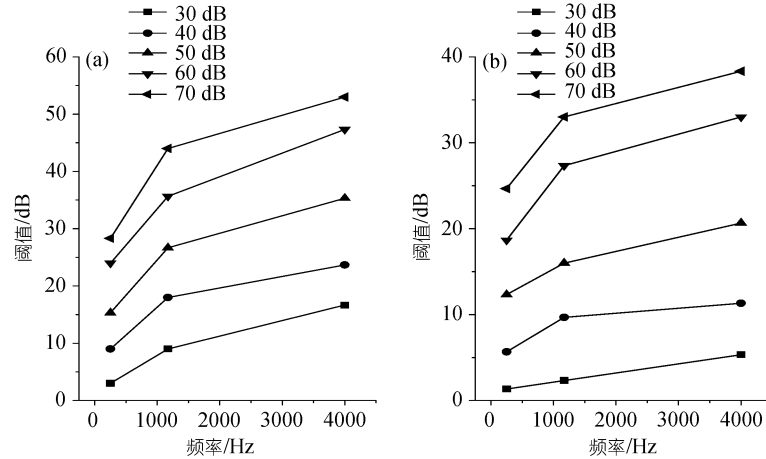
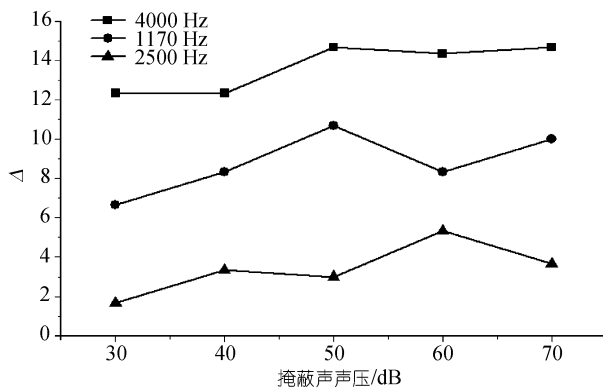


图 2 掩蔽阈值与被掩蔽信号频率的关系

(a) 耳机重发虚拟掩蔽声和被掩蔽声来自 0° 声像位置; (b) 掩蔽声和被掩蔽声分别来自 ±30° 声像位置

形式, 研究被掩蔽声的阈值与频率的关系. 从图 2 中可以看出, 在声压级一定的情况下, 无论掩蔽声和被掩蔽声是否来自同一位置, 被掩蔽声的阈值均有随频率升高的总体趋势.

为了进一步探讨去掩蔽值 Δ 与掩蔽声声压级以及信号频率间存在的关系, 对实验数据作进一步处理, 得到图 3.

图 3 去掩蔽值 Δ 与频率及掩蔽声声压的关系

从图 3 来看, 并没有发现 Δ 有随掩蔽声声压级增加而变化的明显规律. 这可以说明声源位置分离造成的被掩蔽声掩蔽阈值降低量跟掩蔽声声压级的大小无关. 但是从图中明显可以发现, 掩蔽声频率为 250 Hz 时的去掩蔽值 Δ 的曲线(实心三角表示)通常处于图形最底部, 而且去掩蔽值 Δ 都有随着频率升高而增大的总体趋势. 这一新发现与以前同时掩蔽下空

间去掩蔽效应研究的结果一致^[2,14].

对于上述实验结果, 我们认为其主要原因是较优耳作用^[15]. 因为在耳机重发实验中, 实验信号是将原始信号经过个性化 HRTF 及滤波处理后直接馈给倾听者, 所以计算双耳处声信号时, 不用再额外考虑人头, 躯干和环境等的影响, 而可以直接通过分析耳机发出的信号. 由(9)~(14)式得出, 在虚拟声像位于 0° 位置时, 双耳信掩比为

$$(SMR)_{AL} \approx (SMR)_{AR} \approx SMR, \quad (16)$$

而当虚拟声像分别在 ±30° 位置时, 双耳信掩比为

$$(SMR)_{BL} \approx 20 \log \left(\frac{H'_L(r, 30, \omega)}{H'_L(r, 330, \omega)} \right) + SMR, \quad (17)$$

$$(SMR)_{BR} \approx 20 \log \left(\frac{H'_R(r, 30, \omega)}{H'_R(r, 330, \omega)} \right) + SMR, \quad (18)$$

其中 $(SMR)_{AL}$ 代表掩蔽声和被掩蔽声的虚拟声像均位于 0° 位置时左耳处的信掩比, $(SMR)_{AR}$ 则代表此时右耳处的信掩比. $(SMR)_{BL}$ 代表掩蔽声和被掩蔽声的虚拟声像分别位于 ±30° 位置时左耳处的信掩比, $(SMR)_{BR}$ 表示此时右耳处的信掩比. SMR 代表初始信掩比. 比较(16)~(18)式可知, 掩蔽声和被掩蔽声的声像位置分别位于 ±30° 时, 右耳处的信掩比要比左耳处的信掩比以及掩蔽声和被掩蔽声的声像位置均处于 0° 位置时双耳处的信掩比高, 因此判断右耳为较优耳. 表 2 是具体的计算结果.

表 2 较优耳处声压变化

	掩蔽声频带/Hz	被掩蔽声频率/Hz	掩蔽声声压下降量/dB	被掩蔽声声压提高量/dB	信掩比提高量/dB
S1	200~300	250	0.3	0.7	1.0
	1080~1270	1170	4.3	5.2	9.5
	3700~4400	4000	6.5	4.0	10.5
S2	200~300	250	0.1	0.7	0.8
	1080~1270	1170	7.9	2.1	10.0
	3700~4400	4000	6.0	2.1	8.1
S3	200~300	250	0.4	1.0	1.4
	1080~1270	1170	9.0	1.9	10.9
	3700~4400	4000	5.7	3.2	8.9

由上面的讨论可知, 当掩蔽声与被掩蔽声源主观感觉上分离时, 如果原始信号掩蔽声声压 P_{M0} 和被掩蔽声声压 P_{S0} 均保持不变, 则较优耳处的信掩比与掩蔽声与被掩蔽声主观感觉上不开(都在 0°)时的信掩比相比要高一些. 因此, 如果由内耳处决定的掩蔽阈值保持不变的话, 要使掩蔽声能掩蔽被掩蔽声, 需要减小被掩蔽声声压 P_{S0} (即被掩蔽声阈值降低). 由于 HRTF 是因人而异的, 因此各人在较优耳处信掩比的提高量会有不同, 但总体趋势是一样的. 即掩蔽声源与被掩蔽声源在空间分离后, 掩蔽阈值会下降, 产生空间去掩蔽现象. 而且, 空间去掩蔽随频率而变, 频率越高, 空间去掩蔽现象越明显. 值得注意的是, 由于时域掩蔽的非线性, 被掩蔽声阈值的降低量不等于较优耳处的信掩比的提高量.

但是仅利用较优耳处信掩比的提高, 还不能完全定量地解释实验结果. 根据掩蔽阈值随掩蔽声声压级非线性变化的规律, 可以得出: 被掩蔽声的阈值在掩蔽声声压级较低时减少的幅度会小于较优耳处信掩比的提高量, 而在掩蔽声声压级较高时被掩蔽声阈值减少的幅度会接近于或稍大于信掩比提高量. 然而图 3 和表 2 的数据并不符合这一假设, 例如当被掩蔽声纯音的频率为 4 kHz 时, 被掩蔽声的阈值差别即去掩蔽量 Δ 高达 15 dB, 而此时的较优耳处信掩比的提高远没有到达这个程度. 另外从实验数据中发

现掩蔽阈值差别并没有明显随掩蔽声声压而改变的规律. 因此存在除较优耳作用以外的其他因素影响掩蔽阈值的变化, 我们认为这是高层听觉系统对双耳声信息综合处理即双耳效应的结果. 与掩蔽声和被掩蔽声之间的双耳相位差为 0 时的情况比较, 当掩蔽声和被掩蔽声的双耳相位不同即双耳相位差不为 0 时, 掩蔽阈值会下降(时间差所造成的影响太小, 可以忽略). 但双耳效应的详细机理, 它对各频率的影响以及倾听者个体间的差异等问题仍有待进一步的研究.

4 结论

将加以改进了的耳机虚拟声重发方法用于前向掩蔽的实验中, 这样不仅能达到声源位置主观感觉上空间分离的效果, 而且能够很好地虚拟自由场内扬声器重发实验的环境条件. 实验结果为前向掩蔽的空间去掩蔽提供了又一例证. 值得提出的是, 在用耳机重发做实验时, 采用了个性化 HRTF 做信号处理, 因此得到了更为理想和准确的实验结果: 虽然声像位于 0° 位置和分别位于 $\pm 30^\circ$ 位置时掩蔽阈值均随频率和掩蔽声声压级的增加而增加, 但两种情况下的掩蔽阈值是不同的, 存在着空间去掩蔽现象, 最大去掩蔽值可达 15 dB. 这种空间去掩蔽效应可归因于较优耳贡献和双耳效应的结合, 但主要是较优耳的贡献.

参考文献

1 Kopco N, Shinn-Cunningham B G. Spatial unmasking of nearby pure-tone masked signals in a simulated anechoic environment. J Acoust Soc Am, 2003, 114(5): 2856—2870[DOI]
2 Freyman R L, Helfer K S, McCall D D, et al. The role of perceived spatial separation in the unmasking of speech. J Acoust Soc Am,

- 1999, 106: 3578—3588[DOI]
- 3 Saberi K, Dostal L, Sadralodabai T, et al. Free-field release from masking. *J Acoust Soc Am*, 1991, 90: 1355—1370[DOI]
 - 4 Kidd Jr G, Mason C R, Rohtla T L, et al. Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *J Acoust Soc Am*, 1998, 104: 422—431[DOI]
 - 5 Doll T J, Hanna T E. Spatial and spectral release from masking in three-dimensional auditory displays. *Hum Factors*, 1995, 37: 341—355[DOI]
 - 6 Gatehouse R W. Further research on free-field masking. *J Acoust Soc Am*, 1987, 82(Suppl 1): S108
 - 7 Moore B C J. *An Introduction to the Psychology of Hearing*. 5th ed. San Diego: Academic, 2003
 - 8 Gunawan T S, Ambikairajah E. A new forward masking model and its application to speech enhancement. In: 2006 IEEE International Conference on Acoustics, Speech and Signal Processing. New York: IEEE, 2006, 1: I149—I152
 - 9 Oxenham Andrew J, Plack Christopher J. Effects of masker frequency and duration in forward masking: Further evidence for the influence of peripheral nonlinearity. *Hearing Res*, 2000, 1(150): 258—266[DOI]
 - 10 Kim S M, Choi W. On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach. *J Acoust Soc Am*, 2005, 117: 3657[DOI]
 - 11 Stern R M, Shear G D. Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay. *J Acoust Soc Am*, 1995, 100: 2278—2288[DOI]
 - 12 Zwicker E, Flottorp G, Stevens S S. Critical bandwidth in loudness summation. *J Acoust Soc Am*, 1957, 29: 548—557[DOI]
 - 13 Levitt H. Transformed up-down methods in psychophysics. *J Acoust Soc Am*, 1971, 49: 467—477 [DOI]
 - 14 Xie Z W, Yin J X, Rao D. The experiment study of spatial mask effect. *Chin J Acoust*, 2006, 25(1): 75—86
 - 15 Good M D, Gilkey R H, Ball J M. The relation between detection in noise and localization in noise in the free field. In: Gilkey R H, Anderson T R, eds. *Binaural and Spatial Hearing in Real and Virtual Environments*. Mahwah: Lawrence Erlbaum Associates, 1997. 349—376