

# WDRC 释放时间对普通话言语理解度的影响研究

赵建平<sup>1,2,3</sup>, 原 猛<sup>1,3</sup>, 冯海泓<sup>1,3</sup>

(1. 中国科学院声学研究所东海研究站, 上海 200032; 2. 中国科学院大学, 北京 100190;  
3. 中国科学院声学研究所东海研究站-上海市徐汇区中心医院听力与言语医学中心, 上海 200033)

**摘要:** 宽动态范围压缩算法作为助听器非线性听力补偿的核心算法, 其释放时间常数的设定可影响言语理解度。根据普通话的语音特点, 将宽动态范围压缩算法按频率范围划分为低频区间(<2000Hz)和高频区间(>=2000Hz), 并采用不同的释放时间常数。低频区间和高频区间释放时间常数均采用 5ms、100ms 及 2000ms, 但各自交叉配对为实验条件, 共形成 9 组实验条件。以正常听力者为实验对象, 通过仿真平台输出普通话言语仿真声, 进行噪声下普通话言语理解度测试。实验结果显示, 低频区间时间常数过短或高频区间时间常数过长都将影响普通话言语理解度。实验结论对我国助听器 WDRC 参数的设定有一定参考价值。

**关键词:** 宽动态范围压缩算法; 释放时间; 普通话; 言语理解度

中图分类号: R318.04

文献标识码: A

文章编号: 1000-3630(2013)-03-0217-05

DOI 编码: 10.3969/j.issn1000-3630.2013.03.009

## Effect of release time in WDRC on mandarin speech intelligibility

ZHAO Jian-ping<sup>1,2,3</sup>, YUAN Meng<sup>1,3</sup>, FENG Hai-hong<sup>1,3</sup>

(1. Shanghai Acoustics Laboratory, Chinese Academy of Sciences, Shanghai 200032, China;  
2. University of Chinese Academy of Sciences, Beijing 100190, China; 3. Medical Center for Hearing and Speech, Shanghai Acoustics Laboratory & Shanghai Xuhui Central Hospital, Chinese Academy of Sciences, Shanghai 200033, China)

**Abstract:** In digital hearing aids, wide dynamic range compression algorithm (WDRC) is commonly treated as the core algorithm to compensate loudness. Differences between release time constants in WDRC can affect speech intelligibility (SI). Based on unique acoustic characteristics of Mandarin, WDRC algorithm are divided into two parts: low frequency (LF) channel and high frequency (HF) channel. Release time constants are taken as 5 ms, 100ms, and 2000ms in either LF channel or HF channel. Combining a certain LF release time constant with a certain HF one forms totally nine different release time combinations for hearing tests. Normal hearing objects go through the Mandarin SI tests with speech shaped noise. The testing material is Mandarin sentences which are the outputs of WDRC simulation in the 9 different release time combinations. Test results show that too short release time in LF channel or too long release time in HF channel trends to decrease Mandarin SI. The results can be a reference for WDRC parameter setting of Mandarin speaker.

**Key words:** wide dynamic range compression; release time; mandarin; speech intelligibility

## 0 引言

目前数字助听器进行听力补偿的最常用验配处方公式是宽动态范围压缩算法(Wide Dynamic range Compression, WDRC)。该算法对声音进行非线性放大压缩, 将正常的听觉动态范围声音压缩到听力障碍者的动态范围。

WDRC 时间常数的长短反映了算法对外界声音

放大压缩处理速度的快慢。国外不同的助听器厂家所使用的 WDRC 的启动时间(Attack Time, AT)设置从几毫秒到数百毫秒, 释放时间(Release Time, RT)从几毫秒至数秒, 但并未达成统一的设置标准<sup>[1]</sup>。一般认为, 长/短启动时间常以 10ms 为界, 长/短释放时间以 100 ms 为界。对于启动时间的长短基本达成一致, 即选用短的启动时间。短的释放时间可以维持紧邻能量瞬大瞬小的声音可听度, 对于两个能量不同的声音相互交替时有好的效果。但短的释放时间会破坏语音时域包络, 降低语音的调制深度。长释放时间有利于维持语音时域包络, 可以保持短时的能量差变化, 使得双耳能量差的声音定位不被破坏。但长的释放时间会导致紧邻能量瞬大声的瞬时小声得不到及时和足够的增益<sup>[2]</sup>。对于释放时间长短的选择, 国外许多文献探讨了不同的时间

收稿日期: 2012-11-06; 修回日期: 2013-02-07

基金项目: 国家自然科学基金(11104316)、上海市自然科学基金(11ZR1446000)、中国科学院声学研究所所长择优创新基金(Y154221701)、美国斯达克科学研究基金资助项目。

作者简介: 赵建平(1987-), 男, 山东青岛人, 硕士研究生, 研究方向为语音信号处理。

通讯作者: 原猛, E-mail: ym@mail.ioa.ac.cn

常数对于言语易懂度的影响<sup>[1-9]</sup>，但结论也不一致。Jenstad 和 Souza 发现短的释放时间会导致更高的辅元比，对于无语义字节中的 5 个摩擦音和塞音的清辅音有更好的识别率，但是对于另两个清辅音的识别率降低<sup>[3]</sup>。Bentler 和 Nelson 结合不同的释放时间组合，发现不同释放时间对于噪声下的音节识别无影响<sup>[4]</sup>。Hansen 对于正常人和听障者的研究发现，长释放时间下对于连续语句有更好的言语易懂度和言语质量<sup>[5]</sup>。Gatehouse 综述其他的研究者的文献，将时间常数对于言语易懂度和言语质量的影响研究结果归为四类：(1) 时间常数不产生影响；(2) 短时间常数优于长时间常数；(3) 长时间常数优于短时间常数；(4) 不同类别实验对象得出不同的最优时间常数<sup>[6]</sup>。对于助听器厂家来说，虽然各家所采用的时间常数有所差异，但对于同一个品牌(或型号)而言，在通道上所设置的时间常数均相同。

目前尚未有文献研究释放时间对于普通话言语感知效果的影响。普通话相对于英语等印欧语系来说，有音调、清辅音多、音节结构简单且相对固定等声学特点。从吴宗济对普通话元辅音时域频域统计数据<sup>[10]</sup>可以看出，普通话元音和浊辅音的前两个共振峰大部分在 2 kHz 以下，清辅音能量主要集中在的高频区域，频率在 2 kHz 以上。普通话是声调语言，保持元音共振峰轮廓有利于普通话的音调感知<sup>[11]</sup>。同时，普通话为辅-元-辅(CV(C))的音节结构，由于双字词和连续语音中能量大的元音后面紧跟能量小的辅音，过长的时间常数会导致辅音增益不足，因此在 WDRC 算法中要注意选择合理的时间常数，避免清辅音增益降低，使感知度降低。本文提出如下假设：在低频部分，宜采用较长的时间常数以维持元音共振峰特征；在高频部分，宜采用较短的时间常数以使清辅音得到足够的增益。通过将 WDRC 算法的通道划分为低频区间和高频区间，分别选取不同的释放时间，期望能探究不同释放时间对于普通话言语理解度的影响。

### 1 宽动态范围压缩算法

WDRC 算法常用图 1 所示的输入输出曲线描述。算法参数包括压缩拐点(Compression Threshold, CT)、压缩比(Compression Ratio, CR)、增益(Gain)、时间常数(Time Constant, TC)。其中，压缩拐点、压缩比和增益根据听力障碍者的听力损失情况，在助听器验配过程中由处方公式确定。

时间常数(即启动时间和释放时间)是影响 WDRC 算法性能的重要参数之一。ANSI 中对启动

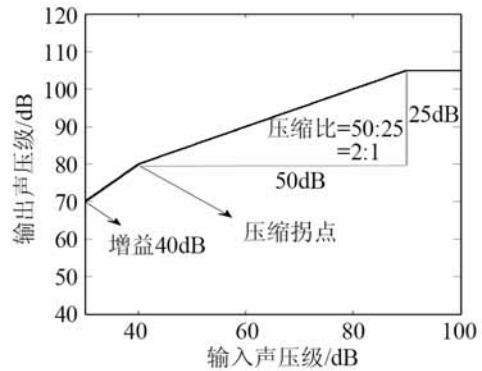


图 1 WDRC 算法输入输出曲线  
Fig.1 Input-output curve of WDRC

时间作了如下定义：当输入信号声压级突然增加某一指定的分贝的瞬间，且助听器输出声压信号稳定在已提高后的稳定声压级，其偏差在±3dB 内的时间间隔，反映了压缩算法对信号强度增加的反应速度。对于释放时间的定义如下：当输入信号从某一指定的声压级突然降到较低声压级的瞬间到助听器输出声压再次稳定到较低的稳态声压级，且其偏差在±4 dB 内的时间间隔<sup>[12]</sup>。例如，图 2 所示为输入信号从 50 dB 声压级突变为 90 dB 声压级，然后再回到 50 dB 声压级时输出信号声压级的变化。图 2(a)为输入信号的声压级变化，图 2(b)为经压缩算法处理后的输出信号声压级变化。对于语音这类能量随时间动态变化的信号来说，其时间常数的长短所引起的语音畸变对于言语易懂度有重要影响。本研究将通过下述实验的设计试图探究时间常数中的释放时间对于普通话言语易懂度的影响。

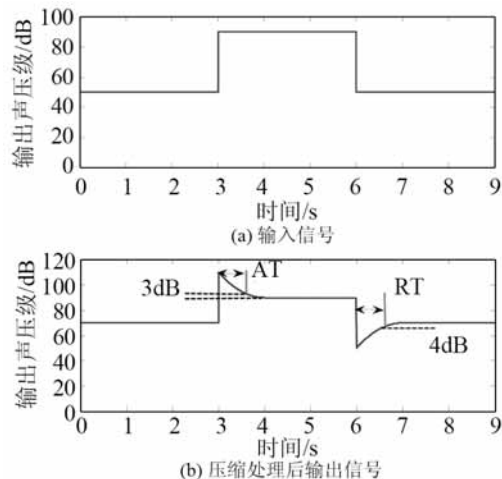


图 2 WDRC 启动/释放时间示意图  
Fig.2 Definition of attack/release time for WDRC

### 2 实验方法

#### 2.1 实验对象

10 名年龄在 20~40 岁、普通话为母语的正常听

力者(4 女, 6 男)参加了本研究。受试者的纯音测听阈值(125 Hz~8 kHz 倍频程)均小于 20 dB HL。

### 2.2 实验设备

测试在本底噪声小于 30 dBA 的隔声室进行。使用计算机连接外置声卡, 并选用插入式耳机(E.A.R. Tone 3A 50Ω)播放刺激声。刺激声经 2CC 耦合腔使用声级计进行校准。采用美国斯达克公司提供的用于研究的助听器压缩算法仿真平台对实验材料进行预处理。测试平台为基于 MATLAB 自主开发的普通话句子测试软件。

### 2.3 实验材料

实验材料为普通话句子<sup>[13]</sup>, 该语音库包含 32 个句表, 每个句表包含 10 个普通话句子, 平均每句话含 7 个汉字。材料录制自一名以普通话为母语的男性播音员。实验中采用掩蔽噪声, 噪声类型为语音型(Speech-Shaped)噪声。

### 2.4 预处理

实验中采用常见典型的缓降型听力损失模型(如图 3 所示)。在 250 Hz~8 kHz 倍频程上的听力损失依次为 30、45、50、60、60、65 dBHL。首先, 将该听力损失图输入 NAL-NL2 处方公式软件<sup>[14]</sup>进行验配, 其输出为 16 个通道不同输入声压级下的输出增益、压缩拐点、及压缩比(如表 1 和表 2 所示)。WDRC 时间常数设定: 启动时间采用普遍公认的短启动时间(10ms)。对于释放时间, 低频区间(即通道 1 至通道 4)选取 5 ms、100 ms 或 2 s; 高频区间(即通道 5 至通道 16)选取 5 ms、100 ms 或 2 s。形成 9 个释放时间组合作为实验条件, 依次为(5ms, 5ms)、(5ms, 100ms)、(5ms, 2s)、(100ms, 5ms)、(100ms, 100ms)、(100ms, 2s)、(2s, 5ms)、(2s, 10ms)和(2s, 2s)。其中, (5 ms, 100 ms)表示低频区间采用 5 ms 释放时间, 高频区间采用 100ms 释放时间, 以此类推。将普通话句子测试材料作为输入, 经仿真平台输出 9 组不同释放时间条件下处理产生的语音音频文件。在所有实验条件下仅释放时间不同, 其他参数由 NAL-NL2 处方公式软件产生, 并在测试过程中保持不变。

### 2.4 实验过程

实验对象佩戴插入式耳机, 在隔声室内进行普通话言语理解度测试。实验开始前, 使用声级计对播放声音进行校准, 使耳机输出声压级保持在 70 dBA。左耳播放目标言语声与噪声的混合声, 信噪比为-15 dB; 右耳使用海绵塞堵塞以达到隔声效果。正式实验前, 由实验操作员引导实验对象进行预实

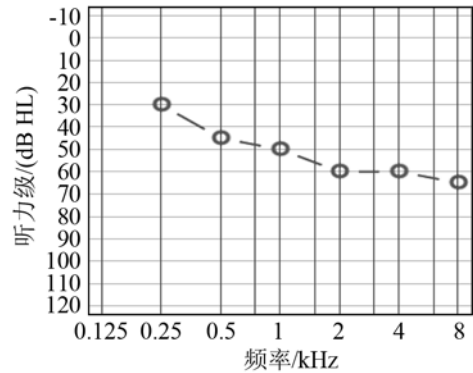


图 3 仿真输入听力损失图

Fig.3 Typical hearing loss for simulation input

表 1 NAL-NL2 处方公式验配输出(通道 1-通道 8)

Table 1 Output of NAL-NL2 description (Band 1- 8)

通道	测试频率/Hz	压缩拐点	压缩比	增益/(50dB)	增益/(65dB)	增益/(80dB)
1	160	46.8	2.3	11.3	2.3	0
2	200	46.7	2.3	11.3	2.3	0
3	250	46.4	2.2	11.3	2.3	0
4	315	47.1	2.2	13.4	4.7	0.3
5	500	48.6	2.2	18.1	10.2	2.7
6	630	47.6	2.1	19.6	11.7	3.8
7	800	44.5	2.2	20.8	12.5	4.4
8	1k	40.7	2.3	21.8	13	4.5

表 2 NAL-NL2 处方公式验配输出(通道 9-通道 16)

Table 2 Output of NAL-NL2 description (Band 9- 16)

通道	测试频率/Hz	压缩拐点	压缩比	增益/(50dB)	增益/(65dB)	增益/(80dB)
9	1250	39.4	2.4	22.6	13.4	4.7
10	1600	38.0	2.5	23.3	13.6	4.3
11	2000	35.0	2.8	23.1	12.8	0
12	2500	34.1	2.8	22.2	9.6	0
13	4000	32.7	2.7	22.4	10.2	0
14	5000	31.5	2.5	22.2	12.6	0
15	6300	30.9	2.0	20.2	11.7	3.7
16	8000	30.4	1.4	9.8	5.6	0.4

验, 熟悉实验流程和实验材料并消除学习效应。正式实验时, 实验对象复述所听到的普通话语句, 由实验操作员统计并记录复述结果。对于每个实验对象, 9 组实验参数条件顺序随机产生, 每组实验参数下测试经预处理的 1 个句表, 句表顺序随机产生。实验对象每隔 20 min 休息 5 min。整个实验持续时间约 2h。

## 3 结果和讨论

10 名听力正常者普通话言语理解度测试实验结果如图 4 所示。从总体趋势可发现: (1) 低频区间(LF)释放时间固定时, 言语正确率随着高频区间(HF)释放时间的增加而减小。(2) 高频释放时间固

定时, 言语正确率随着低频释放时间的增加而增加。

对实验结果进一步采用多因素重复测量方差分析法(ANOVA)分析释放时间对言语理解度的影响。以 LF 释放时间和 HF 释放时间作为影响因子: LF (5 ms, 100ms, 2s), HF(5ms, 100ms, 2s)。ANOVA 分析结果显示, LF 释放时间对普通话言语理解度有显著影响([ $F(2, 18)=15.4, p<0.0005$ ]); HF 释放时间对普通话言语理解度也有显著影响[ $F(2, 18)=13.7, p<0.0005$ ]; 在 LF 释放时间和 HF 释放时间交互影响时并未对实验结果有显著影响( $p>0.05$ )。

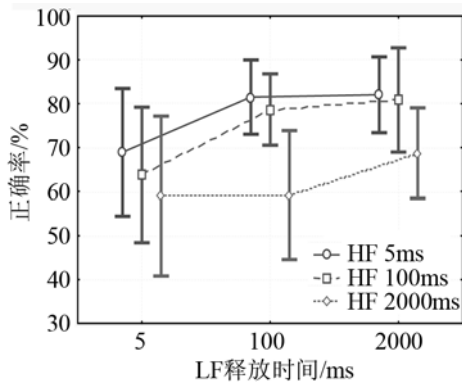


图4 普通话言语理解度测试总体结果图  
Fig.4 Overall results of mandarin speech intelligibility

采用 Post-hoc LSD 测试 9 组条件下 LF 通道和 HF 通道释放时间的个体间实验结果差异, 结果见表 3。对于 LF 区间释放时间为 5 ms 的三种情形, HF 区间释放时间为 5ms 的识别率显著比释放时间为 2s 时高( $p<0.05$ )。平均识别率最高的四组条件分别为: (2s, 5ms)、(100ms, 5ms)、(2s, 100ms)和(100ms, 100ms), 识别率分别为 82.0%、81.5%、80.6%和 78.6%。上述四组识别率最高的实验条件恰是 LF 区间释放时间选取 100ms、2 s 和 HF 区间释放时间选取 5 ms、100ms 的两两组合, 且几乎与其他组有明显统计意义( $p<0.05$ ), (100 ms, 2 s)和(5 ms, 5 ms)之间( $p=0.052$ )除外。平均识别率最低的两组(5ms, 2 s)、(100 ms, 2 s), 识别率分别为 59.0%和 59.2%,

表 3 普通话言语理解度测试 Post-hoc 检测结果  
Table 3 Post-hoc LSD results of mandarin SI

实验编号	LF 释放时间/ms	HF 释放时间/ms	平均识别率/%	显著性差异 $p$ 值								
				1	2	3	4	5	6	7	8	9
1	5	5	68.9		0.265	0.034	0.008	0.037	0.037	0.006	0.012	0.972
2	5	100	63.8	0.265		0.294	0.000	0.002	0.313	0.000	0.000	0.280
3	5	2000	59.0	0.034	0.294		0.000	0.000	0.966	0.000	0.000	0.037
4	100	5	81.5	0.008	0.000	0.000		0.532	0.000	0.908	0.885	0.007
5	100	100	78.6	0.037	0.002	0.000	0.532		0.000	0.460	0.631	0.034
6	100	2000	59.2	0.037	0.313	0.966	0.000	0.000		0.000	0.000	0.040
7	2000	5	82.0	0.006	0.000	0.000	0.908	0.460	0.000		0.795	0.005
8	2000	100	80.6	0.012	0.000	0.000	0.885	0.631	0.000	0.795		0.011
9	2000	2000	68.8	0.972	0.280	0.037	0.007	0.034	0.040	0.005	0.011	

且与其他大部分参数有着显著差异( $p<0.05$ )。上述实验结果表明, 在低频采用释放时间过短(5 ms)或在高频采用释放时间过长(2 s)均会使普通话言语理解度显著降低, 本实验结果验证了本文提出的假设。

以普通话句子“燕子吃不到小虫子”为例, 对不同释放时间处理的效果进行声学分析。如图 5 所示, 图 5(a)为原始语句时域波形图, 图 5(b)为原始语句、释放时间(2s, 5ms)和释放时间(5ms, 2s)处理的时域包络对比图。其中, 为方便对比, 将原始语句进行加权放大 20 dB。释放时间(2 s, 5 ms)和释放时间(5 ms, 2 s)为普通话言语识别率最高组和最低组。实线为原始语音包络, 短虚线为释放时间(5 ms, 2s), 长虚线为(2s, 5ms)。虚线框区域为语句中清辅音段。从图中可见, 清辅音在(2s, 5ms)条件下能量大于(5ms, 2s)。同时, 元音和浊辅音的包络在(5ms, 2s)条件下的变化幅度大于(2s, 5ms), 与原始信号的包络相差明显, 说明此条件下的包络失真更大。

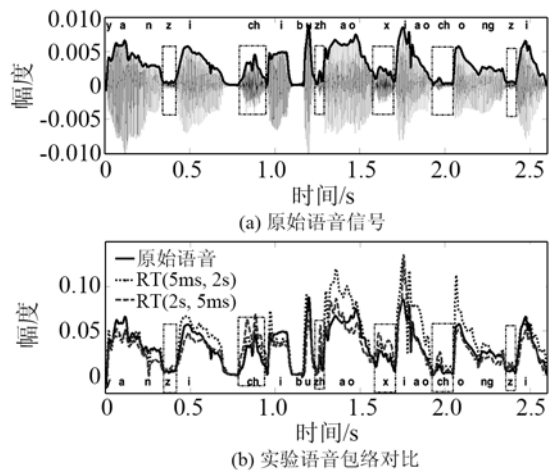


图5 实验语音材料对比  
Fig.5 Comparison between testing materials of different RT combinations

实验结果验证了本文根据普通话声学特征分析提出的假设。普通话元音共振峰携带音调信息, 辅音能量小但携带大量言语信息。划分低频、高频

双区间处理, 为不使元音共振峰结构遭到破坏, 低频区间不应采用短的释放时间, 维持原始共振峰包络结构; 普通话的 CVC 结构, 高频区间不应采用较长的释放时间, 以防止清辅音由于紧邻的元音处于 WDRC 算法压缩状态, 过长的释放时间致使清辅音部分或者全部处于压缩状态, 获得增益减少, 导致感知下降。

## 4 结论

本文通过对于听力正常者不同释放时间条件下普通话句子测试表明, 低频区间释放时间不宜设置过短, 对应高频区间释放时间不宜设置过长。本文通过对听力正常人普通话言语理解度测试, 所得结果可为我国助听器 WDRC 参数优化提供一定借鉴意义。

致谢: 在此, 感谢参与本次实验的志愿者们。同时, 感谢美国斯达克公司提供的研究用助听器压缩算法仿真平台, 并特别感谢 Sridhar Kalluri 博士对本研究提供的建议和帮助。

### 参 考 文 献

- [1] Gilbert G, Akeroyd M A. Discrimination of release time constants in hearing-aid compressors[J]. *International Journal of Audiology*, 2008, **47**(4): 189-198.
- [2] Moore B C. The choice of compression speed in hearing aids: Theoretical and practical considerations and the role of individual differences[J]. *Trends in Amplification*, 2008, **12**(2): 103-112.
- [3] Jenstad L M, Souza P E. Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility[J]. *Journal of Speech, Language, and Hearing Research*, 2005, **48**(1): 651-667.
- [4] Bentler R A, Nelson J A. Assessing release-time options in a two-channel AGC hearing aid[J]. *American Journal of Audiology*, 1997, **6**(1): 43-51.
- [5] Hansen M. Effects of multi-channel compression time constants on subjectively perceived sound quality and speech intelligibility[J]. *Ear and Hearing*, 2002, **23**(4): 369-380.
- [6] Gatehouse S, Naylor G. Linear and nonlinear hearing aid fittings-1. Patterns of benefit: Adaptación de auxiliares auditivos lineales y no lineales-1. Patrones de beneficio[J]. *International Journal of Audiology*, 2006, **45**(3): 130-152.
- [7] Moore B C, Glasberg B R, Stone M A. Optimization of a slow-acting automatic gain control system for use in hearing aids[J]. *Brit J Audiol* 1991, **25**(3): 171-182.
- [8] Henning R L W, Bentler R A. The effects of hearing aid compression parameters on the short-term dynamic range of continuous speech[J]. *Journal of Speech, Language and Hearing Research*, 2008, **51**(2): 471-484.
- [9] Davies-Venn E, Souza P. Effects of audibility and multichannel wide dynamic range compression on consonant recognition for listeners with severe hearing loss[J]. *Ear and hearing*, 2009, **30**(5): 494-504.
- [10] 吴宗济. 普通话元音和辅音的频谱分析及共振峰的测算[J]. *声学学报(中文版)*, 1964, **1**(1): 34-40.  
WU Zongjing. Spectrum analysis and formant measurement of Mandarin vowel and consonant[J]. *Acta Acustica*, 1964, **1**(1): 34-40.
- [11] LUO X, Fu Q J. Enhancing Chinese tone recognition by manipulating amplitude envelope: implications for cochlear implants[J]. *J. Acoust. Soc. Am*, 2004, **116**(6): 3659-3667.
- [12] ANS/S3.22-2003, Specification of hearing aid characteristics[S].
- [13] 郝昕, 赵阳. 嘈杂语噪声下汉语整句识别的同质性研究[J]. *中华耳科学杂志*, 2008, **6**(1): 35-39.  
XI Xin, ZHAO Yang. Homogeneity of mandarin sentence intelligibility in babble noise[J]. *Chinese Journal of Otology*, 2008, **6**(1): 35-39.
- [14] Keidser G, Dillon H. The NAL-NL2 prescription procedure[J]. *Audiology Research*, 2011, **1**(1): 88-90.