

普通话快速噪声下言语测听材料的等价性评估

任丹丹¹, 张华¹, 武文芳¹, 曹文², 郭连生¹, 陈建勇³, 李玉玲¹, 王硕¹, 吴丹¹

(1 首都医科大学附属北京同仁医院耳鼻咽喉头颈外科, 北京市耳鼻咽喉科研究所, 耳鼻咽喉头颈科学教育部重点实验室(首都医科大学), 北京 100730; 2 北京语言大学对外汉语研究中心, 北京 100083; 3 上海交通大学医学院附属新华医院耳鼻咽喉头颈外科, 上海交通大学耳科学研究所, 上海 200092)

[摘要] **目的** 开发一套标准的普通话版快速噪声下言语测听材料, 评估依据听力学要求和语言学特征选取的300个语句材料的等价性, 并获得其在50%正确识别时所对应的信噪比(记为SNR-50)的正常值。**方法** 首先将这些语句材料按照数字化录制原则录制为女声音频材料, 背景噪声为4人同时谈话的言语噪声。言语声和噪声混合在一起。选取30例正常受试者, 采用听说复述方式, 分别在+6、+3、0、-3、-6 dB信噪比条件下测试。对测试所得结果使用SPSS17.0统计软件进行数据分析和强度-识别率函数曲线的拟合。**结果** 300个语句在噪声环境下测试结果分析显示, 斜率和阈值不呈正态分布。剔除斜率过大(>0.72)的语句, 及曲线拟合相关系数过小($r<0.7$)的语句, 精选出等价性较好的78句。后者呈正态分布。**结论** 经过对300个语句的等价性评估, 经过严格的筛选及精细调整, 最终保留的78个语句在正常人中的等价性较好, 下一步可进行这些语句复测信度的测试。

[关键词] 语言试验; 言语识别测验; 噪声; 语句材料; 等价性

Equivalence study of mandarin quick speech in noise test materials

REN Dandan¹, ZHANG Hua¹, WU Wenfang¹, CAO Wen², GUO Liansheng¹, CHEN Jianyong³, LI Yuling¹, WANG Shuo¹, WU Dan¹

(1 Department of Otolaryngology Head and Neck Surgery, Beijing Tongren Hospital, Capital Medical University, Beijing Institute of Otolaryngology, Key Laboratory of Otolaryngology Head and Neck Surgery (Capital Medical University), Ministry of Education, Beijing, 100730, China; 2 Center for Studies of Chinese as a Second

Language, Beijing Language and Culture University, Beijing, 100083, China; 3 Department of Otolaryngology Head and Neck Surgery, Xinhua Hospital, School of Medicine, Ear Science Institute, Shanghai Jiao Tong University, Shanghai, 200092, China)

Corresponding author: ZHANG Hua (Email: a-zhang@263.net)

[ABSTRACT] **OBJECTIVE** To select equivalent sentences and get their SNR-50 value, in order to develop the Mandarin Quick Speech in Noise Test (M-Quick SIN). **METHODS** All the 300 sentences were recorded into the digital materials with a female voice at first, the background noise was 4-talker babble noise, and they were mixed together. Then they were presented to 30 subjects at +6, +3, 0, -3, -6 dB SNR. The subjects should repeat the sentences that they heard. Evaluation and analysis of P-I functions of all sentences were done by SPSS 17.0. **RESULTS** The Histogram and the analysis showed that neither the thresholds nor the slopes were in the normal distribution. The sentences, whose regression coefficients were below 0.7 and slopes were greater than 0.72 were abandoned. The retained 78 sentences had good equivalence. **CONCLUSION** Through the evaluation and selection, the retained 78 sentences, which were adjusted for their SNR (signal-to-noise ratio), had good equivalence among normal hearing people. Next, test-retest reliability should be done for these sentences.

[Key words] Language Tests; Speech Discrimination Tests; Noise; sentences; equivalence

基金项目: 国家自然科学基金资助项目(81070784)

第一作者简介及通讯: 任丹丹, 女, 河南人, 在读医学硕士, 主要从事听力学及耳科学的相关基础和临床研究。

Email: rdan123589@163.com

通讯作者: 张华 (Email: a-zhang@263.net)

人们日常生活中的工作和学习大多是在噪声环境下进行的, 噪声下的言语交流是一项基本技能。听力障碍者的主要表现之一是理解言语困难, 噪声环境下尤其如此, 即便是使用助听器或(和)人工耳蜗以后也难以克服。如何准确评估患者在噪声下的言语感知能力就变得尤为重要。鉴于纯音听力测试等主、客观检查并不能反

映听力损失者的这种能力^[1], 而且安静环境下的言语测听结果也不能推测出他们的此种能力^[2]。而噪声下的言语测听, 由于能较好模拟真实生活中的言语交流环境, 能在一定程度上反映和评估听力障碍者在噪声环境下的言语感知能力。Killion^[3]认为要想得知听力障碍者噪声下的言语感知能力, 最好的办法就是直接测试出来。2001年出现了英文版的快速噪声下言语测试 (Quick speech in noise test, Quick SIN) 测试, 该方法简便、快捷, 能在较短时间内获得听力障碍者的信噪比损失 (signal-to-noise ratio loss, SNR loss) 情况, 评价其噪声下言语交流能力^[4]。由于汉语与英语分属两种不同的语言系统, 以英文为测试材料的Quick SIN并不能直接用来测试我国日常以汉语普通话为主要交流方式的人群。因此, 本课题拟开发一套汉语普通话版的快速噪声下言语测试材料, 以期快速、简便、准确评价患者的信噪比损失, 为助听器的选配和人工耳蜗植入者的康复提供一定的参考, 使患者在早期就能够建立正确的期望值, 为制定言语康复策略提供一定的参考信息。

1 资料与方法

1.1 测试材料。本研究所采用的测试材料是300个语句材料, 每个语句有5个关键词。这些材料是依据听力学要求和语言学特征, 参照英文版Quick SIN材料的特点, 根据我国的普遍受教育情况及汉语普通话本身的特点, 确定一定的原则, 并根据这些原则, 从合适的语料库中选取^[5], 并请语音学与语言学专家共同修订后而获得。这300个语句材料的选取原则为: ①可预见性低; ②针对初中以上文化程度的成人; ③每句话有5个关键词; ④没有专业性的术语, 所选词语均是日常所熟知, 没有生僻、晦涩、难懂的词语; ⑤语句自然, 符合语法及逻辑; ⑥句式变化少, 且都为简单的单句; ⑦均为现代用语。

所有300个语句材料均在中央人民广播电台采用数字化录制原则, 由专业的录音师录制完成。录制时本底噪声≤45 dB (A)。语句材料由有多年播音主持经验的年青女播音员以正常说话语速朗读, 以录自调音台信号发生器的1 kHz校准纯音的强度做参考, 使所有语句的音频材料

句与句之间的强度变化不超过±3 dB。语句与语句之间的间隔为5 s。本课题的背景噪声选用另外4位音色各不相同的专业播音员 (3女1男) 分别录制, 然后混合在一起成为4人同时谈话的言语噪声。使用Cool Edit Prof 2.1专业声音处理软件进行后期处理, 使言语声和噪声归一化, 并保证每个语句都与它所对应的噪声在时间上精确匹配。噪声和言语声混合在同一个声道内, 且噪声总是先于语句前2 s出现。受试者有5 s的时间复述所听到的语句, 语句结束后噪声也同时结束。所有的300句话随机分为5组 (每组为60个语句, 以组一、组二、组三等表示), 且每组均采用+6、+3、0、-3、-6 dB 5种信噪比合成。录制合成时固定言语声强度, 调节噪声强度以改变信噪比。

1.2 测试对象。受试者30例, 男13例, 女17例。年龄18~25岁之间, 平均年龄为 (23.7±1.8) 岁。无耳科病史, 耳科检查正常。纯音听力测试250~8000 Hz各个倍频程频率的听阈均在25 dB HL以下, 声导抗检查正常, 鼓室图均为A型, 声反射均能引出。选取受试者较好耳进行测试。其中左耳13例, 右耳17例。受试者均为大专以上文化程度, 日常以汉语普通话为主要交流方式, 口齿清晰。

1.3 测试地点及测试设备。测试地点为北京同仁医院临床听力学中心标准隔声屏蔽室, 环境噪声<20 dB (A)。测试采用电脑控制声音的输出, 为使测试时音频的播放更灵活且易控制, 测试材料采用Cool Edit Prof 2.1声音软件播放, 并送入GSI-61 听力计的外接音频输入端口, 通过TDH-39耳机进行测试。整个测试系统按照国标《GB/T 17696声学 测听方法 第三部分: 语言测听》校准。

1.4 测试方法。单耳耳机给声, 采用受试者纯音听阈较好耳进行测试。言语声的固定强度为每个受试者的最舒适阈 (most comfortable level, MCL)。所有的300个语句都要在5种信噪比条件下 (+6、+3、0、-3、-6 dB) 被每个受试者测试到。即每个受试者共听到300句×5种信噪比=1500次语句。受试者的具体测试顺序 (表1)。

对每例受试者进行测试前指导, 向受试者详细讲解测试过程。如可告诉受试者: “你将会听到一位女士说话的声音, 背景同时也有其他几个人在讨论着。刚开始这位女士的声音很大很清晰, 渐渐就被其他人说话的噪声所淹

表1 受试者的测试顺序 (组次)					
受试者	信噪比 (dB)				
	+6	+3	0	-3	-6
1	一二三四五	二三四五一	三四五一二	四五一二三	五一二三四
2	二三四五一	三四五一二	四五一二三	五一二三四	一二三四五
3	三四五一二	四五一二三	五一二三四	一二三四五	二三四五一
4	四五一二三	五一二三四	一二三四五	二三四五一	三四五一二
5	五一二三四	一二三四五	二三四五一	三四五一二	四五一二三
6	一二三四五	二三四五一	三四五一二	四五一二三	五一二三四
...
30	五一二三四	一二三四五	二三四五一	三四五一二	四五一二三

没,但是要集中注意力聆听这位女士在说什么,然后把她所说的全部重复出来,有听不清的地方可以联系整个语句的句意进行猜测。”在获知受试者的最舒适阈之后,要给予其3张练习表进行练习,便于受试者进一步熟悉测试流程。尽量使受试者在测试过程中注意力集中,心态平和,避免焦躁、疲劳、走神、漫不经心等情况出现。由于本次实验测试时间较长,每例受试者均分3次测完,每次测试1.5 h,中间安排休息。每次测试大约间隔1周。

1.5 统计学分析。在记录测试结果时以受试者答对的关键词的个数采用“全或无”的方式计分,整个关键词答对记1分,答错记为0分。无论单字、2个字、3个字或4个字短语,只要有一个字答错就记为0分。同时将30例受试者的所有测试数据建成数据库,利用SPSS 17.0统计软件的“非线性拟合功能”,以Logit曲线作为P-I函数(即识别率-强度函数曲线)模板^[6],拟合出每个语句的识别率-信噪比函数曲线。其拟合公式为 $P = \exp(a + b \times i) / [1 + \exp(a + b \times i)]$,P为识别率,i为信噪比值。从这个公式上可以计算出任何识别率所对应的信噪比强度,包括获得50%正确识别时的信噪比强度,即SNR-50值和它所对应的斜率,同时也可从图像上观察其走势是否一致。分别考察300个语句的P-I函数的阈值(SNR-50值)和斜率分布情况。

2 结果

30例受试者均顺利完成实验。根据实验所得数据,将所有300个语句的结果都拟合成P-I曲线,其曲线拟合相关系数各不相同。计算出所有语句的斜率和SNR-50值,对其进行正态性检验。经检验无论是其SNR-50值还是斜率, P 均 < 0.05 ,不服从正态分布。分别绘制其斜率和SNR-50值所对应的分布图(图1),300个语句的变异度较大,少数语句即使在最低的信噪比强度下仍能被100%正确识别。反之,个别语句即使在最高的信噪比强度下得分也趋近于0,存在地板效应。所有这些语句的SNR-50值均由拟合公式计算而来,因此,其值分布较广,从-25~+3.75 dB信噪比。斜率分布也为偏态分布,变异度较大(图2)。

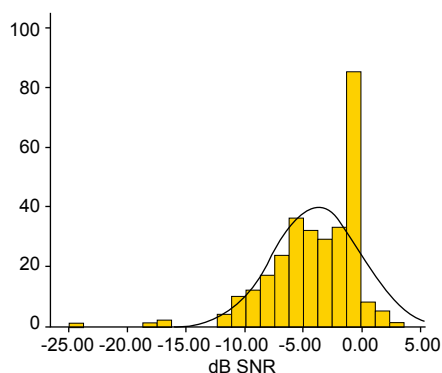


图1 300个语句噪声下言语识别SNR-50分布图

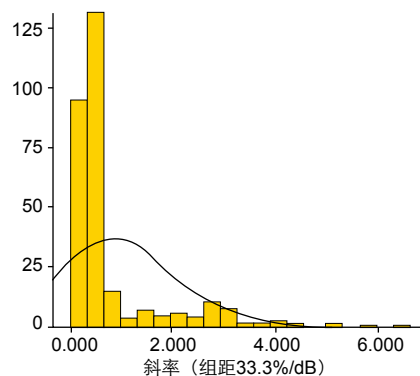


图2 300个语句噪声下P-I函数曲线斜率分布图

根据以上统计分析结果,为筛选出同质性好的语句,剔除了语句中变异度大的。剔除标准如下:①P-I函数曲线拟合相关系数 $r < 0.7$ 的语句^[7];②语句斜率 > 0.72 即斜率过陡的语句,即敏感度过高者。经过以上筛选,保留78句符合以上条件的语句。经检验,78个语句的等价性尚可,其阈值和斜率都呈正态分布,且较为集中(图3,图4)。其中SNR-50值为 (-2.00 ± 1.75) dB,95%置信区间为 $[-2.40, -1.60]$ 。而斜率为 (46.4 ± 1.0) %/dB。对所选出的语句进行进一步的同质性精细调整,使其SNR-50值都落在-2 dB SNR上。利用Cool Edit Prof 2.1专业声音处理软件,对这些选出的已经经过语句-噪声归一化处理后的语句再次进行调整。所采用的调整方法与英文的Quick SIN一致。例如,语句1的阈值为-3 dB SNR,则与它相匹

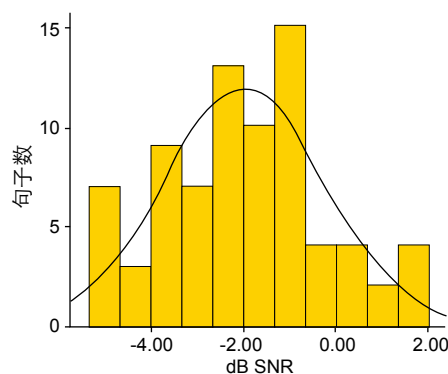


图3 78个语句噪声下P-I函数曲线SNR-50分布图

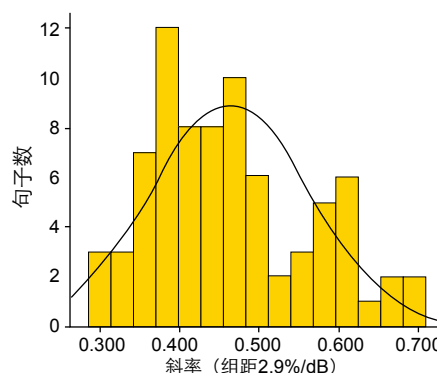


图4 78个语句噪声下P-I函数曲线斜率分布图

配的噪声则相应的下调1 dB, 调整后的语句仍然要保证与噪声在时间上的精细同步^[8]。

原有300个语句的测试结果经过函数拟合后所得图形比较杂乱, 无规则(图5)。但经过对变异性较大的语句进行剔除, 并对所保留的语句进一步调整后, 其同质性有了很大改观, P-I曲线较为整齐, 走势也较为一致(图6)。

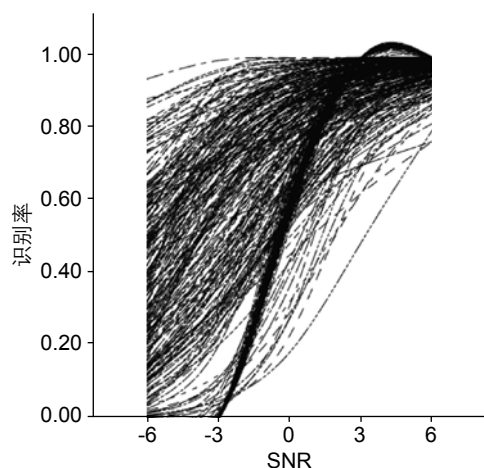


图5 原始300个语句的P-I曲线

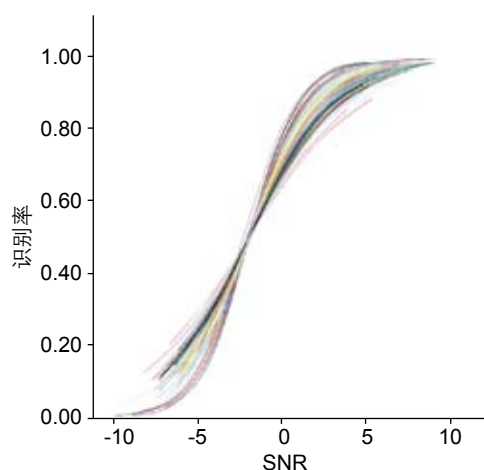


图6 经进一步调整后78个语句的P-I曲线

3 讨论

噪声环境下的交流困难一直是困扰听力障碍者(尤其是感音神经性听力损失患者)的常见问题, 它既影响了助听器的使用满意度, 同时也对人工耳蜗植入者的术后效果产生不利影响。本课题拟开发一套标准化的汉语普通话版噪声下快速言语测试材料, 以评估听力障碍者在噪声环境下有无言语理解困难及其困难程度, 使患者建立正确的期望, 为助听器的选配和人工耳蜗植入者术前、术后评估及言语康复训练策略的制定提供参考。

听力损失可以分为对声音敏感度的下降及清晰度的下降。听敏度与对声音信号的感知有关, 而清晰度则涉及对声音信号的辨别。前一个是由听力损失引起的, 后一个则

与声音的畸变有关。对声音敏感度的下降可由纯音听力测试的听力图上直接反映出来, 而由于声音畸变所引起的清晰度下降就要靠噪声下的言语测听来评价^[9]。本研究作为普通话快速噪声下言语测试材料研发的一部分, 最终的测试材料是用来测试患者的信噪比损失, 即与正常人平均噪声下言语理解能力相比, 受试者能听懂50%言语所额外增加的分贝数^[9]。信噪比损失能直观反映患者的噪声下言语感知情况。听力图相似的患者, 其噪声下言语理解能力会有很大差异。据Killion报道, 一部分平均听力损失40~60 dB患者可能基本上没有信噪比损失, 而另外一些同样的轻中度听力障碍者的信噪比损失却可以达到15~20 dB^[1]。本研究期望最终所开发出来的噪声下言语测听测试材料和方法能最大程度的模拟真实生活中的交流环境。因此, 测试材料选用语句材料。相对于字与字之间或词与词之间较为孤立的单、双音节词而言, 自然状态下的语句有较大的动态范围, 较能代表日常的语谱分布, 有正常语言的语气起伏和停顿, 还带有自然的语调, 更接近日常生活中的言语交流^[10]。而背景噪声选用4人同时谈话噪声, 有报道认为这更能代表日常生活社交环境中的噪声^[8]。这种掩蔽噪声不仅有语谱能量上的掩蔽, 也有声音信号所携带信息的干扰。

为使测试材料能够更方便应用临床和科研, 本课题拟开发出来的噪声下言语测听材料要能在短时间内(最好几分钟内)测试出听力障碍者在噪声下言语理解情况, 对所需测试材料的等价性要求较高。因此, 我们从300个语句中只选出了同质性好的78句。对于这300个语句材料, 即使在最初选句的时候, 虽然尽量在语言学上使其难度一致, 但在背景噪声中及听力学上的等价性只能在实验中进行评估。另外, 本研究所用语句材料的特点为可预测性低, 冗余度较小。所有语句材料都遵守着相应的语法规则和句子结构的限制。虽然受试者在听不清楚的时候可根据这些规律进行一定的猜测, 但语句的句意所包含的上下文信息较少, 冗余度并不高, 因此材料较为关注受试者对声音线索的感知能力。尽管在选取的时候尽量使各个语句在难度上保持一致, 但它们所包含的信息量依然不尽相同, 加之可预测性目前尚无一定的评价标准, 很难在一开始就保证语句能绝对一致。

英文版Quick SIN所用材料是从美国电子电气化工程协会(Institute of Electrical and Electronic Engineers, IEEE)语料库中选取的。这些语句原来为音位平衡语句, 但在Quick SIN开发的过程中并未刻意维持。它们针对高中文化水平的成年人, 可预测性低, 所含的语意信息较少, 很难根据上下文从前面推测出后面一个词是什么。更多的是依赖受试者对声音信号的感知和以往的知识背景、生活经验等来判断所听到的内容^[4]。本研究在选取语句材料的时候虽然也保留其可预见性低的特点, 但是并未遵守音位平衡的原则。本研究最终目的是要获得能在较短时间内(1~2 min内)获知听力障碍者噪声下言语感知能

力的测试结果, 而每张表所用音节不可能完全覆盖人们日常生活中所有常用音节。因此, 正如英文版的Quick SIN在开发的过程中摒弃了音位平衡的原则一样, 本研究在选材过程中虽然选取的也是人们日常较常用语句, 但并未刻意考虑音位平衡。

本次研究目的在于获得300个语句材料的SNR-50值, 并评估其等价性。目前的测试结果表明SNR-50均值为 (-2 ± 1.75) dB, 要低于英文版的Quick SIN, 即正常值为+2 dB信噪比^[8]。出现这种情况可能与汉语本身的特点有关。由于汉语与英语是两种完全不同系统的语言, 汉语为声调语言, 声调在辨义过程中起着重要的、甚至是决定性作用^[11], 更有助于对语言的理解。另外两者的文化背景和语句结构等的不同也会使测试结果有所差异, 在普通话噪声下言语测试 (mandarin hearing in noise test, MHINT) 的开发过程中也发现类似情况, 汉语普通话版语

句识别阈 (reception threshold for sentences, RTS) 的阈值要比英文版低^[12]。此外, 实验条件不同可能对此也有影响, 在录制测试材料的时候, 发音人的不同可能也会对此有所影响。言语测听词表的等价性研究是言语测听标准化的重要组成部分^[13]。本研究对初步选出的300个将要应用在快速噪声下言语测试中的语句材料进行等价性评估, 并拟合出这些语句的P-I函数曲线。在剔除一部分斜率过陡, 曲线拟合相关系数过小的语句后, 筛选出78个语句。经进一步调整后, 这些筛选出的语句材料有较好的同质性和较高的等价性, 达到预期的实验目的。下一步可进行其复测信度的测试, 以对这些语句进行再次验证。

志谢 感谢中国社会科学院语言研究所的李爱军老师、完权老师和贾媛老师及中央人民广播电台的周伟老师等在本文的写作过程中给予的帮助

参考文献

1. Mead KC, Patricia NA. What can the pure-tone audiogram tell us about a patient's SNR loss? *Hearing Journal*, 2000, 53: 46-53.
2. Brian T. Speech-in-noise tests: How and why to include them in your basic test battery. *Hearing Journal*, 2003, 56: 40-46.
3. Killion MC. New thinking on hearing in noise: a generalized articulation index. *Semin Hear*, 2002, 23: 57-76.
4. Etymotic Research. QuickSIN: Speech-in-Noise Test (Version 1.3) [Audio CD]. USA Elk Grove Village: Etymotic Research, 2001.
5. 曹文, 张劲松. 面向计算机辅助正音语音语料库的创制与标注. *语言文字应用*, 2009, 4: 124-133.
6. Nissen SL, Harris RW, Jennings LJ, et al. Psychometrically equivalent Mandarin bisyllabic speech discrimination materials spoken by male and female talkers. *Int J Audiol*, 2005, 44: 379-390.
7. Munro BH, Visintainer MA, Page EB. *Statistical Methods for Health Care Research*. USA Philadelphia: JB Lippincott Co, 1993: 181.
8. Killion MC, Niquette PA, Gudmundsen GI, et al. Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*, 2004, 116: 2395-2405.
9. Killion MC. SNR loss: I can hear what people say, but I can't understand them. *Hearing Review*, 1997, 4: 8-14.
10. 韩娜. 噪声下的言语识别能力和评价方法. *听力学言语及疾病杂志*, 2005, 13: 202-204.
11. 张华. 研发汉语测听材料的重要性. *中华耳科学杂志*, 2008, 6: 11-12.
12. Wong LL, Soli SD, Liu S, et al. Development of the Mandarin Hearing in Noise Test (MHINT). *Ear Hear*, 2007, 28: 70s-74s.
13. 王靓, 张华, 王硕, 等. 普通话单音节词言语测听材料的等价性分析. *中国耳鼻咽喉头颈外科*, 2006, 13: 397-401.

(收稿日期: 2011-12-05)

编辑 沈懿