汉语普通话CMnBio测听句表的编制及在成年人工耳蜗植入者中的验证

郗昕 1 王烨 1 石涯 2 高瑞 3 李思齐 3 仇心悦 3 王倩 4 徐立 5

1 解放军总医院耳鼻咽喉头颈外科医学部；国家耳鼻咽喉疾病临床医学研究中心，北京 100853；2 浙江医院耳鼻咽喉科，杭州 310007；3 浙江中医药大学医学技术学院，杭州 310053；4 首都医科大学生物医学工程学院，北京 100069；5 北京语言大学语言康复学院，北京 100083；6 首都医科大学附属北京安贞医院，北京 100193

美国俄亥俄州立大学沟通科学与障碍专业，美国俄亥俄州阿森斯 45701，美国

邮编和王烨对本文具有同等贡献

通信作者：徐立，Email: xul@ohio.edu

【摘要】目的 为克服现有中文言语测听材料在成年耳蜗植入者中的天花板效应，依据英文AzBio句表的开发流程，开发一套更难的、更贴近现实生活的中文普通话句表，并在成人耳蜗植入者中验证其等价性。方法 选取成人日常生活口语素材，邀请2男2女播音专业学生在本底噪声低于20 dBA(0分贝)的隔声室内录音1020句，以声码器技术仿真成5通道人工耳蜗的音效，编撰成51张表，以拉丁方式设计17名听力正常人进行言语识别率测试，从均值±1标准差( x ± 1, 61.9% 至 92.2%)的范围内挑选出同质性良好的600句(每个播音人150句，识别率均值78%)，编撰成每表20句共30张CMnBio句表。另行招募30例使用年限12个月-12年不等的成年人工耳蜗植入者，在65 dB SPL声级下依序测试30张表的言语识别率。采用拉丁方设计方案，每张测试表都有同样的机会成为首张测试表，以避免练习效应和疲劳效应的干扰。使用SPSS 18.0统计软件进行数据分析。结果 (1)30例人工耳蜗植入者的言语识别率呈现较大的个体间差异，但每例植入者对30张表的言语识别率相对稳定。(2)对言语识别率进行合理的化反正弦变换后，计算每例植入者每表识别率与其30张表平均识别率的差值；以30张CMnBio表作为考察对象，对30例植入者的上述差值进行重复测量方差分析，F = 3.503, P < 0.001；进而采用Post Hoc Tukey两两比较，除第4, 5. 10, 29号表外，其余26张表之间差异无统计学意义(P > 0.05)。(3)26张等价CMnBio句表在30例成年耳蜗植入者中的识别率为(65.4±13.3)%，呈良好的正态分布。(4)采用n=65的二项分布模型可很好地反映表间的变异性。结论 本研究建立了面向成年人工耳蜗植入者，具有良好等价性的26张普通话CMnBio句表，并提供了使用单张表和两张表时识别率得分95%置信度下的临界差值。为普通话成人人工耳蜗临床实践及开展跨语言种成效对比研究，提供了一个可避免天花板效应的实用工具。

【关键词】 测听法, 语 言： 语句识别；耳蜗植入术；语后聋；汉语普通话

基金项目：国家重点研发计划(2020YFC2004005)；国家自然科学基金(61370023, 81460099)

DOI: 10.3760/cma.j.cn115330-20230103-00003

收稿日期：2023-01-03

本文编辑：金昕

引用本文：郗昕，王烨，石涯，等. 汉语普通话CMnBio测听句表的编制及在成年人工耳蜗植入者中的验证[J]. 中华耳鼻咽喉头颈外科杂志，2023，58(11): 1112-1122. DOI: 10.3760/cma.j.cn115330-20230103-00003.

本文曾以英文在线发表：Xi X, Wang Y, Shi Y, et al. Development and validation of a Mandarin Chinese adaptation of AzBio sentence test (CMnBio)[J]. Trends Hear, 2022, 26: 23312165221134007. DOI: 10.1177/23312165221134007.
Development and validation of a Mandarin Chinese adaptation of AzBio sentence test (CMnBio) for adult cochlear implant users

Xi Xin¹, Wang Ye², Shi Ya², Gao Rui², Li Siqi², Qiu Xinyue¹, Wang Qian¹, Xu Li²

¹Department of Otolaryngology Head and Neck Surgery, Chinese PLA General Hospital, National Clinical Research Center for Otolaryngologic Diseases, Beijing 100853, China; ²Department of Otolaryngology, Zhejiang Hospital, Hangzhou 310007, China; ³School of Medical Technology, Zhejiang Chinese Medical University, Hangzhou 310053, China; ⁴School of BioMedical Engineering, Capital Medical University, Beijing 100069, China; ⁵School of Communication Science, Beijing Language and Culture University, Beijing 100083, China; ⁶Communication Sciences and Disorders, Ohio University, Athens, OH 45701, USA

Xi Xin and Wang Ye contributed equally to the article

Corresponding author: Xu Li, Email: xul@ohio.edu

【Abstract】 Objectives The goal of the present study was to develop and validate a set of Mandarin Chinese adaptation of AzBio sentence test (CMnBio) used for the assessment of speech recognition in hearing-impaired listeners and cochlear implant (CI) users. Methods Following the procedures of development of the English AzBio sentence materials, the present study was conducted in two stages. In the first stage, a total of 1,850 sentences were compiled and recorded by four Mandarin-speaking adult talkers (two males and two females), from which 1,020 sentences with better sound quality and higher naturalness were retained. All the 1,020 sentences were processed through a 5-channel noise vocoder and presented to 17 normal-hearing Mandarin-speaking adults for recognition. A total of 600 sentences (150 from each talker) in the range of approximately 62% to 92% correct were subsequently selected to compile 30, 20-sentence lists. In the second stage, 30 postlingually-deafened adult CI users were recruited to verify the list equivalency. Finally, the binomial distribution model was adopted to account for the inherent variability in the lists and to generate the lower and upper limits of the 95% critical differences.

Results The average sentence recognition of the 5-channel vocoder-processed 30 lists of Mandarin Chinese AzBio sentences (600 sentences total) by the 17 normal-hearing listeners was 78.0% correct. The validation based on the performance scores of the 30 CI participants revealed that 26 of the 30 lists were equivalent through a repeated-measures analysis of variance and a following post hoc Tukey's test. The binomial distribution modelling indicated that the inter-list variability could be accounted for with a 65-item binomial distribution model. The lower and upper limits of the 95% critical differences for one- and two-list recognition scores were generated.

Conclusions The Mandarin Chinese adaptation of AzBio sentence test (CMnBio) is developed and validated for speech recognition assessment of CI users. The final set of 26, 20-sentence lists shows high inter-list reliability or equivalency. The elevated level of difficulties of the CMnBio sentences can help eliminate the ceiling effects when testing sentence recognition in Mandarin-speaking CI users. The lower and upper limits of the 95% critical differences derived from the binomial distribution modelling can provide guidance for detection of a significant difference in recognition scores in clinical settings.

【Key words】 Audimetry, speech; Sentence recognition; Cochlear implantation; Postlingual deafness; Mandarin Chinese

Fund program: National Key Research and Development Project of China (2020YFC0204005); National Natural Science Foundation of China(61370023, 81460099)
播音人会被要求重复该语句，语速及意外状态。监控播音人的声软件界面。Adobe Audition 0 bits kHz，同步将语音实时显现。44分辨率采样速率。SD格式记录在其内置的。使用2250声级计B&K配装，避免过度强调和夸张的韵。

词汇或句式结构等语言学特征进行限制。未对语句的音地域特征以及成语或专有名词，话句长~16低预判性的语句，如句意清晰，AzBio再从国内电视剧及社交媒体的对话中挑选出居海外且英语水平较高的华人听力学家审听，以协同法命名为句表。AzBio2021年期间建立了一套具有较高难度的中—27版本也得到开发和验证，AzBio语句的识别能力的最简言语测试集。植入者在·27~24BKB⁃SIN句表和还选用了更高难度的

植入者言语CI⁃25FrBio⁃22句表和还选用了更高难度的

CNC除继续保留第一版中最低难度的AzBio句表开发研究。AzBio句表的编制流程

本文参照英文AzBio句表的开发流程，2013—2021年期间建立了一套具有较高难度的中文普通话版AzBio句表，命名为CMnBio，以协同法语版FrBio和希伯来语版HeBio的命名体系，并在普通话为母语的成年CI受试者中验证其等价性。

### 资料与方法

1. 语句编撰、录音及声学处理：由普通话为母语、久居海外且英语水平较高的华人听力学家审听英文AzBio语句，再从国内电视剧及社交媒体的对话中摘编语言信息量与之相近，语意清晰、结构合理、低预判性的语句1800句，句长7~16个字，如“二个人在走廊上认出了对方”“入口的红酒也不一定好喝”“电影票太贵了，还是在家看电视吧”！话题内容涵盖成年人日常生活场景，不含有明显的时代、地域特征以及成语或特有名词。未对语句的音位构成，词汇或句式结构等语言学特征进行限制。

邀请经过专业发声训练、普通话甲级一级的播音主持专业学生2男2女，年龄20~23岁。在本底噪声≤20dB(A)的隔声室中，以生活化口语风格播音，语调和逻辑重音要自然，避免过度强调和夸张的韵律。使用b&k2250声级计(配装4155声场传声器，罩以防风球，其位置离播音人的口唇约30cm)，将语音以.wav格式记录在其内置的SD卡中，采样速率为44.1kHz，分辨率为16bits。同步将语音实时显示在Adobe Audition 3.0软件界面，监控播音人的声强、语速及意外状态。如果出现错漏字，杂音或任何中断，播音人会被要求重复该语句。一名男性播音人录制了350条语句，其他三人各录制了500条语句。

应用Adobe Audition 3.0软件，将录音切分成一句句的.wav文件，并将每句电平归一化，以获得相同的均方根(root-mean-square，RMS)量级。具有相同RMS量级的校准音也一并生成。4位富有语言测听经验的听力师逐句审听所录语音文件，剔除有口误、音质差或语义不自然的语句，最后选出1020句(每名播音人255句×4人)。各播音人的平均语速为每秒4.7~5.7个音节，与英语AzBio原句相似。均衡话题，句式等因素，从4人播音中各选取5句构成一张张20句的测试表，共计51张表。另备80句以同样方式编选成4张练习表。采用关键词计分方式，每句含4~9个关键词，平均(6.5±1.0)个。(x ± s)。

使用付前杰提供的AngelSim™软件，选中语音作为载波，对上述语句进行声码器( vocoder)处理，模拟5通道CI的信号处理。声码器处理时将总带宽为200~7 000Hz的语音信号分割成5个通道，频段间隔根据Greenwood公式确定(90~190Hz，每个频段的包络通过半波整流及低通滤波进行提取，低通滤波的截止频率为160Hz。提取的5个包络分别对5个带通滤波后的噪声进行调制，带通滤波的频段间隔与前述一致。最后，两个通道的噪声相加起来，合成的信号便携带有5个通道的时域包络信息，送往耳机，听力正常人便可听到CI的声学模拟。5通道噪声声码器仿真处理曾用于其他语种AzBio句表开发研究(15，16)。我们的预实验也表明，它对听力正常的普通话母语者也可有效避免天花板效应。将处理后的仿真语句，再串接成51张实验用表和4张练习表。每表20句，每间隔4s播放一句。

2. 声码器仿真5通道CI的语音识别测试：招募17名母语为普通话的成年受试者，男9人，女8人，年龄(25.6±1.4)岁(x ± s)，无耳科疾病史，耳镜检查正常，日常交流均使用普通话，双耳在250~4 000Hz各频听阈频率的听阈均≤10dBHL。

笔记本电脑播放经声码器处理后的语句，送至GSI 61听力计，以70dB SPL强度由ER-3A插入式耳机送入受试者的双耳。受试者聆听并尽可能地重复语句中的每一个字词。正式测试前受试者须至少完成3张练习表，以使其熟悉仿真后的语音。了解整个测试流程并能积极配合。为均衡疲劳，练习效率等干扰，以拉丁方实验设计确定每名受试者
的句表聆听顺序。测试在 3 d 内分 3 次接续完成，每次测 17 张句表，约持续 45~60 min。

3. 统计分析及语句遴选：以关键词计分，计算 17 名听力正常受试者听取仿真处理后 1020 句的识别率(77.0±15.1)% (x±s)，范围为 9.2%~100% (图 1)。从每个播音人的语句中选出正确识别率在其均值加减一个标准差范围内的 150 句。4 个播音人(男 1、女 1、男 2 和女 2) 每人 150 句识别率的总体正态分布更符合正态分布的条件，以每个播音人识别率的总体均值为正态分布的均值，得其标准差可粗略保证各表难度大体一致。20 句(100%、78%、77%、76% 和 75%) 的识别率得分在 96%~100%。平均 (65.5±13.9)% (x±s)。

三、测试条件

受试者处于隔声室中央，扬声器位于受试者正前方 1m 左右，同其坐姿时耳部齐高。由 GSI 扬声器播放语句，给声强度为 65 dBSPL。要求 CI 使用者描述所听到语句中的每一个字词，听不清或语义不连贯时大胆猜测。以关键词为分项，计算每句乃至每表的正确认识概率。在安静和 +10 dB 信噪比条件下进行的语句识别率预实验显示，上述受试者无一人能在 +10 dB 信噪比的噪声背景下完成本语句材料的测试。因此实验只在安静条件下进行。

四、实验流程

正式测试前，受试者须完成 3 张练习表(每表 20 句，不包括在这 30 张测试句表中)。每完成 10 张正式测试句表，安排受试者至少休息 15 min。为避免学习效应及疲劳效应等干扰因素带来的偏倚，句表测试顺序在不同受试者之间进行了均衡。整个测试(包括休息时间)历时约 3 h。二、成年 CI 受试者

招募 30 例成年植入或自幼植入现已成年的单侧 CI 使用者，男 10 例、女 20 例，平均年龄 (32.1±12.7) 岁。植入前确诊为感音神经性听力损失，实验时至少使用 CI 装置 12 个月以上并已融入主流社会。音位平衡的普通话单音节表高识别率得分在 40% 以上者方可纳入。最终入选的 30 例 CI 受试者的最高单音节识别率为 44%~96%，平均 (65.5±13.9)% (x±s)。

五、统计学处理

对所有识别率分值进行了“合理化”的反正弦 (RAU) 变换后，采用 SPSS 18.0 统计软件，进行单因素重复测量的方差分析以检验 30 张表的等效性，P<0.05 为差异具有统计学意义。

结 果

一、CI 受试者识别率分布情况

30 例 CI 受试者均能完成测试，但 30 张表平均后的识别率得分有差异，介于 43.8%~89.3%，平均 (65.0±13.0)% (x±s)。图 3 按照从低到高的顺序逐条显示了 30 例受试者识别率得分的分布情况；30 例中只有 2 张的总体得分在 85% 以上；同时可见 30 张表在 30 例中的识别率变异范围介于 9.3%~36.9%，30 张表表识别率的标准差为 2.7%~8.8%。
三、表间等价性验证
图4所示为30张句表的标准化RAU值，对其进行单因素重复测量的方差分析以检验30张表的等效性。结果显示，表号存在显著的主效应（$F_{29,29}=3.08, P<0.001$），即30张表之间存在显著的差异。Post Hoc Tukey检验显示，有4张表与其他至少一张表有显著差异，虽然差异不大，但表4和表29与其他表更难，而表5和表10更容易。拆除这4张句表后，对所余26张表再进行单因素重复测量的方差分析，它们之间似乎依然存在统计学上的差异（$F_{25,25}=1.68, P=0.02$）。但Post Hoc Tukey检验结果显示，没有任何表与其他任一张表有显著差异，26张表的平均识别率为62.6%~67.0%（$\pm$1.2%）。因此这26张句表被确定为汉语普通话CMnBio句表的等效版。句表平均11.9个字，9~16字不等。

四、以二项分布模型估计变异度

言语识别率测试中，患者的应答要么判定为“对”要么为“错”，故Thornton和Raffin指出，“假设受试者对各测试项的应答是彼此独立的，则测试结果呈现二项分布，可据此进行统计推断”，并提出了一个二项分布的数学模型来描述测试项数目（n）与变异度的关系。变异度（用标准差$\sigma$表示）是识别率得分与测试项数目的函数。得分50%对应的变异度最大，得分极低或极高的变异度最小。测试项数目（n）越多，得分的变异度就越小。Thornton和Raffin给出了若干n值条件下各言语识别率所对应的5%置信区间的上、下限数值对照表。治疗/干预前后的测试得分须超出此区间，差异才可能确认不再是由随机误差构成的，可以得出治疗/干预有效的结论。

语言测试时受句法、语法及上下文线索的影响，句中的关键词并非完全独立，所以需要依据实测结果来反推句表的有效测试项数目。Spahr等提出了英语AzBio句表的变异度参考范围（$\pm$1.96$\sigma$），预测了言语识别率得分的95%置信区间
对数据的拟合效果好，只有少量散点超
之间。结果表明，以n=40构建的模型能覆盖15例CI
受试者中95%置信区间的数据。由此该作者认为，
英语AzBio句表得分的变异度能以相当于40个独
立测试项的二项分布模型来估计[25]。西班牙语版
AzBio句表也报告了类似结果[26]。

本研究首先采取与英语和西班牙语版AzBio
相同的方法[25, 26]，将30例CI受试者得分的标准差
绘制在基于n=40的二项分布95%置信区间的参考
范围内。结果见图5左图，虚线是n=40对应的标
准差平均值，实线是用引法法(bootstrap)获得的变
异度的预测。每个散点的横纵坐标就是每个CI受试者
26张CMnBio句表识别率的均值和标准差。图5左
图显示CMnBio句表的变异度较小，30例CI受试者
中约2/3的标准差数据低于n=40二项分布所预测
的参考范围，因此n=40并不适合CMnBio句表关于
变异度的预测。我们进一步增大二项分布模型中
的测试项数目，目的当二项分布模型中的n在65~
80之间时，对数据的拟合效果好，只有少量散点超
出参考范围。图5右图显示了30例CI受试者的变
异度数据散点及n=65时二项分布模型95%置信度
下的参考范围。

接下来，我们选择另一种方法，再次确认
CMnBio句表中的有效测试项数目。根据30例CI
受试者对最终26张句表的句词识别率，我们首先
创建了任意单张句表得分的配对散点图（图6左图）
。对应这26张句表，就有26×25/2=325个可能的配
对，因此散点图有975 0个数据点（235×30例CI受
试者），句表的配对顺序是随机的。假设我们在同
样条件下一次测了双表并得到合计识别率，那么
就有14 950(C%)种双表的配对。为了与单张表的
条件相一致，我们随机选择了325个双表合计分
值，图6（中图）显示了在双表条件下的9 750个数
据点（即325×30例CI受试者）散点图。随后我们根
据Thornton和Raffin[34]的二项分布模型计算了识别
率分数的95%临界差值的下限和上限。临界差值
取决于测试项数目(n)。对于CMnBio句表，每表有
20个语句，平均130个关键词。句表中的关键词并非完全独立，最极
端的情况是整句中各关键词彼此牵
制，要么全对要么全错，只能以整句
作为一个计分项；另一极端则是所
有关键词都是彼此独立的。所以
CMnBio语句测听的有效测试项数
目介于20~130。我们尝试以5为步
距将n=20~130时的临界差值的下
限和上限绘制在单表得分的配对散
点图上，10为步距将40至260时
临界差值的下限和上限绘制在双表
合计分值的散点图上。最后，我们
计算数据散点落在n的每一步距所
对应的临界差值参考范围之外的百
分比，以找寻能使约5%的数据点超
出临界差值时的测试项数目，见
图6（右图）。图中双表条件下的数
据被叠印到了单表条件下的数据
上。结果表明，二项分布模型的n值
在单表条件下为65（图6左图），
在双表条件下为130（图6中图），即
每表n=65的二项分布模型最能反
映汉语普通话AzBio句表的变
异度。

表1列出了CMnBio句表识别

图 4 30张CMnBio句表的标准化反正弦(RAU)分数 句表依据在30例受试者中的总体制表率标准化RAU分数，从低到高排序。每个小提琴图显示了30例受试者识别率数据的概率密度，每个灰点代表一例受试者的识别率，红色的方形和粗垂直线表示每个句子列表的平均值±1倍标准差

图 5 用目测引算法来估计二项分布模型中的n值变异度(用标准差σ表示)是识别率得分和测试项数目n的函数。用基于二项分布的引算法得到预测的σ（虚线）及其95%置信度下的参考范围（实线）, 每个散点的横纵坐标就是本研究中每名人工耳蜗受试者26张CMnBio句表识别率的均值和标准差
率得分基于Thornton和Raffin二项分布模型的95%临界差值的下限和上限，分数的步为5%。查阅该表，对临床医师正确理解言语测试分有指导意义。例如，若某张表的识别率为50%，而用另一张表测得的识别率在33%~67%之间（即＜17%的差异），则两次分差不能认为有统计学意义。然而，若使用两张语句进行测试而患者合计识别率仍为50%，那么12%的差异（即＜38%或＞62%的识别率）就被认为差异具有统计学意义。

讨 论

与单音节识别测试相比，语句材料更能代表日常交流，并体现自然语音流中的协同发音和语调，因此更能反映听者在日常生活中的语言理解能力。参照英语HINT语句，Wong等在早开发了普通话版MHINT（Mandarin HINT），由12张语句组成，每张20句，语句的内部变异性低且每句等价性高，并提供了安静和噪声下测试的语句识别阈常模数据。Fu等在为C1受者设计汉语语句测试时，采用C1仿真处理后的语句进行验证，构建了100张音素平衡语句，每表包含10句，每句每个字，命名为MSP（Mandarin Speech Perception）语句测试。Xi等建立了一个在四种语言背景下的普通话版BKB语句库，并经心理测量学评估实现了同质性优化，可用于儿童和成人C1用户语句识别测试。近来Hu等还开发了一种矩阵式汉语普通话（CMNmatrix）语句测试，由一组句法固定但语义上不可预测的陈述语句构成，可用于安静及噪声下言语识别测试。

应用现有的中文语句测听材料评估成人C1患者的语句识别率时，常出现天花板效应。王炳等使用MHINT和普通话BKB语句在安静条件下测试了32例后聋性C1受试者的语句识别率，分别有4例和14例受试者的MHINT和普通话BKB语句识别率达到100%，超40%的受试者在这两项测试中的识别率均达到85%以上。Li等在使用MSP语句测试时也发现，35例C1成人中表现居前1/3的受者在安静条件下的识别率也达到90%以上。天花板效应也同样见诸于英语为母语的C1成人。Gifford等发现，多达71%的成人C1受者在安静下HINT语句识别率的95%不临界常模低于80%。

表1 CMNBio语句识别率得分对应的95%临界差值的下限和上限值

| 得分 | 使用单张表时(n=65) | 使用双张表时(n=130) |
|------|-------------------|-------------------|
| 0    | 0                 | 0                 |
| 5    | 0                 | 1                 |
| 10   | 2                 | 4                 |
| 15   | 6                 | 8                 |
| 20   | 9                 | 12                |
| 25   | 13                | 16                |
| 30   | 16                | 20                |
| 35   | 20                | 24                |
| 40   | 25                | 28                |
| 45   | 30                | 33                |
| 50   | 33                | 38                |
| 55   | 38                | 43                |
| 60   | 43                | 48                |
| 65   | 48                | 53                |
| 70   | 54                | 58                |
| 75   | 59                | 64                |
| 80   | 65                | 70                |
| 85   | 71                | 76                |
| 90   | 78                | 82                |
| 95   | 85                | 88                |
| 100  | 95                | 98                |

图6 表示变数四分布模型 左图；30例C1受者975次句(325×30)26个CMNBio语句中两两句表识别率分值的配对散点图。横竖代表了基于Thornton和Raffin二项分布模型的95%临界差值的下限和上限，测试项数目n=65；右图；30例C1受者975次句(325×30)句表识别率分值的配对散点图。横竖代表了基于Thornton和Raffin二项分布模型的95%临界差值的下限和上限，测试项数目n=130。
语句的识别率能高达85%以上，28%的受试者甚至能达到100%。Bassim等[9]也报告，在植入1年后，语后聋CI受试者的CUNY语句和HINT语句测试中的平均识别率分别能达到96%和87%[38,39]。

因此，有必要为植入成效良好的成人提供更高难度的言语测试材料。英文AzBio语句提供了一个很好的范例[21]。它由亚利桑那州立大学开发，初衷是用于比较不同品牌CI高水平植入者的言语识别能力[38]。其语句由2男2女以日常生活语气录制，语句较长，且上下文线索有限。原本包含33张句表，每表20句，经15名CI成人进行语句识别率测试，确认其中29表等价。Gifford等[20]报告156例语后聋CI使用者中，只有1例的AzBio语句识别率能达到100%。由于AzBio语句具有避免天花板效应的优势，因而除了15张等价的句表被编为AzBio语句测试日独立出版和发售外，等价性居于其后的8张句表被纳入成人CI植入者最简言语测试集[21]。

本研究依据类似原则和流程，开发并验证了汉语普通话版本CMnBio。

一个面向成人的言语测听材料需要满足以下条件[39,39]：(1) 难度较高，避免受试者尤其是语后聋及成人的语前聋植入者受到天花板效应的影响；(2) 句表长度较多，能够避免反复使用少数几张表后患者记忆效应的影响，确保长期有效地对植入者康复效果进行评估；(3) 不同测试表之间难度等价，保证长期评估受试者的可靠性；(4) 测试语料应尽量接近日常生活中的交流场景。

一、较高的测试难度

为了达到较高的测试难度，语句由4名播音人（2男2女）以生活化口语风格播音，而不是像其他大多数材料那样字正腔圆的“播音腔”。开发CMnBio句表时，语句结构的复杂性和对词汇量的要求都比普通话BKB语句高[38]。虽未与MHINT语句进行详细的语言学比较[41]，但二者在句长上存在明显差异。MHINT每个句子都是10个汉字，而CMnBio语句长度虽不固定（9~16字），但平均接近12个字。CMnBio语句的较高难度，从正常人聆听5通道声码器仿真语句的识别率中就能明显看出。30张CMnBio句表的总体平均识别率78.0%（图2），而本课题组此前在同样的5通道声码器条件下测试，正常人对普通话语句的识别率则可轻松达到90%以上。最近我们还比较了声码器仿真CI语句和HINT和CMnBio语句的识别率[40]。在5通道噪声声码器条件下，两类语句材料的识别率分别为93.4%和73.9%。显而易见，普通话CMnBio语句的难度已经高于MHINT和普通话BKB语句。同时，本研究招募的30名CI受试者，入组前要求至少有中等程度的言语识别能力（单音节识别率>40%），但CMnBio识别率所得的2例也只介于85%~90%（图3），说明CMnBio语句能有效避免安静条件下语句测试的天花板效应。当然，针对预成效不佳的成人患者，仍可以使用相对容易的其他语句测试材料。我们后续会在成人CI患者中进行难度验证，尽可能以循证方式来准确界定CMnBio句表的适用人群及其言语理解能力和认知水平。

与更早期开发的其他语种AzBio语句相比，CMnBio语句的难度也略高。本研究中，17名母语为普通话的听力正常成人，聆听5通道噪声声码器仿真语句的识别率78.0%。而使用类似方式，英文AzBio语句的识别率为85.2%[21]，西班牙语AzBio语句的识别率为85.0%[39]，希伯来语HeBio语句的识别率为82.0%[21]。此前我们比较过英语和汉语版本AzBio语句测试，发现CMnBio语句识别率比英文AzBio语句低9%~12%[40]。CMnBio相较于AzBio，完全没有天花板效应，而AzBio对个别明星CM使用效果却不一定好。Bergeron等[20]发现，所有非法语的听力正常人在使用5通道声码器仿真语句的识别率均为100%，因此该研究最终选择用5通道声码器仿真方案，仿真语句的平均识别率约为85%。普通话CMnBio语句似乎难度最高，这可能是由于汉语语法结构松散、字与词的边界模糊、“声调”具有表意作用而导致同样音节数量的句子比拉丁语系具有更丰富的语义[41]。总之，虽然英语、西班牙语和希伯来语版本的AzBio语句显示出相似的难度，但中文普通话版和法语版与它们存在差异，故在进行CI效果的语前语句比较时，应注意这一难度上的差异。

二、较多的句表数量

CMnBio语句材料的另一个特点是句表数量更多。最终的CMnBio语句集有26张测试表和4张练习表。这个数量比英语（33张句表）、法语（30张句表）、西班牙语（42张句表）和希伯来语（33张句表）略少[23,24,25]，但比MHINT语句（12张句表）多很多[41]。这一特点使得实施涉及多种实验条件的大
规模研究成为可能，因为这些研究常常须使用不重复的句表以避免受试者记忆的介入。

三、较高的表间等价性

经验证，CmBio 句表具有较高的表间等价性。30 例 CI 受试者测试 30 张表句的平均识别率为 62%~69% (标准差=1.7%)。相比之下，Spahr 等[23]观察到 15 例 CI 受试者对 33 张英语 AzBio 句表的平均识别率为 62%~79% (标准差=3.8%)。CmBio 句表较小的表间差异，对临床实践极为有利，它为观测患者在多个时间，多种条件下识别率的变化提供可靠工具。

句表间较大的变异也体现在 30 例 CI 受试者对最终 26 张句表识别率的标准差相对较小(图 5)。使用二项分布模型进行分析的结果显示，英语和西班牙语 AzBio 语句材料对应的 n=40 的测试项数目并不能符合 CmBio 语句材料[23,26]。当我们将 n 增加到 65~80 时，目测二项分布模型才能符合实验结果。使用量化方法来推测句表中的独立测试项数目时，发现 n=65 更吻合 CmBio 句表(图 6)。每张 CmBio 句表约含 130 个关键词，而 n=65 恰好是关键词数量的一半。对包含 50 个关键词的澳大利亚英语版 BKB 句表，Keidser 等[42]也曾预言，关键词数量的一半 (25 个) 是对测试项数目恰当的估计值。值得注意的是，英语或西班牙语版 AzBio 句表的计分是基于单词数而不是关键词数，平均单词数为 142 个，而 CmBio 的计分是基于关键词。这或许可以解释为什么本研究估计的独立测试项数目 (65) 比英语或西班牙语 AzBio 句表材料中的独立测试项数目 (40) 多。

另一个重要的区别是本研究使用的推导临界差值下限和上限的方法不同。本研究完全参照Thornton 和 Raffin[44] 的方法和术语得出 95% 临界差值的下限和上限，而 Spahr 等[23]和 Rivas 等[26]则使用二项分布置信区间的正态近似法来推导 95% 置信区间的下限和上限。事实上，我们在测试项数量为 65 和 130 时得出的下限和上限 (表 1) 与他们在测试项数量为 40 和 80 时得出的界限非常相似。因此，不同研究中报告的独立测试项数目即使不尽相同，也不一定意味着语义信息难度在多语种 AzBio 语句中有多大不同。在临床上，只要得出的临界差值是准确的，独立测试项的确切数目可能并不重要。在本研究中，临界差值是基于 30 例具有良好语音识别能力的成人 CI 受试者的数据计算出来的。后续还需要在人口学特征不同的 CI 人群中进行大规模的验证，为未来的临床使用建立更加可靠的临界差值。

四、贴近真实交流场景

CmBio 句表的编撰，本身就选自影视剧及社交媒体中的对话，以日常口语化的方式播音，克服了以往测听句表的播音腔。同时编排上又将男 1、女 1、男 2、女 2 四名播音人的语句交替呈现，使得测试过程不再单调呆板，更贴近现实生活的生活场景。但针对内容效度、结构效度方面的分析，尚需与汉语普通话语音能力测试专家进行后续探讨。

总之，CmBio 句表测试是基于英语 AzBio 语句测试而开发的汉语普通话版语句测听材料，有 26 张表，每表 20 句，其表间等价性 (复本信度) 在成人 CI 使用者中得到了证实。有关复测信度、内部一致性信度和评分信度的验证，我们今后将在更大样本的成人 CI 患者中进行。与英语、西班牙语、法语和希伯来语的类似语句测试材料相比，汉语普通话话材料的难度似乎略高，这种难度的提升有助于更久远地避免成人语句测听时面临的天花板效应。我们基于二项分布建立模型出了识别率得分的 95% 临界差值的下限和上限，为临床判定某患者干预前后识别率分差是否有临床意义提供依据，有助于 CmBio 语句在 CI 植入成人言语感知能力临床评估和研究中的应用。

利益冲突 所有作者声明无利益冲突

参考文献

[1] 胡向阳，郑晓瑛，马芙蓉，等. 我国四省听力障碍流行现况调 查[J]. 中华耳鼻咽喉头颈外科杂志, 2016, 20(11): 819-825. DOI: 10.3760/cma.j.issn.1673-0860.2016.11.004.
[2] Liang Q, Mason B. Enter the dragon--China’s journey to the hearing world[J]. Cochlear Implants Int, 2013, 14(Suppl 1(Suppl 1)): S26-S31. DOI: 10.1179/146701013X134000080.
[3] Zeng FG, Rebscher SJ, Fu QJ, et al. Development and evaluation of the Nurotron 26-electrode cochlear implant system[J]. Hear Res, 2015, 322: 188-199. DOI: 10.1016/j.heares.2014.09.013.
[4] Zeng FG. Celebrating the one millionth cochlear implant [J]. ASA Express Lett, 2022, 2(7): 077201. DOI: 10.1121/10.0012825.
2023年本刊可以直接用缩略语的常用医学词汇（耳科）

为节约篇幅，以下临床医师比较熟悉的常用词汇在本卷文章的正文及摘要中首次出现时给出中文全称和缩略语即可，不再要求给出英文全称。

毛细胞（hair cell，HC）
内毛细胞（inner hair cell，IHC）
外毛细胞（outer hair cell，OHC）
声压级（sound pressure level，SPL）
听力级（hearing level，HL）
峰值等效声压级（peak-equivalent sound pressure level，peSPL）
正常听力级（normalized hearing level，nHL）
预估听力级（estimated hearing level，eHL）
感觉级（sensation level，SL）
总和电位（summing potential，SP）
动作电位（action potential，AP）
耳蜗微音器电位（cochlear microphonics，CM）
听性脑干反应（auditory brainstem responses，ABR）
脑干听觉诱发电位（brainstem auditory evoked potential，BAEP）
自动听性脑干反应（auto auditory brainstem response，AABR）
电刺激听性脑干反应（electrical auditory brainstem response，EABR）
短声诱发听性脑干反应（click auditory brainstem response，click-ABR）
短纯音诱发听性脑干反应（tone burst auditory brainstem response，tb-ABR）
听觉稳态反应（auditory steady-state response，ASSR）
皮层听觉诱发电位（cortical auditory evoked potential，CAEP）
耳声发射（otoacoustic emission，OAE）
畸变产物耳声发射（distortion product otoacoustic emission，DPOAE）
瞬态声诱发耳声发射（transient evoked otoacoustic emission，TEOAE）
纯音测听（pure tone audiometry，PTA）
视觉强化测听（visual reinforcement audiometry，VRA）
游戏测听（play audiometry，PA）
良性阵发性位置性眩晕（benign paroxysmal positional vertigo，BPPV）
前庭-眼反射（vestibular-ocular reflex，VOR）
前庭-前庭反射（vestibulo-collic reflex，VCR）
前庭诱发肌源性电位（vestibular evoked myogenic potential，VEMP）
颈肌前庭诱发肌源性电位（cervical vestibular-evoked myogenic potential，cVEMP）
眼肌前庭诱发肌源性电位（ocular vestibular-evoked myogenic potential，oVEMP）
视频头脉冲试验（video head impulse test，vHIT）
视频眼震图（video nystagmography，VNG）
高分辨CT（high resolution computed tomography，HRCT）
锥形束CT（one beam computer tomography，CBCT）
弥散加权成像（diffusion weighted imaging，DWI）
功能性磁共振成像（functional magnetic resonance imaging，fMRI）
前庭导水管扩大（enlarged vestibular aqueduct，EVA）
大前庭导水管综合征（large vestibular aqueduct syndrome，LVAS）
梅尼埃病（Meniere’s disease，MD）
前庭神经炎（vestibular neuritis，VN）
分泌性中耳炎（otitis media with effusion，OME）
听神经病（auditory neuropathy，AN）
人工耳蜗（cochlear implant，CI）
部分听骨赝复物（partial ossicular replacement prosthesis，PORP）
全听骨赝复物（total ossicular replacement prosthesis，TORP）
骨锚式助听器（bone-anchored hearing aid，BAHA）
振动声桥（vibrant soundbridge，VSB）

本刊编辑部