

汉语降频助听算法的实验分析与改进*

梁瑞宇^{1)2)†} 奚吉¹⁾²⁾ 赵力²⁾ 邹采荣²⁾ 黄程韦²⁾

1) (河海大学计算机与信息学院, 南京 210098)

2) (东南大学信息科学与工程学院, 南京 210096)

(2011 年 10 月 17 日收到; 2011 年 11 月 21 日收到修改稿)

降频助听算法是改善听障患者声音辨识能力的最安全有效的方法. 本文以主观测试实验为手段, 通过分析当前算法的声音识别能力的不足, 提出一种自适应慢放降频算法. 算法结合慢放算法和频移算法的优点, 并能根据信号的频谱结构, 自适应调整慢放因子, 降低时域不同步性. 并且, 通过分析含噪信号和噪声信号的频谱关系, 提出一种噪声下的慢放因子评估方法. 实验结果显示, 同其他降频算法相比, 该算法可以提高 15% 到 20% 的识别率. 在对听障患者的测试中, 同传统的助听设备相比, 平均识别率也获得显著改善.

关键词: 降频, 汉语助听器, 高频听力损失, 自适应慢放

PACS: 43.71, 43.72

1 引言

中国有一亿三千万的听障患者, 其中 90% 的人存在中高频听力损失. 此外, 耳机的过度使用和急剧恶化的老龄化问题, 使得青少年和老人中的高频听损患者比例更高. 高频听力损失意味着对于低频都能听的很好, 而高频听的很少甚至根本听不到^[1]. 成人无法感知这些高频信息, 就没法辨识语言, 进行交流. 而对于幼儿来说, 则无法学习发音和许多语法规则^[2], 变得既聋又哑. 从技术上来说, 高频处的听力损失越严重, 就需要越多的增益才能获得可听性. 但是, 这些增益要受到回声、低频掩蔽特性或者放大系统的输出性能的限制^[3], 几乎不可能获得足够的增益. 此外, 研究证明, 高频听力损失超过 70 dB 时, 内毛细胞的功能已损失殆尽^[4]. 此时, 传统的振幅放大技术不但对语言理解没有作用, 反而会产生失真, 不适及反馈啸叫^[5,6].

目前, 改善听障患者的高频损失的技术主要有两种: 降频助听和电子耳蜗. 这两种技术都要对语言重新编码, 并且经过语言培训后, 患者才能达到

最佳效果. 移植电子耳蜗后, 大部分的患者在语音感知识别方面有了一定改善. 但是, 电子耳蜗存在手术危险, 并且费用昂贵, 一般是降频助听器的十到二十倍. 其最大的潜在危害是, 电子耳蜗的破坏性的手术植入, 改变了患者的中耳结构, 使患者失去了对新技术的选择机会.

降频助听技术不需手术, 技术升级容易, 成为高频患者的首选技术. 目前, 降频助听技术主要包括多通道声码器、慢速回放、频率转移和频率压缩四类^[7]. 多通道声码器是最早的降频技术, 利用信号的高频包络调制噪声或纯音信号进行降频. 但是, 由于调制后的信号屏蔽了一些重要的低频信息^[8], 该技术没有显示出显著的高频性能改善. 最早的商业助听技术是慢速回放, 其原理是以比采集速度慢的速度来播放音频信号. 虽然可以保留频率成分的比例关系, 但是信号在时间轴上的拉伸, 造成输入输出间的不同步, 从而引起失真. 实验表明, 慢速回放的改善效果不太明显^[5]. 丹麦唯听公司的 Inteo 助听器集成了频率转移技术, 将高频信息转移到低频部分, 并与原来低频信号相加^[9]. 频移技术有更自然的声音质量, 长期的实验证实表明,

* 国家自然科学基金 (批准号: 60872073, 60975017, 51075068)、中央高校基本科研业务费 (批准号: 2009B32614) 和广东省自然科学基金 (批准号: 10252800001000001) 资助的课题.

† E-mail: lly1711@163.com

对改善患者的高频感知能力有显著的作用. 虽然频率成分间存在交叠, 但是对低频信息 (如元音) 的影响并不明显^[10]. 相比于上述三种技术, 目前研究最多的是非线性频率压缩技术, 其原理是保持低频信号不变, 高频信号以较大比例压缩. 不同于频移技术, 非线性频率压缩技术的频率信息间没有交叠, 尤其保留了元音信号的理解度. 2005 年, Simpson 等^[6] 设计了非线性频率压缩技术的截止频率和压缩因子的选择方法, 在随后的测试中, 大约 50% 的被试表现出明显改善. 2009 年, Scollie 等^[11] 改善了听力适配方法, 在 13 位儿童和 11 位小孩的测试中, 87% 的患者显示出明显改善. 2010 年, Bohnert 等^[3] 评估了噪声下非线性频率压缩算法的性能, 结果是 63% 的患者改善明显, 而其余患者并没有任何改善, 甚至比传统助听器差. 另一种频率压缩方法是线性压缩, 即按照固定比例压缩所有频率信息. 但是研究显示该方法并没有明显改善患者的高频听力^[12]. 此外, Arioiz 等还提出一种分频段处理的方法: 幅度放大低频信号, 移动中频信号, 非线性压缩高频信号^[13]. 在患者的语音、环境和音乐的测试, 患者的高频可听性获得了较大提高. 但是, 非常遗憾, 该方法没有同传统的助听技术做比较.

上述研究虽然获得了一些成果, 但只是针对母语为英语的患者进行测试, 针对汉语的降频技术的报道多为综述或介绍性的论文. 目前的研究结果表明, 汉语和英语在听觉感知过程中有很大差别. 首先, 英语是语调语言, 而汉语是声调语言. 声调信息是汉语的重要组成部分, 对提高汉语语言的可懂度有重要贡献^[14]. 其次, 汉语中包含较多的清辅音^[15], 而清辅音虽然能量较少, 但频率较高 (一般都不低于 2 kHz)^[16], 无法被高频听损患者感知. 而且, 像汉语中的 /j/, /q/, /x/, 在英语中并无相同发音. 辅音对语言清晰度的贡献达到 95%, 听不清这些辅音, 就会使听辨语言的能力下降甚至丧失^[17]. 另外, 在对听障患者非语音声的识别上, 很多学者只是做了比较, 没有具体研究影响降频声音识别的客观因素, 更没有提出相应的改善方法.

本研究以主观测听为手段, 首先比较了不同的降频算法下, 母语为汉语的被试对英语和汉语的语言要素的识别能力, 分析了现有算法在汉语识别中的不足, 并结合生活实际, 提出一种自适应慢放频移算法. 算法能根据信号频谱分布决定慢放因子, 然后将慢放后的信号进行频移. 针对汉语特点, 该

算法只对小部分语音做慢放处理, 提高识别率, 对大部分语音保持不变, 保持时域同步性. 而对于高频能量较高的环境声, 也能以合适的慢放因子进行降频, 提高声音的可听性和识别率. 此外, 通过分析噪声和含噪信号间的频谱关系, 本文提出一种噪声情况下的慢放因子的估计方法, 增强算法的实用性. 实验结果显示, 被试的声音识别率提高 15% 到 20%. 在对听障患者的测试中, 该算法也比传统助听设备获得更明显的听力改善.

2 常用降频算法分析

2.1 实验设置

为了保证所有算法处在相同的测试条件下, 本文采用 10 位正常听力者作为测试者, 8 男 2 女. 因为不同高频听损患者的截止频率是不同的, 即使是相同的, 其内部的听力损伤也存在差异. 如理论上, 患者的高频听损达到 70 dB 以上, 其内毛细胞会严重缺失. 但是, 其缺失的程度是不同的, 残留了一些内毛细胞的患者的测试结果要明显好于无残留的患者. 在这种情况下, 所得到的算法评估会出现较大偏差, 从而无法客观的评价算法质量. 为此, 在提供给被试测试声音前, 加一级全相位低通滤波器. 低通滤波器的截止频率就是设定的截止频率, 从而保证受试信号中不存在高于截止频率的成分. 但是, 如果是听障患者的测试则不需要低通滤波器. 测试流程如图 1 所示.

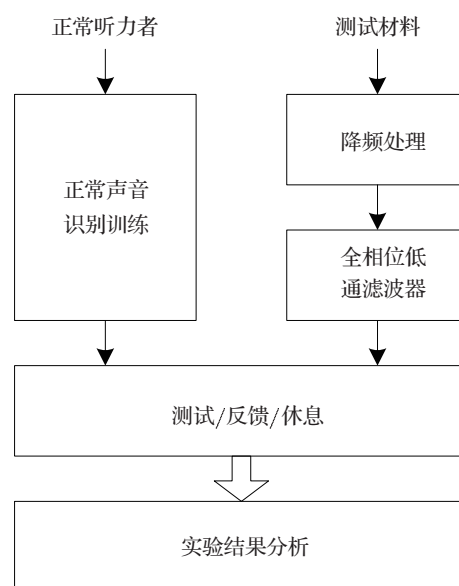


图 1 正常听力者测试流程

首先,所有正常的听力测试者必须经过正常声音识别训练,减少因为缺乏专业知识(如乐器)或没有概念(如鲸鱼叫声)而导致的识别偏差.训练采用正常声音进行,这和听障患者的训练是不同的.训练声音播放3遍后,对被试进行简单测试,如果识别率达到85%以上,则认为通过训练,否则继续训练.

表1 测试反馈表及评分标准

声音编号	1	...	评分标准
①首选的内容	...		正确: 1分; 错误: 0.25分
②可能的内容	...		正确: 0.75分; 错误: 0.25分
③听到但不知内容	<input type="checkbox"/>	...	选择: 0.5分
④几乎听不到	<input type="checkbox"/>	...	选择: 0分

测试测试环境为中等大小的消声室.测试中,每个被试坐在位于扬声器1 m处的位置.每个声音样本的长度为5 s.发音平均强度维持在55—60 dB,这与日常交谈的强度相似.测试过程分为三个部分:听测试声音、识别结果反馈和休息.识别结果反馈采用三级可选记录方式,表格如表1所示.每组语音或环境声间隔为5—10 s,以保证被试有足够的时间记录结果.被试可以根据听到的声音,填写①、②、单选③或单选④.实验一般是每测试45 min,休息20 min,防止被试听觉疲劳.但在含噪声音测试时,每测试30 min,休息30 min.测试完成后,根据被试所填的表格,进行识别率评估.为了全面分析算法的性能,本文采用如表1所示评分标准.这里,设置③大于①和②填写错误的分数,是因为选择③有两种常见原因:1)被试没听过(训练不充分);2)因为降频,声音失真,听起来与众不同.这两种情况后期可以通过训练改善.而将声音误听为其他声音是因为降频后的失真造成的,这种错误即使是训练也难以纠正,只能通过改变算法细节实现,因此分数较低.

实验所有声音都是16 kHz采样,选自数据库中的声音如不是16 kHz采样,都通过CoolEdit软件进行转换.语音测试材料都选自《普通话水平测试实施纲要》,由具有标准汉语发音的男声或女声录制,并根据文献[18]进行选择组织并生成音节、单词和句子测试表.实验中选择的噪声都来自于NoiseX-92噪声库,根据语音长度按照不同的信噪比叠加到测试信号中.十五种噪声的类型为:(1)喷气飞机(190海里),(2)喷气飞机(450海里),(3)驱逐舰引擎,(4)驱逐舰操作间,(5)F16座舱,(6)工

厂车间1,(7)工厂车间2,(8)高频信道,(9)机枪,(10)军用车辆,(11)粉红噪声,(12)乱糟糟的语音,(13)坦克,(14)车辆内部,(15)白噪声.环境声有四类:动物、自然气候、日常生活和乐器.其中自然声音5种,生活声音33种,乐器声音19种,动物声音29种.每种声音包含4-5组相似的语音片段,每段语音长度为5 s.

针对正常被试的实验中,所有算法的截止频率均设为1 kHz.比较的慢放算法^[5]和线性压缩算法^[19]的比例因子都设为2.多通道声码器的实现方法是将语音通过0.8 kHz的低通滤波器,然后将四种不同中心频率的带通噪声分别调制转移到低频的输出频带上,其生成的信号作为高频部分,最后与通过的低通滤波器的语音信号混合产生降噪信号^[8].频率转移算法采用的是Widex Inteo公司的AE助听器中使用的线性频移算法,即高于截止频率的一个倍频程被转移到低频区与原先的低频信号叠加^[20].非线性压缩算法的输入频率和输出频率的映射关系按照文献[21]设置.

2.2 测试量表的信效度分析

言语测试的信度和效度对于测试材料的选择和测试结果的解释都至关重要,可解决数据的准确性和可靠性的评价问题.本文采用SPSS 17.0软件对所用量表进行信效度分析.信度指标包括重测信度^[22]、分半信度和Cronbach α 系数.效度包括表面效度和关联效度.

本文所选用的汉语测试量表选自张家驷编写的言语清晰度测试材料^[18].该测试材料不仅声、韵、调是平衡的,而且在单词长度上也是平衡的,即单音词、双音词、三音词与多音词的出现频率相平.其中,清晰度测试词表包含在国家标准GB 15508-1995中.因此,该测试量表表面效度较好.同时,根据被试对自己测试分数的评估,所得分数与计算得分作Pearson相关分析,所得系数为0.82($p < 0.01$),说明量表的关联效度较好.

经测试分析,量表的重测信度的相关系数范围在0.77—0.94之间,说明量表在时间上的稳定性较好;分半信度在0.80—0.90之间,说明量表跨指标的一致性较好;Cronbach α 系数范围在0.72—0.82之间,显示了指标体系问题的内部一致性较好.

本文采用的测试数据的筛选方式为,凡某一被试得分与全体平均得分相差大于标准差三倍时,则删除该记录,并重新计算平均得分.

表 3 英汉基本语言要素识别

语种	类型 (个数)	声码/%	慢放/%	频移/%	线压/%	非线性压/%
英语	元音 (20)	72	80	83	81	89
	辅音 (28)	65	70	77	68	73
	韵母 (39)	69	76	78	74	76
汉语	声母 (21)	56	65	60	56	60
	声调 (4)	85	55	94	40	93

2.3 英汉语言要素识别比较

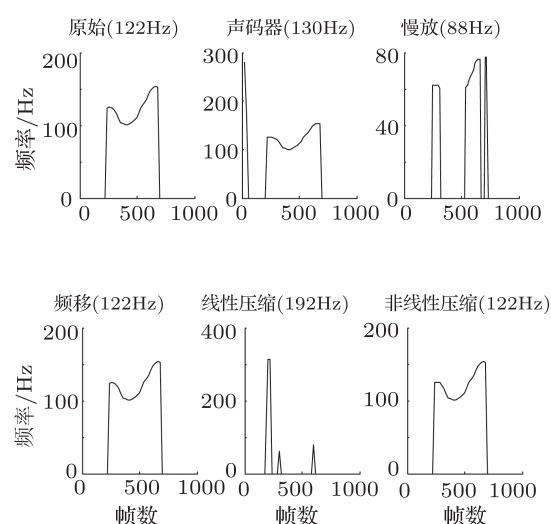
音标是构成语言的最基本要素,通过对音标的识别分析,有助于改善对语音的识别.更重要的是可以学习发音,了解语法规则.此外,对于汉语来说,声调是很重要的要素.本文比较了五种降频算法的英汉语言要素的识别情况,设定的截止频率为 1 kHz,识别结果如表 3 所示.此处,慢放算法和线性压缩方法的比例因子都设为 2. 如果将 8 kHz 的信号都压缩到 1 kHz 内,需要的因子为 8,此时,时域上的过度拉伸导致无法辨认出任何声音.

如表所示,各种算法对英语的识别率要好于汉语的识别率,元音比韵母高 6.4%,辅音比声母高 11.2%. 实际上,识别结果可能会更高,因为被试的母语是汉语,可能对一些英语辅音的判断上本身就存在不足,如 */t f/* 和 */f/*, */ts/* 和 */dz/* 等. 对汉语语音标识别率偏低的原因主要有两点: (1) 被试常将 */i/* 误听为 */u/*, 从而影响复韵母和鼻韵母的识别,如 */ia/*, */iao/*, */in/* 和 */ing/* 等; (2) 汉语有 17 个清辅音,而英语只有 10 个,因此其识别结果要低于英语. 而且对于 */j/*, */q/*, */x/* 识别,常被误听为 */z/*, */c/*, */s/*.

在对声调的判断中,频移方法和非线性压缩方法明显好于其他两种方法. 图 2 所示为 */shi/* 的基频变化曲线. 这两种方法由于保留了低频部分不表,所以保持了较好的基频曲线,因此其声调识别率要好于其他 3 种方法. 而线性压缩方法最差,声码器方法较好.

从上面可以看出,对于汉语的识别上,尤其是声调的识别上,频移方法和非线性压缩方法较好. 而声调对于汉语识别非常重要,因此后继实验主要针对这两种方法,进行比较. 声码算法虽然声调识别率也比较高,但是过去的研究指出该算法对于改善高频听力并无多大作用. 从表 3 也可以看出该算法对汉语音标的识别率也不高. 此外,慢放方法作

为一种有成熟产品的方法,在后继的比较中,也考虑了该算法.

图 2 */shi/* 的基频变化曲线

2.4 环境声识别性能比较

本实验比较了 2 倍慢放方法、频移方法和非线性压缩方法对四类非语言声音的识别. 由于语音信号的能量主要集中在低频区,所以这些非语言声音的识别与语音相比是有很大区别的^[23]. 本实验的目的就是验证降频算法对能量集中在高频区的环境声的识别. 识别率如图 3 所示,非线性压缩方法和频移法对环境声的识别相差不大. 慢放方法的起伏较大,而且被试反应所听到的声音自然度较差,拖沓感明显. 通过分析实验数据,得出一些结论:

1) 有两类声音识别效果较差: (1) 自然声音 (风、雨等): 此类声音实验中常作为噪声使用,而降频后的声音较沉闷,两者易混淆; (2) 高频声音,尤其是高低频能量对比明显的声音 (鹰, 海豚等): 从频移或频率压缩的原理看,频率越高的区域压缩的比例越大,因此失真越大. 所以降频后,几乎无法

辨认甚至是听不到声音.

2) 图中显示的慢放速率是两倍, 但是通过实验发现, 适当改变慢放速率可以改善高频声音的识别率. 以动物声音为例, 鹰和鲸鱼的慢放因子为 2, 鸟的慢放因子为 3, 蟋蟀的是 4. 此外, 其他动物声音, 不经过降频, 人耳也可以较好的辨识.

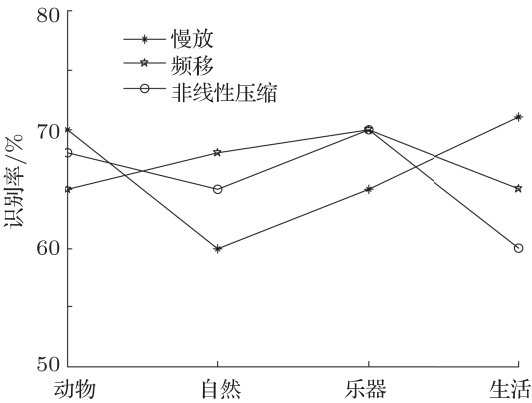


图3 环境声的识别

对韵母的识别率在 60%到 80%之间, 而对声母的识别率普遍低于 70%, 这对于学习语法和发音是不够的. 而且, 对于一些高频能量集中的声音, 目前的算法几乎没有作用. 在日常生活中, 每当别人听不清我们说话时, 我们总是试图降低语音速度, 同时增强音量来使对方听清, 这和慢放的道理大同小异. 从 2.3 节的研究也可知, 适当的慢放有助于提高声音的识别率. 可是, 如果我们每句话都如此说的话, 情况就会变得非常奇怪. 事实上, 对于绝大多数韵母和浊音来说, 频移算法和非线性压缩算法的识别率都比较高. 此时, 采用慢放法, 反而使语音听起来奇怪, 而且过多的慢放, 不光会使延长语音时间, 造成连续语音的同步问题, 也会使语音变得难以识别. 只有不同的慢放速率相结合才可以得到最佳的识别效果. 过大或过小都会影响声音的识别, 甚至无法识别.

表 4 动物的慢放因子范围

动物	可听慢放因子
蟋蟀	3, 4
小鸟, 鹰	2, 3
鲸鱼	1.5, 2, 3
其余	1, 1.5, 2

3 自适应慢放频移算法

3.1 慢放因子的评估

从 2.2 和 2.3 节的分析可知, 现有的降频算法

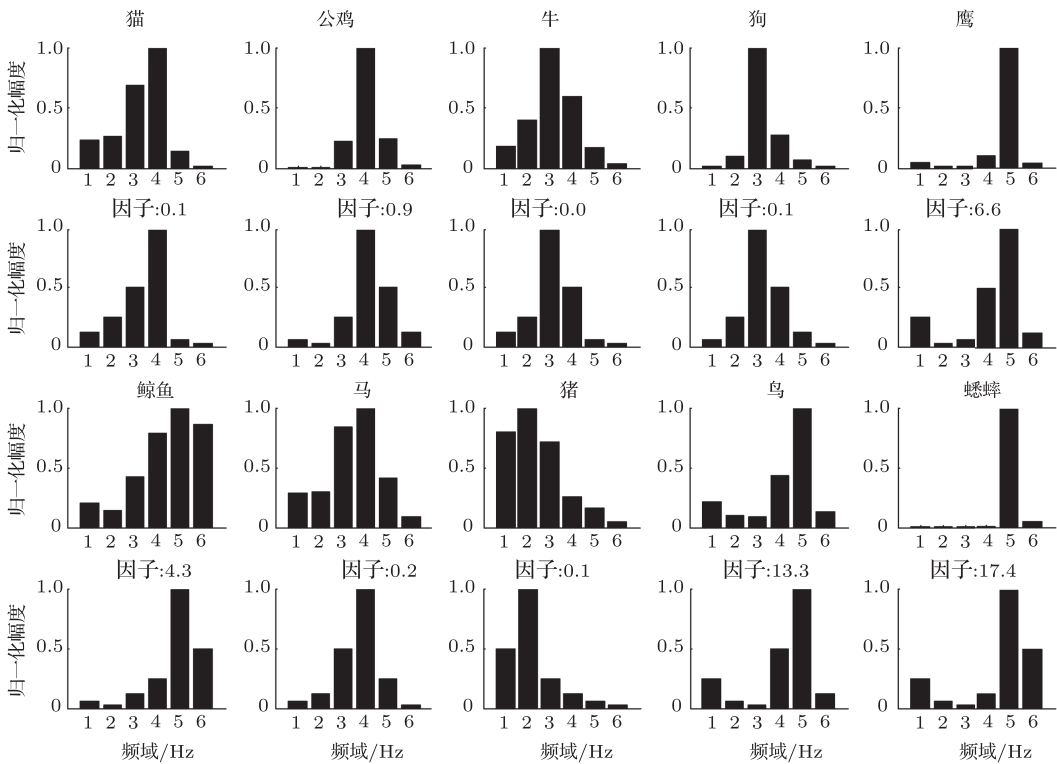


图4 十种动物的频谱直方图

以动物声音为例,图4列出了实验中10种动物声音的频谱直方图和标准化后的频谱直方图,图中的横轴1到6代表6个频域范围:0—250 Hz, 250—500 Hz, 500—1000 Hz, 1000—2000 Hz, 2000—4000 Hz, 4000—8000 Hz. 此处将频谱按照倍频程划分,是为了减少谱失真^[13]. 表4为实验获得的,保证十种动物声音可识别的慢放因子范围. 联合图4和表4分析可知,影响慢放因子的因素主要有两类. 首先,信号中的高频成分. 从降频算法的原理可知,信号中的高频成分越多,需要慢放的因子则越多,如果没有高频成分,则根本不需要慢放. 其次,截止频率所在的频段. 如果截止频率为1 kHz,那么对应图中横轴的3所代表的频域. 从图4可以看出,鲸鱼和蟋蟀的标准化后的2000—4000 Hz区间的高频成分比例相同,但是其实际的慢放比例蟋蟀要大于鲸鱼,而鸟类的标准化后的高频成分比鲸鱼还小,但是其慢放因子也大于鲸鱼. 原因可能在于,3号和4号频域代表的中频范围,对识别度影响较大^[17],如果该频域成分较高的话,会部分削弱高频损失带来的影响. 针对其他数据实验也显示了截止频率所在的频段对降频因子的削弱作用. 此外,研究还显示,高于截止频率的第一个频带对慢放因子的影响不

大,可能是由于频移算法的作用,该频带的信息会搬移到截止频率以下. 同时,结合人耳对不同声音区域的理解能力,对标准化系数进行修正^[17]. 计算公式如式1所示.

$$\text{coef} = \frac{\sum_{i=J+2}^6 c_i x_i}{c_J x_J}, \quad (1)$$

这里, x_i 表示图所示的归一化幅度, J 代表截止频率所在的频域, c_i 表示人耳对不同声音区域的理解系数^[17]. 此处, $J+2$ 代表,忽略高于截止频率的第一个频带对慢放因子的影响. 图中显示的因子为计算得到的因子,其大小关系与表4实验得到的关系是匹配的.

3.2 噪声下的算法改进

现实生活是充满各种噪声和背景,必然影响信号的频谱,从而导致慢放因子计算错误. 图5列举了-15 dB时,五种典型噪声下的鲸鱼声音的频域直方图. 第一行显示的是五种噪声的频域直方图,由图可见,喷气式飞机、乱哄哄语音和车辆内部属于低频噪声,高频信道是高频噪声,而白噪声的频

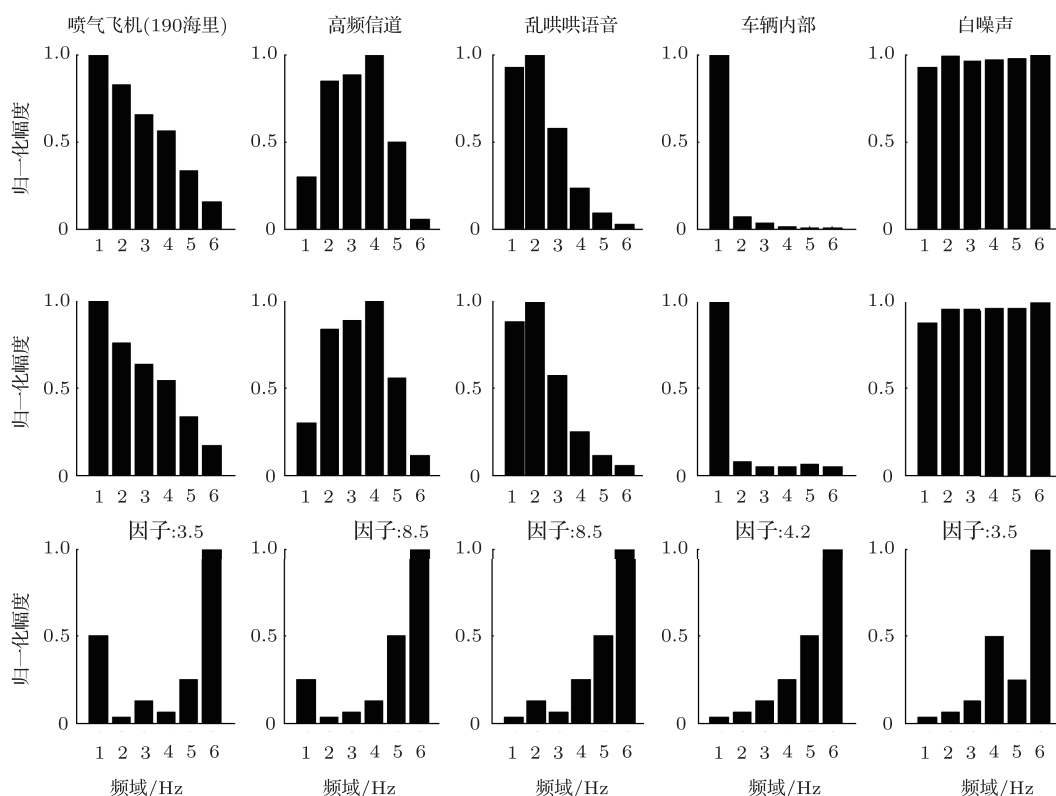


图5 五种噪声下的鲸鱼的频谱图

谱分布均匀. 而且, 车辆内部噪声的频率成分较单一. 第二行是含噪的鲸鱼叫声的频域直方图. 对比图 4 和图 5, 可以看出鲸鱼的频谱被噪声严重污染, 所得的频域直方图几乎和噪声本身相同. 但是, 通过对比两行数据, 我们可以看出, 原先信号的频率成分的比例关系可以从含噪信号的频谱中获得. 因为鲸鱼频谱的高频成分高, 所以获得含噪信号的频谱的高频成分就比原始信号的频谱要高. 将含噪频谱除以噪声频谱, 可以较精确的获得原先信号的频率成分的比例关系. 第三行数据显示了这种比例关系, 虽然与纯净信号的频谱关系存在一定差别. 但是, 从所获得频谱可以看出信号中高比例的高频成分, 并且慢放因子处在一定范围内变化, 有一定的抗干扰性. 更重要的是, 该算法实现方便, 实时性高, 符合助听器应用背景.

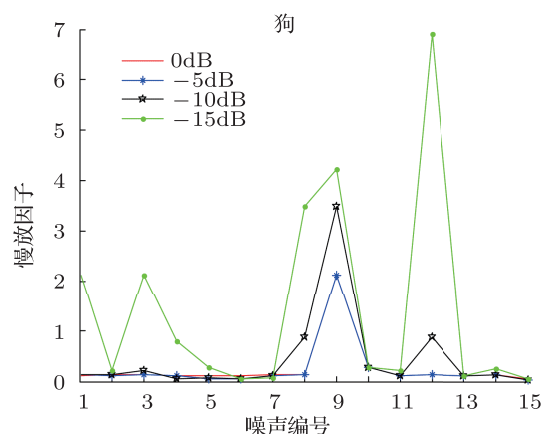


图 6 不同噪声下狗叫声的慢放因子变化规律

表 5 慢放因子对应关系

计算因子	实际因子
< 1	1
1—3.5	1.5
3.5—6.5	2
6.5—14	3
> 14	4

图 6 显示, 低频占主要成分的狗叫声在 15 种噪声下的慢放因子情况. 从图上可以看出, 慢放因子的取值主要与信噪比和噪声类型有关. 从图上可以看出, 在 -15 dB 时, 慢放因子计算的偏差较大. 此外, 在乱哄哄语音和机枪噪声下的, 偏差也较大.

根据原始信号的慢放因子和含噪信号的慢放因子变化范围, 确定实际慢放因子与计算的慢放因

子的对应关系, 如表 5 所示. 从表 4 可知, 每种动物的可听的慢放因子有一定范围. 所以总的说来, 只有在 -15 dB 的乱哄哄语音下, 其慢放因子的计算超出可听范围. 所有的声音统计下, 慢放因子的计算正确率平均为 89%, 主要是一些高频声音的识别率较低. 对于汉语语音来说, 高频能量低, 慢放因子低, 正确识别率可达 95% 以上.

3.3 算法实现原理

综合上述的实验和分析, 本文提出一种自适应慢放频移算法, 其实现流程如图 7 所示. 语音信号首先经过傅里叶变换获得其频谱信息. 自适应慢放频移算法所需的截止频率由患者的听力评估获得或在实验中指定, 而慢放因子则通过分析信号频谱获得. 经过降频后, 通过傅里叶反变换可获得所需的语音信号. 如果是针对正常听力者的实验, 可在降频后先通过全相位滤波器, 滤除高于截止频率频率的信号, 以保证实验的可靠性.

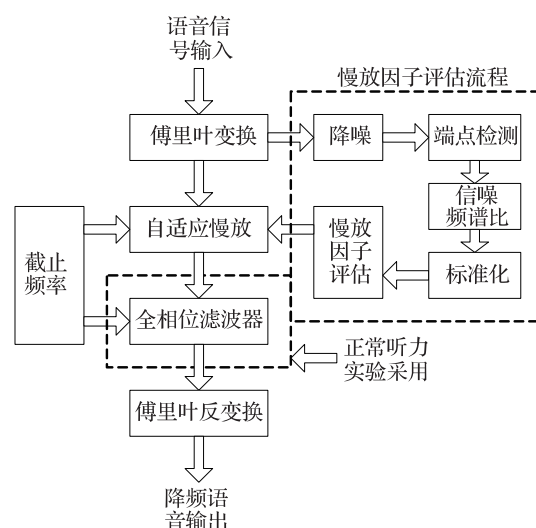


图 7 算法流程图

从流程图我们可以看出, 算法中最关键的步骤就是慢放因子的评估. 信号首先进行降噪, 本文采用相位补偿法^[24], 该法基于短时傅里叶变换, 不会显著增加计算量. 并且, 此处降噪的目的只是用来进行端点检测, 判断声音段和噪声段. 因为降噪处理会破坏信号的频谱成分, 所以用来评估噪声频谱和声音频谱的声音段仍是降噪前信号. 将按照倍频程划分的 6 个频域范围内的频谱叠加, 并归一化获得的频域直方图. 然后, 将信号的域直方图与噪声的域直方图相比, 并将获得比值进行标准化, 即按

照各频域范围内的归一化幅度大小排序, 并按照 2 的幂次等比排列. 标准化的目的在于控制慢放因子的取值范围, 便于比较分析. 通过处理标准化的频域直方图, 获得算法的慢放因子.

实际上, 对于汉语语音来说, 其最大的慢放速率只需调节为 1.5 倍 (信号先三倍升采样, 再两倍的降采样) 即可, 从而基本解决了普通交谈中的时间不同步问题. 而且, 日常交谈中, 往往不存在不间断说话的情况.

4 算法性能评估

4.1 汉语识别性能

图 3 显示了当慢放速率为 2 倍时, 音节/shi/的基频曲线与原始声音相比, 发生较大变化. 图 8 为自适应慢放频移算法下, 音节/shi/的四声基频曲线. 从图上可以看出, 基频曲线的变化规律和声调对应的曲线规律基本相似, 尤其是上声, 对照图 3, 可以看出明显的改善.

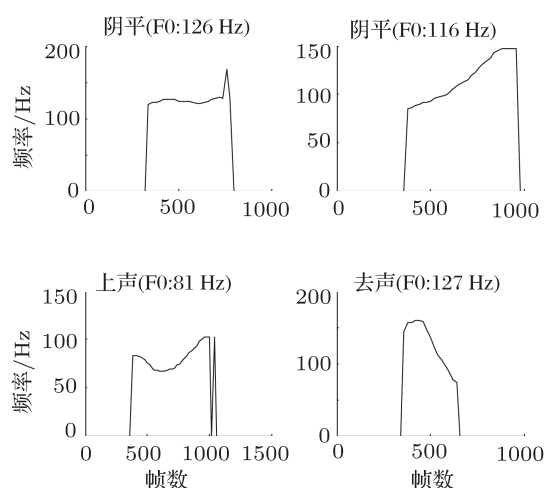


图 8 音节/shi/的四声基频曲线

表 6 声音识别性能比较

算法类型	音节/%	单词/%	句子/%
频移	75	70	80
非线性压缩	70	74	85
自适应慢放频移	86	92	92

为了进一步验证算法对汉语识别能力的改善, 实验比较了频移算法、非线性压缩和自适应慢放频移算法对汉语音节、单词、句子的识别能力. 所有听力材料按照声、韵、调平衡的方式构建. 表 6

为测试结果. 自适应慢放频移算法要好于其他两种算法, 平均改善 15%. 此外, 对句子的识别率要好于对音节或单词的识别率, 说明被试的推断能力起到了作用.

4.2 含噪声的识别性能比较

助听器作为听障患者的日常必备品, 在现实生活中, 必然遇到各种各样的噪声和背景声. 本文不讨论降噪算法对提高助听器性能的作用, 只是分析降噪算法在噪声环境下对声音的识别能力, 以及什么样的噪声对降噪算法的影响最大. 实验只比较了信噪比在 -10 dB 和 -5 dB 下的不同降噪算法下声音的识别能力. 在 -15 dB 情况下, 即使不降噪, 人耳也只能辨识 50% 左右的声音, 而 0 dB 时, 噪声的作用又不太明显. 图 9 显示了被试在 -10 dB 和 -5 dB 下对语音和环境声的识别结果.

从图中可知, 自适应频移算法的性能要优于其他两种算法, 但低于不降噪信号的识别率. 这说明降噪方法可以改善患者听力, 要达到正常人的听力水平很难. 在五种声音中, 对自然声音的识别较低, 因为自然声音的特性和噪声 (如工厂噪声) 很相似, 所以难以辨认. 因为语音的低频成分较多, 其慢放因子的变化较小, 不会因为慢放速率调节的误差导致对声音的识别错误. 因而, 算法对语音的识别率较高.

在对噪声研究上, 白噪声的频谱分布均匀, 对识别率的影响适中, 这说明在信噪比相同的情况下, 每个区域对识别率的干扰是不同的. 机枪噪声、军用车辆和车辆内部噪声对识别率的影响较低, 相比于其他噪声识别率至少高 15%, 其频谱分布较单一, 且都集中在 500 Hz 以下的低频区域; 识别最差的是高频信道噪声, 其频谱主要集中在 250 到 2000 Hz 之间, 且各频域的比例几乎相同, 而人类对 500 Hz 到 2000 Hz 的频率范围最敏感. 从其他噪声也可以看出, 频域 3 和频域 4 的能量高, 其识别率就低. 虽然识别率也有起伏, 如粉红噪声与 F16 座舱噪声相似, 识别率差别也在 15% 之内.

5 听障患者比较测试

最后, 本文采用自适应慢放频移方法对三位高频听障患者进行声音识别实验. 被试的基本信息如表 7 所示. 实验为期一月, 被试每周测试一次. 根据

被试的适配情况, 将降频声音刻成 CD, 要求被试每天至少听 1 到 2 个小时.

所有被试首先进行听力测试来确定截止频率, 使用的听力计为 Interacoustics AD226, 截止频率设定为患者听力增益达到 90 dB 的最小频率. 但是由于患者是未经过训练的, 所以初次测试时, 被试的识别率在 20% 以下, 将首先降低频率压缩后的语音级别 3—5 dB. 如果被试仍然不满意语音质量, 则增加截止频率. 然后重新生成降频语音, 并测试. 然后根据测试结果, 选择一个识别率最高的频率作为最终的截止频率. 截止频率的测试一般要做 3 到 5 组不同类型声音的测试, 以降低偶然性.

听障患者平均识别率如图 10 所示, 测试声音包括语音和环境声. 在第四周, 使用慢放频移算法比使用传统幅度放大助听器, 患者在识别语音和环境声上都获得一定改善, 平均 13%. 三位被试中, 平

均改善最大的是 1 号被试, 改善 19.5%; 最小的是 2 号被试, 改善 6.5%.

表 7 被试基本信息

被试	年龄	性别	听损原因	适配情况	截止频率
1	58	女	老年性	双耳	1.5/2
2	33	男	药物性	左耳	2.5
3	55	女	老年性	双耳	2/2

从图上可以看出, 基本上随着测试时间的增加, 患者的识别率增加, 这说明训练对提高声音识别率具有一定作用. 在对含噪声声音的测试中, 二号被试的识别率略有起伏, 这是正常的. 因为训练的作用也是有限的, 训练只能使患者逐渐辨识可以听到但不知道类型的声音, 对于本来就听不到的声音, 再训练也是没用的. 此时, 注意力的不集中, 往往就会造成识别率的轻微下降.

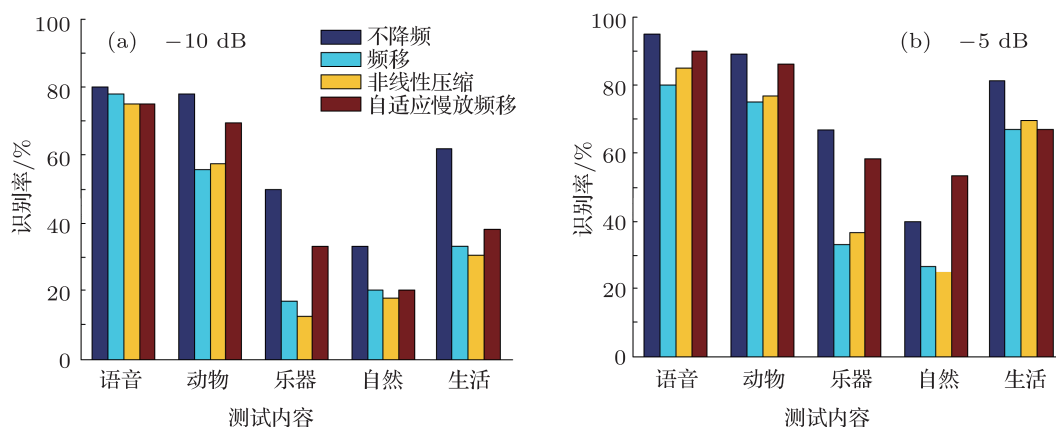


图 9 不同信噪比下的声音识别比较

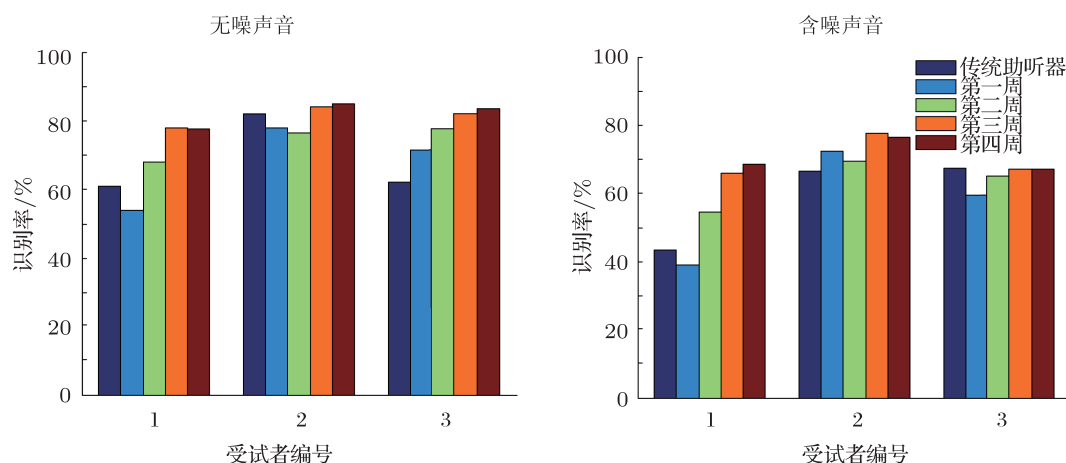


图 10 听障患者声音识别性能

6 结 论

本文通过分析传统降频算法的不足, 结合生活实际, 提出一种自适应的慢放频移方法. 算法能根据信号的频谱分布自适应的调整慢放速率, 并结合频移算法, 减少传统慢放算法时域不同步的影响. 而且, 通过分析含噪声音的特点, 提出一种噪声情

况下的改进算法. 在无噪声音和含噪声音测试中, 该算法同其他降频算法相比至少改善 15%. 同传统的助听设备相比, 患者的声音识别率也有显著改善.

将来, 应增加测试声音的数量, 进一步改善算法的细节, 如慢放因子的比例关系. 同时增加听障患者的受试人数, 深入分析患者的听力损失特征, 研究更细致的, 更有效的降频助听算法.

-
- [1] Stelmachowicz P G, Lewis D E, Choi S, Hoover B 2007 *Ear Hearing* **28** 483
 - [2] Stelmachowicz P G, Pittman A L, Hoover B M, Lewis D E 2002 *Ear Hearing* **23** 316
 - [3] Bohnert A, Nyffeler M, Keilmann A 2010 *Eur Arch OTO-RHINO-L*. **267** 1045
 - [4] Kluk K, Moore B C J 2006 *Hearing Res.* **222** 1
 - [5] Miller H D R, Nelson P B, Widen J E, Simon S D 2003 *J. Am. Acad. Audiol.* **12** 106
 - [6] Simpson A, McDermott H J, Dowell R C 2005 *Hearing Res.* **210** 42
 - [7] Simpson A 2009 *Trends in Amplification* **13** 87
 - [8] Posen M P, Reed C M, Braida L D 1993 *J. Rehabil. Res. Dev.* **30** 26
 - [9] Kuk F, Peeters H, Keenan D, Lau C 2007 *Hearing J.* **60** 59
 - [10] Kuk F, Keenan D, Peeters H, Korhonen P, Auriemma J 2008 *Hearing Rev.* **15** 32
 - [11] Glista D, Scollie S, Polonenko M, Sulkers J 2009 *Hearing Rev.* **16** 20
 - [12] McDermott H, Dean M 2000 *Brit J. Audiol.* **34** 353
 - [13] Arioiz U, Arda K, Tuncel U 2011 *Comput. Meth. Prog. Bio.* **102** 277
 - [14] Liu Zh J, Shao J, Zhang P Y, Zhao Q W, Yan Y H, Feng J 2007 *Acta Phys. Sin.* **56** 7064 (in Chinese) [刘赵杰, 邵健, 张鹏远, 赵庆卫, 颜永红, 冯稷 2007 物理学报 **56** 7064]
 - [15] Yang L, Zhang J P, Yan Y H 2010 *Acta Acustica.* **35** 248 (in Chinese) [杨琳, 张建平, 颜永红 2010 声学学报 **35** 248]
 - [16] Bao Z W, Wei R J 1960 *Acta Phys. Sin.* **16** 338 (in Chinese) [包紫薇, 魏荣爵 1960 物理学报 **16** 338]
 - [17] Zhang H 2004 *Hearing aid* (Beijing: People's Medical Publishing House) p11 (in Chinese) [张华 2004 助听器 (北京: 人民卫生出版社) 第 11 页]
 - [18] Zhang J L, Qi Sh Q, Yu G 1998 *Acta Acustica.* **23** 19 (in Chinese) [张家驷, 齐士铃, 俞舸 1998 声学学报 **23** 19]
 - [19] Turner C W, Hurtig R R 1999 *J. Acoust. Soc. Am.* **106** 877
 - [20] Kuk F, Keenan D, Peeters H, Korhonen P, Hau O, Andersen H 2007 *Hearing Rev.* **14** 60
 - [21] Simpson A, Hersbach A A, McDermott H J 2005 *Int. J. Audiol.* **44** 281
 - [22] Wang N, Chen K A, Huang H 2009 *Acta Phys. Sin.* **58** 7330 (in Chinese) [王娜, 陈克安, 黄凰 2009 物理学报 **58** 7330]
 - [23] Chen K A, Wang N, Wang J Ch 2009 *Acta Phys. Sin.* **58** 5075 (in Chinese) [陈克安, 王娜, 王金昌 2009 物理学报 **58** 5075]
 - [24] Wojcicki K, Milacic M, Stark A, Lyons J, Paliwal K 2008 *IEEE Signal Proc. Lett.* **15** 461

Experimental study and improvement of frequency lowering algorithm in Chinese digital hearing aids*

Liang Rui-Yu^{1)2)†} Xi Ji¹⁾²⁾ Zhao Li²⁾ Zou Cai-Rong²⁾ Huang Cheng-Wei²⁾

1) (College of Computer and Information, Hohai University, Nanjing 210098, China)

2) (School of Information Science and Engineering, Southeast University, Nanjing 210096, China)

(Received 17 October 2011; revised manuscript received 21 November 2011)

Abstract

Frequency lowering algorithms are the most effective and secure for improving the auditory identification capability of hearing-impaired patients. Through analyzing the current algorithms by the test of subjective auditory approach, in the paper we propose an adaptive algorithm combining slow playback algorithm with frequency transition algorithm. The algorithm can adaptively adjust playback ratio according to the signal spectrum structure to reduce the desynchronization in time region. Furthermore, a method of evaluating the playback ratio in noise is designed based on the frequency spectrum relation between noise and noisy signal. Results showed that compared with other frequency lowering algorithms, the proposed algorithm can improve the auditory identification capability from 15% to 20%. In addition, in the test for hearing-impaired person, the significant improvement on the average auditory identification capability is also obtained compared with the conventional hearing aid device.

Keywords: frequency lowering, Chinese hearing aid, hearing loss in high frequency, adaptive playback

PACS: 43.71, 43.72

* Project supported by the National Natural Science Foundation of China (Grant Nos. 60872073, 51075068, 60975017), the Fundamental Research Funds for the Central Universities of Ministry of Education of China (Grant No. 2009B32614) and the Natural Science Foundation of Guangdong Province, China (Grant No. 10252800001000001).

† E-mail: lly1711@163.com