

当前中文言语测听的几个误区

郝昕¹

DOI:10.3969/j.issn.1006-7299.2012.06.001

网络出版时间:2012-11-01 12:50

网络出版地址: <http://www.cnki.net/kcms/detail/42.1391.R.20121101.1250.023.html>

【中图分类号】 R764.04 【文献标识码】 A 【文章编号】 1006-7299(2012)06-0509-03

中文言语测听材料历经 60 年的曲折发展,近十年来取得了长足的进步^[1],建立起了成人普通话言语测听的完整体系,基本能满足日常临床言语测听的需求。但在全国推广中文言语测听的过程中,笔者发现许多耳科医师、听力师对言语测听的基本概念和临床意义存在着一定的误区。

1 纯音测听与言语测听的关系

人类听觉最重要的功能是接收和理解言语,如果只能选择一种测试来评价听觉功能,那就一定是言语测听^[2]。换句话说,受试者的言语识别能力是评判其听功能的金标准。

众所周知,人类的言语是一种其强度和频率都随时间不断变化着的声信号,其频率范围约为 100~6 000 Hz,且世界各语种的长时平均会话语谱都大体一致^[3](图 1)。但正是由于世界上(从高度文明的西方社会到尚处荒蛮的原始部落)存在着多种语言,使得听力学家要寻求更易于为全人类共同使用的测听方法,这就催生了纯音测听技术。纯音信号易于标准化,国际上已形成了一套完备的技术标准和测试规范,因此可以作为各种客观听阈检查(如 ABR、ASSR)的参照系。但必须意识到,纯音测听只是测试了言语频率范围内若干个倍频程频率处的听阈,它并不能反映言语识别过程中的诸多细节,纯音听阈并不一定与言语识别能力相匹配,否则就不会有今天大家都在热议的“听神经病谱系障碍”了。

回顾一下听力学的早期发展史,就可以很明晰地看到:言语测听与纯音测听技术始终并驾齐驱。二战结束后,以英语为主的言语测听在评估退伍老兵的听力伤残和选配助听器时占有重要地位,在美国、英国得到迅速发展并走向规范化。随后的 60 多年,围绕言语测听的研究热潮一直没有降温,各国基于英文测听范式开发的多语种言语测听日渐丰富^[4]。我国幅员广阔、民族众多、方言庞杂,开发中

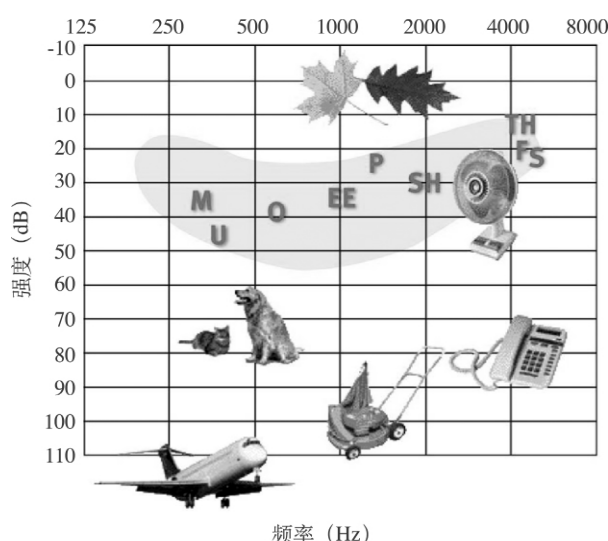


图 1 长时平均会话语谱在纯音听力图上的分布呈香蕉状,故俗称香蕉图

文言言语测听材料的任务十分艰巨。自 20 世纪 50 年代尝试开展中文言语测听以来,一些语音学家、心理学家、耳科医师和听力学工作者倾注了大量心血,在 21 世纪的第一个十年终于初步建立起中文普通话言语测听的标准化体系^[5]。

2 标准化不等于惟一化

随着人工耳蜗等听觉康复技术在我国广泛开展,学术界、政府及慈善家都希望能真切地了解助听器、人工耳蜗的康复成效,为规范临床诊疗实践、规划国家中长期的残疾人辅助政策提供依据^[6]。由于认识到开展中文言语测听的紧迫性,国内以解放军总医院^[7]、北京同仁医院^[8]为代表的多家研发团队已先后研发出不同版本的单音节字、双音节词和短句测听材料。这本是十分可喜的局面,但许多单位在启动言语测听临床检查项目时却首先遇到了一个困惑——如何选择恰当的测试材料?

仅就单音节表而言,甲单位每表 25 个字并正式出版了 CD,乙单位每表 50 个字并进行了临床验证,丙单位仅仅是邀请播音员照着前人的字表念了一遍;就短句表而言,甲单位可在安静及噪声下测试,

1 解放军总医院耳鼻咽喉头颈外科(北京 100853)

同时提供发音人的视频,允许在视、听、视+听三种模式下进行,乙单位在同一张语句表中的句子有长有短……由此不少耳科医师建议中华耳鼻咽喉头颈外科学会应规范中文言语测听,建立起一套全国通用的标准化的言语测听材料。

但事实上,追溯标准化的起源,典型的事例之一是:当初不同的工厂都在各自生产螺钉与螺母,但由于缺乏统一的规格尺寸,A厂生产的螺钉,无法装配到B厂生产的螺母上,导致极大的浪费。后来国际标准化组织制订了螺钉与螺母的系列性的标准尺寸(直径、螺距等),众厂家都采用此系列标准,彼此的产品就可以通用了。从上述例子可以看出,标准化并不意味着螺钉与螺母只能由一家生产,只要出厂的螺钉与螺母符合标准尺寸,该工厂的产品就可以被广泛使用。

3 如何定义言语测听的标准化

言语测听材料可以看作是一个测量工具,借助它可以测试患者的言语识别能力。而言语识别能力是软性指标,只有理论上的识别率,量程范围为0%~100%。既然言语测听材料是一个测量工具,它就涉及到测量效能(称为效度)、结果的可靠程度(称为信度)、区别差异的能力(称为敏感度)等问题^[9]。

3.1 效度 所谓效度,反映的是测量工具的有效性,即该测试工具是否能真实测量出需要测试的内容。具体到言语测听,效度考察的就是使用该言语测试材料能否准确地反映出患者言语识别能力的残障程度。从语前聋儿童到语后聋成人,其言语认知能力跨度很大,这就要求针对不同的测试对象或测试目的开发多种难度水平的言语测听材料。换句话说,不能用中学生的试卷来考小学生,否则无论优劣,小学生的得分都会很低;相反,也不能拿小学生的试卷来考中学生,否则能力平庸和能力出众的中学生都会得高分,无法真实地反映他们的能力差异。

对新开发的言语测听工具的效度进行评价,可以选用一个为业界公认的言语测听“金标准”,将两者的测试结果进行平行对照,以印证新开发的测听材料的有效性。但当业内尚未确立公认的言语测听材料时,多采用表面效度的方式进行。表面效度评价的是测量方法或观测结果所说明的问题是否符合专家和公众共识^[10]。

3.2 信度 信度反映的是测试结果的可靠性。若只是由于担心受试者对测听语料存有记忆而在前后两次测试中采用了两张彼此等价的测试表,其它所有测试条件都保持不变,而受试者在前后两次测试时得分的变异度(variability)很小,则可以认为该测听工具具有良好的可靠性。测试材料的可靠性,蕴

含了两方面的内容——得分在多次测试中的稳定性和多张表之间的一致性。当用同一张表重复测试同一组受试者时,希望前后两次得到的分数是稳定的,可用复测信度^[11]来描述。当用同一套测试的多个词表测试同一受试者时,希望各个词表的得分都是一致的,可用复本信度(也常被称为表间等价性)来描述。

有关信度指标可分别在听力正常及听力减退的人群中进行重复测量,只要其得分的变异度能与统计学上的随机误差的量级大体对应,并符合临床上对测量误差的宽容度^[12](如纯音听阈可允许±5 dB的误差),就是符合信度要求的测试工具。

3.3 敏感度 敏感度反映的是测量工具所能区分出的最小差异值。当改变测试条件(如言语强度、信噪比)时,受试者言语分辨能力上的“实际”差异,若能通过某一言语测听工具反映出来,则该测听工具是敏感的。

言语测听材料的敏感度通常体现为识别率随言语声级(或信噪比等)增减而出现的分值变化,即识别率—强度(P-I)函数的斜率。采用敏感度高的测试材料,则言语声级(或信噪比)上的稍许提高就足以使识别率的提升量超出随机误差的波动范围,而易于得出可靠的结论。因此言语测听材料在编制过程中应追求较高的敏感度。

用另外一种测量工具——温度计为例,可以更生动地说明效度、信度和敏感度指标在标准化中的意义。在市场上可以买到不同厂家生产的多种温度计——水温计、气温计、体温计,它们的特征见表1。针对不同的测量目的,三种温度计采用了不同的材质并具有特定的量程范围和长短不一的外形;也为了测量可靠性的需要,出厂前均经过质量检验,保证了测量误差在允许的范围之内。水温计本也可以用来量体温,但为什么还要生产专门的体温计呢?关键的原因在于两者的敏感度差异很大。体温每升高1℃,体温计中的水银柱会增0.6 cm,而水温计中的酒精柱只增0.2 cm,对人眼常规目视能力而言,显然体温计对微小的温度变化是最敏感的。只要满足上述测试效能(效度)、测量可靠性(信度)和区分度(敏感度)的要求,温度计可以由不同的厂家生产。

表1 各类温度计的量程范围(效度)、误差(信度)与敏感度

功用	材质	刻度长度 (cm)	量程范围 (℃)	误差 (℃)	敏感度 (cm/℃)
体温计	水银	4.2	35~42	0.1	0.6
气温计	酒精	15	-10~40	1	0.3
水温计	酒精	24	-10~110	1	0.2

4 未经标准化验证的言语测听材料应停止使用

如上所述,标准化的言语测听材料应是经过临

床验证的、符合效度、信度和敏感度指标的测试材料。不论它是 A 单位还是 B 单位开发的,也许它们都是单音节表,只是在编排上、测试项的数量上存有差异,但只要是经过了效度、信度和敏感度验证的测听语料,它们就都是标准化的言语测听材料。

随着我国听力学的发展,特别是助听器、人工听觉植入技术的广泛开展,面对不同年龄、认知水平的儿童及成人患者,面对不同言语康复阶段的聋儿,还需要花大力气研发更多的标准化言语测听材料。但在制订某一具体的研究方案时,比如开展全国多中心的国产人工耳蜗研究时,就应针对测试对象(成人或儿童、儿童的认知和言语能力)和实验目的(成效评估或中文编码策略改进)选定某一种最适宜的标准化言语测听工具。

必须充分认识到研发标准化中文言语测听材料的艰巨性。卜行宽教授^[13]曾说,在科技史上恐怕很难找到一个像中文言语测听一样经历了 50 年仍然未能较好推广的项目了。原因正如从事了 60 年中文语言声学研究的张家骥研究员所言,敢于从事言语测听材料的开发,是需要极大的勇气和耐心的。仅就测听材料的文字编撰而言,就涉及到语言学(音系学、语音学、词法、句法、语法、语用学)、心理学(儿童发展、心理测量)、统计学(测试项的数目及同质性)等多学科的知识^[14]。录音及后期处理又涉及到许多声学或信息技术。测听材料的表间等价性、测量的随机误差范围等性能指标,则需要经过严格的多中心的临床实验的验证^[15]。但遗憾的是,某些进口听力计的厂家,在缺乏理论根基的情况下,只是请播音员将前人编制好的词表文字稿录了音,未经任何效度、信度和敏感度认证,就将其扩充到号称具有内置式中文言语测听的听力计中,实在是“无知者无畏”的行为,应引起学界的高度警惕。

围绕言语测听材料的研发工作,各单位应扬长避短、相互切磋,合理布局、避免撞车。建议由各级与听力相关的学会,通过举办学习班、网络教学等渠道,大力推广标准化的中文言语测听,厘清基本概念

和澄清错误认识,使中文言语测听真正能够成为可与纯音测听比肩的常规测试项目。

5 参考文献

- 1 郝昕. 中文言语测听材料的新进展[J]. 中国眼耳鼻喉科杂志, 2008, 8: 341.
- 2 Hall JW, Mueller HG. Speech audiometry[M]. In: Hall JW, Mueller HG, eds. Audiologist's Desk Reference. San Diego: Singular Publishing Group, 1997. 115~174.
- 3 Byrne D, Dillon H, Tran K. An international comparison of long-term average speech spectra[J]. J Acoust Soc Am, 1994, 96: 2 108.
- 4 Wilson RH, McArdle R. Speech signals used to evaluate functional status of the auditory system[J]. J Rehabil Res Dev, 2005, 42(4 Suppl 2): 79.
- 5 亓贝尔, 张宁, 刘博. 中文言语测听材料概述[J]. 中华耳鼻咽喉头颈外科杂志, 2012, 47: 607.
- 6 郝昕, 黄高扬, 冀飞, 等. 计算机辅助的中文言语测听平台的建立[J]. 中国听力语言康复科学杂志, 2010(1): 31.
- 7 张华, 王硕, 王靓, 等. 普通话言语测听材料的数字化录制与等价性分析[J]. 临床耳鼻咽喉科杂志, 2006, 20: 1 011.
- 8 张宇晶, 郝昕. 成人人工耳蜗植入相关的中文言语识别评价体系的建立[J]. 听力学及言语疾病杂志, 2012, 20: 387.
- 9 郝昕. 言语测听工具的效度、信度与敏感度[J]. 中华耳科学杂志, 2008, 6: 1.
- 10 颜艳, 徐勇勇. 数据处理的其它统计方法[M]. 见: 孙振球, 徐勇勇, 主编. 医学统计学. 北京: 人民卫生出版社, 2002. 485~509.
- 11 冀飞, 郝昕. 言语测听材料的复测信度评估[J]. 中华耳科学杂志, 2008, 6: 50.
- 12 Killion MC, Niquatte PA, Gudmundsen GI. Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners[J]. J Acoust Soc Am, 2004, 116: 2 395.
- 13 卜行宽, 倪道凤. 推进中文言语测听材料的标准化和临床应用[J]. 中华耳科学杂志, 2008, 6: 9.
- 14 郝昕, 顾瑞, 冀飞. 发展言语识别率测试材料的理论框架[J]. 听力学及言语疾病杂志, 2006, 14: 401.
- 15 冀飞, 郝昕. 影响言语测听的若干因素[J]. 听力学及言语疾病杂志, 2009, 17: 209.

(2012-10-08 收稿)

(本文编辑 周涛)