

中文语音处理在数字助听器系统的开发

目前国外对助听器研究发展的一个热点则是集中在中国,确切地讲是基于对汉语语言和语音研究,开发相关的语音识别技术和产品。以汉语为中心的中文听力学也不例外。我们已经知道听觉科学是一门发展迅速、知识更新很快的一门学科,它所研究的对象以人的听觉为中心,现在我们将介绍和讨论科学家和听力学家更关心的是怎样将听觉科学运用到中国人的听觉和言语实际中去。

目前研究的结果表明,不同的语系,如汉语和英语有各自特点,在听觉感知过程中有很大差别。英文和汉语在语音和口语上有重要区别, Ming-Xi Tsai et al (2000) 认为汉语和英文语音在结构上特征区别很大。汉语的词、字、音节和声、韵母分节含有不同层次的信息,并保持复杂的关系。在口语中,汉语发音差异也很大,在不同会话条件下,受到这些结构中不同层次信息的影响。

对中文语音识别和中文语音音调的研究表现在人工耳蜗的算法上面。言语处理策略是人工耳蜗帮助患者理解语言核心技术,已有大量研究。但对于言语声尤其是声调、语调的研究,比如对以语调为基础的汉语研究还是很少。在最近的一个试验中,他们用澳大利亚人工耳蜗来观察对汉语语音理解的影响。结果表明在某些言语处理策略中汉语的使用的理解度要高于其他时间策略。如果能提高刺激率,加强对语音和音调的理解,他们也认为不同的言语处理策略对汉语说法也有理解。研究再次证明,汉语应该有一定的语音系统处理自己的语言,尤其是对听障人士尤为重要。

美国麻省理工学院研究者 Michael Qin 在其《在噪音背景发音和音调的辨认》的试验中,对汉语普通话音调的辨认和噪音的关系进行了研究。他认为不同的语言利用不同类型的声调使我们口语富于不同的意义,在噪音环境下这些有意义的声调会受到影响,因此他需要发现说普通话的中国人怎样在噪音环境里辨认不同的声调。在试验中他使用 6 个辅元音的音位,同时使用 4 个声调:阴阳上去。其结果表明在信噪比降低的情况下,对汉语声调和元音的识别受到很大影响,从而影响降低言语的理解能力。因此信噪比影响理解中文很重要的因素。这个试验对听觉康复和设计有针对性的助听器具有重要意义。

同时,最近美国成立了一个综合性专家研究小组,开始研制适合中文语音的助听器。该小组成员包括世界闻名的豪斯耳研究院、香港中文大学耳鼻喉科等。与上述研究类似。他们认为在聆听以声调作为识别语音和语义的语言时,如普通话、广东话和泰国语等,可能听觉更重要的是依靠基频相关的信息来理解语言,这是与别的语言不同的。因此,在研制助听器时,我们应该考虑到这些患者的语言特点。

一. 中文语音技术处理原理

中文语音处理策略的英文词有“Chinese speech processing strategy”或“Chinese speech recognition”,(中文语音识别)和“hearing aid algorithm”(助听器算法)等。其中,algorithm,即“算法”一词

使用较多,尤其涉及到数字助听器的开发,“算法”代表了某一特殊技术的核心。“算法”可以被简单地看作为实现某些特定信号处理功能的指令序列。中文语音特征可以通过算法研究来形成的。数字信号处理器和算法构成了数字助听器的 DSP 线路。包含多通道动态范围压缩、噪音衰减等处理,设计助听器的算法的主要目标是利用中文语音处理技术,即使在不同的听音环境中,须确保言语被听见并聆听舒适。同时,利用数字助听器改善汉语可懂度,使有听力损失的中国患者能更容易地理解汉语。

汉语是声调单字语言,声调是汉语的重要语音特征之一。声调特征主要体现在嗓音基频随时间变化的模式上。Eady 技术(1982)曾考察过声调语言——汉语的基频模式与重音语言——英语有什么不同。汉语的声调在词语中具有辩意作用,在生活实践当中,大家也都能体会到声调有助于我们听懂别人的话,而“南腔北调”常表示不易听懂和不大好懂和不大好听的意思。

对于连续言语来说,长时间平均的正、负颤动因数,各种语言和男女发音人是差不多的。只是负颤动总是比正颤动大,而且出现频率也较高。Eady 的测量结果表明,汉语的说话速度要比英语慢一些。这可能是由于说汉语时,说话人要花更大的努力在每一个音节上来控制声带运动,也就是说声调语言的音节喉运动控制有较大的语言学负荷,因而花的时间就多一些。结果就表现为说话慢一些。

因此,声调信息主要存在于基频随时间的变化中,强度变化对声调信息有补偿作用,以及清辅音的存在与否对声调清晰度是有一定的影响的。

1 原理 (Principles)

本文介绍一种可应用于数字助听器的提高汉语可懂度的语音处理方法,其目标是使以汉语为母语的听残人士能更容易地理解语言。增强言语可懂度的思路来源于人们的实践经验。回想一下,当你为使一个有听力障碍的人更容易听懂自己说话时所采用的方法:你不仅仅要提高音量,而且还要改变发音方式,说得更慢和更清晰。一些研究表明清楚地读无意义的语句,比在日常会话句子,大约能提高 17% 的单词可懂度。这里所谓说得更清晰是指强调言语信号中的某些暗示,这些暗示有许多不同的形式,如特定音段的持续时间,元音的共振峰位置或者音素之间的过渡等。

不是所有人都会简单地、方便地对听力损失患者“清楚地”讲话。因此,我们要采用言语增强的方法就是在说话人和听话人中间构造一个处理模型,该模型能强调并突出语句中的特定成分,使语句听起来更清晰。

一切语音之所以能够表达意义,是由于各个音之间存在区别。这些区别产生于声腔内部的器官和肌肉等活动决定的发音方法和发音部位的差异,同时又表现为语音的声学特征的差异。本文提出的言语增强的方法正是通过对语音信号的重构来强化这些差异。所谓重构是指对语音信号中不同性质的信号进行识别并有针对性地予以处理,强调其中对人的感知起作用的特征,从而达到提高言语清晰度的目的。该方法可以简单地概括为:放大辅音、强调重音和突出声调。

2 汉语语音信号的感知特点

2.1 声调

声调的调类。

声调的感知。

主要依据基频的变化。

声调音高的变化对音长和音强都可能产生影响。

2.2 重音

轻重音的声学特性。

与实际音强有密切关系，但并不相等。

还要受音色、音高和音长的制约。

感知特点：分辨轻重音时，音强往往并不是决定性因素。

1) 辅音放大 (Consonant Amplification)

言语知觉的心理实验证实了如下特点：人在言语感知过程中，对语音信号载荷的关于发音方法和发音部位的分辨信息的知觉能力存在强弱差别。总体上说，人对发音方法比对发音部位有更好的分辨能力。而方法清晰度与辅音清晰度的关系很相近。在汉语辅音的发音方法的知觉重要性中，存在清与浊、送气与不送气、摩擦与非摩擦的从强到弱的位次关系。研究表明，相对加强辅音有助于改善语音清晰度。

Kates 描述了放大辅音的方法，图 1 是其中被广为采用的一种模型。该系统把信号分解成几个波段，在每个波段检测短时谱形，根据谱形识别元音和辅音，对辅音则给予放大。需要指出的是，杜利民等提出了汉语语音导引特征的概念，从声学信息计算检测的角度为汉语自动语音识别系统提供了一种辅助匹配结构。

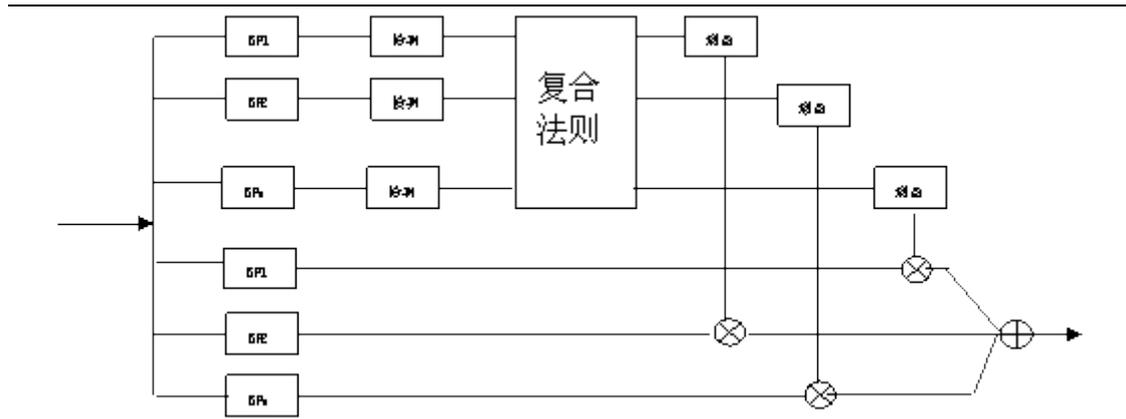


图 1 辅音增强系统

2) 重音 (Stress)

组成一段语流的各音节声音响亮程度并不完全相等。有的音节在语流中听起来声音比其他音节响亮，这就是重音音节。有的重音和语义、语法有密切关系，如汉语普通话中的词重音。词重音出现在词中，是由于词的含义不同，重读音节的位置也不同。如“技术”和“计数”，重音分别在第一音节和第二音节。这种语意的区别是通过“超音段特征”来表达的。

在汉语中，重音对韵律特征参数的影响倍受关注。语流中“韵律特征”（prosodic feature）是通过音高、音长和音强的变化，即“超音段特征”表现出来。从语图上观察，音域明显扩张重音的特点。高明明对普通话语句汇总强调重音的声学表现进行了研究，指出：

（1）“音高升高是普通话语句中强调重音的重要韵律特征”。

（2）音高和时长对于强调重音的实现具有同样重要的作用。它们之间的关系是对立互补的。

语音合成的经验告诉我们，音高是调节重音最有效的手段，所以强化重音的方法主要是提高音高。

3) 声调 (Tone and Internation)

一个音节除了包括由元音和辅音按时间顺序排列成系列的音质单位以外，还必须包括一定的音高、音强和音长。在一些语言里，音高在音节中起的作用可以说是和元音、辅音同样重要，这种能区别音节的意义的音高就是“声调”。根据声调的有无可以把世界上的语言分为声调语言和非声调语言两大类。汉、藏语系语言最突出的一个特点就是有声调。

汉语普通话的声调起着构词辩意的作用。对于具有相同拼音的一个音节，由于声调不同，可以具有不同的含义。普通话单音节的声调变化共有四种模式，不同的声调反映在语音参数上是基音频率轨迹的变化不同。根据实验观察所定义的

一些规则，可以认为基音频率轨迹的某一参数超越某一预先确定的门限时，则可判为某一声调类型。在此基础上，黄泽镇、杨行峻提出的识别模式采用基音轨迹曲线的一、二次斜率、谷点和平坦度对四种声调有很强的区别性，实验表明，这一算法的结果识别率可达到 99%。

林茂灿指出声调信息主要存在于主要元音（及其声学过渡）上。考虑到声调音高的变化，对音长和音强都可能产生影响，即：去声最短、最强，上声最长，最弱，阴平和阳平举重，阳平又往往比阴平略长一些。声调的增强不能简单地对主要元音进行放大，而应该不同的声调在音高和音强上有不同的处理。实际应用中我们采取如下策略：

- (1) 对去声增强音强。
- (2) 对上声加大音长。
- (3) 对阴平和阳平不改变。

图 3 展示的 4 条声学曲线分别描述了四声在不同时间里的频率特征。



图 3 汉语四声的声调声学特征

2. 方法 (Methodology)

数字助听器的核心部分是增益计算，基于频域的处理过程，它建立了各频率段的输入瞬时能量与增益的函数关系，如图 3 所示，对每个频段的瞬时能量进行短时能量累计和长时间慢速平均可获得信号识别和分类所必要的的数据。其中：

- (1) $E_j(n) = a E_j(n-1)$ 式中： a 是时间常数。
- (2) 使用倒谱算法提取基频， 512 个点 FFT ， 40ms 汉明窗，窗移为 10ms。
- (3) 用一个简单的滑动平均算法对每个音节测到的基频进行平滑处理，剔除那些平滑段内偏离均值过大的值。
- (4) 音高和音长分别进行归一化。

(5) 采用一个二次曲线在最小均方误差的意义下逼近基音轨迹。并计算曲线的一次斜率、二次斜率、谷点和平坦度。

上述算法采用基于 TOCCATA 指令系统的汇编语言实现。14 位 A/D ，采样率设为 32KHz 。

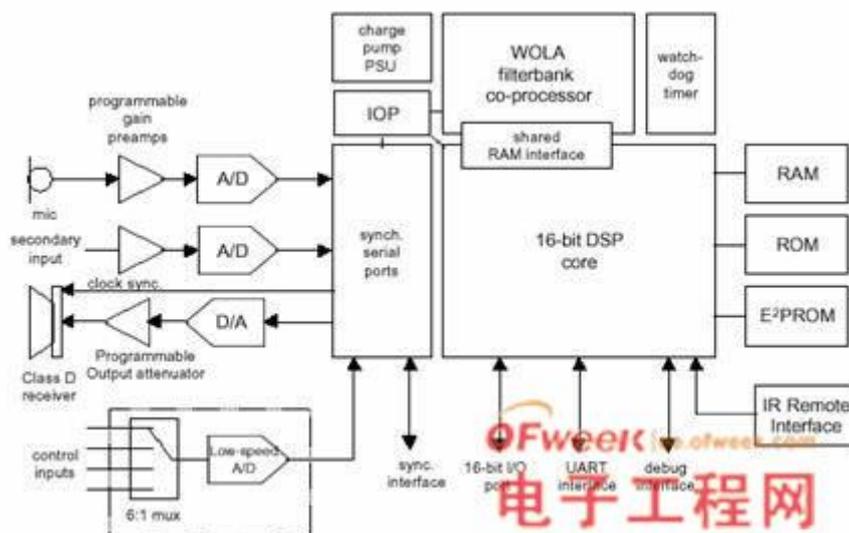


图 3 . 汉语言语增强系统处理结构图

1). 语音的切分 (Classifications of Phonemes)

声波由音质（即音色）、音高、音强和音长四部分组成，这四部分在语音中起着不同的作用，但在时间上又是同时并存的。

音质成分 —— 按音节划分，如元音、辅音。

超音质成分 —— 由音高、音强和音长三部分组成，附着于一个音节或音段上。

从声波特性上看，可以由基频确定音高，根据振幅确定音强，根据时间确定音长。

2). 处理原理 (Algorithm Principles)

中文语音处理主要体现在：

在验配过程的中，考虑中文语音长时间频谱覆盖的频率作加权处理，抬高目标曲线中言语频率的部分，可以达到加强语音理解的作用。

在助听器的信号处理程序中，对压缩控制器做特别的设置，使对高频的信号压缩的启动时间和释放时间很短，做到使辅音清晰化的效果，增强使用者对言语的理解度。

在降噪处理中，根据中文语音在噪音环境中的采样分析，得出了为中文语音优化的降噪策略。实验证实，该策略最高可以提高信噪比 18dB。

二. 中文语音处理技术在涉及助听器的应用

下面是将中文语音技术应用到设计助听器的具体实例。这项技术采用了目前世界上最先进的 DSP 数字技术，包括低功耗的数字芯片。

1. TOCCATA 数字信号处理系统

Toccata TM 系统是微型、超低功耗、高效率的数字信号处理系统。它包括一个高保真加权叠加滤波器组（WOLA filter bank）、一个 16 位 DSP 核心、两个 14 位 A/D 转换器、一个 14 位 D/A 转换器和其它外围设备。Toccata TM 技术提供标准的软件可编程的 DSP 开发平台和采用 0.18 μ 工艺制造的微型超大规模集成电路。它不但为音频处理系统制造商也为其它基于 DSP 的微型、低功耗产品的开发提供了便利。

1.1 硬件结构 (Hardware Structure)

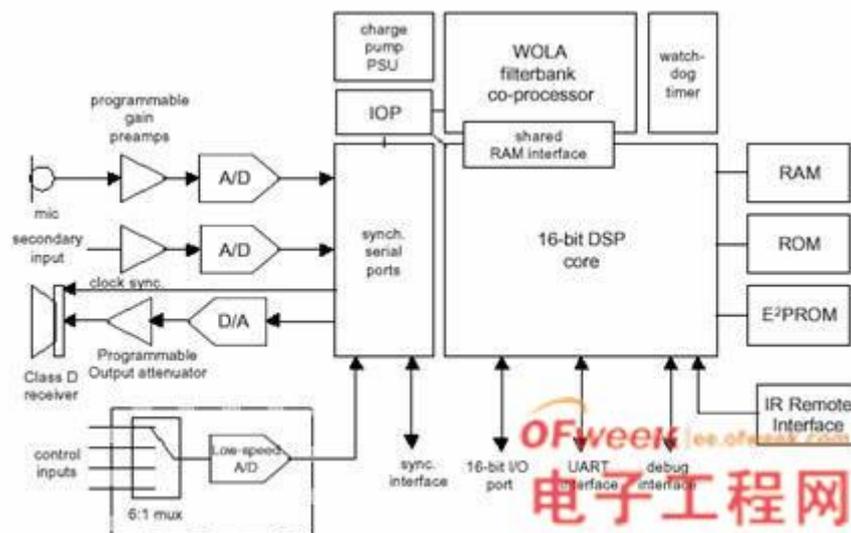


图 4 硬件系统结构图

TOCCATA 系统由三块芯片组成，一个“模拟”芯片（ALPHA），一个“数字”芯片（DELTA），和一个用于无电存储的 E2 PROM 芯片。

1.2 ALPHA 芯片

ALPHA 芯片包括输入和输出放大器，二个 A/D 转换器，一个 D/A 转换器，以及主时钟及供电系统。

1.3 DELTA 芯片

DELTA 芯片包括了 16 位软件可编程 DSP 核心，一个 WOLA 滤波器组协处理器，一个 DMA 控制器（输入输出处理器或 IOP）和存储器（RAM 和 ROM）。可编程核心和灵活的滤波器的组合允许通过软件改变信号的处理方式。因而，该结构可以执行传统的音频处理系统处理方案（例如双通道压缩），当然通过 DSP 核心，也可以执行更强大的处理方案（例如 16 通道乃至更多通道的压缩，降噪，抑制反馈等）。

1.4 DSP 核心和指令系统 (DSP Core)

RCORE 是一个灵活的 DSP 核心，采用带单周期连乘累加操作和 40 位累加器的双哈佛结构。外围组件通过一个由扩展寄存器、存储器映象寄存器和共享存储器组成的复合体提供。

1.5 信号路径

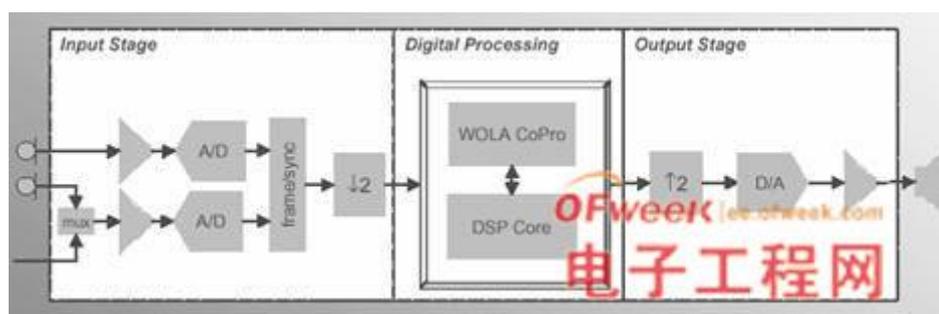


图 5 . Toccata 系统提供的信号路径:

2 Intelligia 数字助听器结构

Intelligia 全数字助听器是基于上述介绍的芯片的技术特点设计的，其结构可由图 6 表明。尽管同模拟助听器一样，数字助听器也使用麦克风和接收器作为能量转换器，但在数字信号处理器中经 A/D 采样后，电平信号已被转化为数字编码。数字编码能够非常灵活地被加以利用来提供增益、改善频响，或按患者听力的要求作其它处理。当 DSP 算法完成后，数字编码又被 D/A 转换为电平信号，并经由接收器转换为声音。

数字助听器的关键在于具有信息处理系统，这里以一款基于现进数字信号处理系统 Toccata TM，开发出的全数字助听器 Intelligia，具有独特的中文语音处理功能。助听器在设计中将信号分解成 16 个波段滤波处理，再将 16 个波段的信号组成 10 组通道，每个通道独立使用输入自动增益控制方法 (AGCi)，对信号进行压缩处理，每个通道使用快慢两个时间侦测器，快速时间侦测器用以

监视信号较快的变化，而慢速时间侦测器侦测较慢的信号变化，也就是音节的变化，并且选用与中文语音变化相匹配的压缩、释放时间常数，达到更好的听觉效果。

全数码助听器技术特点：

1) 中文语音信号处理

深入研究中文和其他声调语言的发声特征后，我们把独创的中文语音处理的技术置入，使它能大大提高在中文语言环境中听音的可懂度。

2) 更快

采用专为数字助听器设计的第 3 代数字助听器处理系统 TOCCATA，它强大的运算能力使能快速处理各种语音信号。

3) 更省电

工作电流不足 1 毫安，并且它能在无信号输入时自动进入省电模式，如此低的能耗免去配戴者经常更换电池之苦。

4) 完全可编程

通过其可编程的优点为听障者配置最适合的听力补偿程序和参数，从而保证配戴者可以获得最佳的听音效果。

5) 多通道独立压缩

把外界的声音按频率细分成多个波段和通道，并对每个波段和通道的信号进行不同的处理，从而保证配戴者听到更清晰、更逼真的声音。

6) 降噪处理

能有效地抑制环境噪音，提高辨别语言的能力，从而保证配戴者无论是在嘈杂的大街还是在喧闹的超市都能听到清晰的声音。

7) 方向性处理

可配置一个方向性麦克风系统和相应的软件，能使降噪的效果更好，从而保证配戴者听到更清晰、更自然的声音。

8) 声反馈抑制

助听器在使用过程中容易引起啸叫，这种现象即声反馈。采用了声反馈抑制技术，能有效地抑制声反馈的出现，使配戴者听到更为舒适的声音。

9) 可轻松升级

由于采用全开放的数字信号处理（DSP）平台 TOCCATA 技术，提供了可编程的能力，具有充分的适应性以及升级能力，因此，配戴者只要采用我们的软件，就可以立刻享受到最新的功能。下面是这款中文语音处理的技术指标对比：

	中文语音技术处理助听器	其它上市产品	其它上市产品	其它上市产品
功能	Intelligia	助听器A	助听器B	助听器C
增益/最大输出	120/60	123/54	124/55	125/56
通道	10	14	2	4
波段	15	14	7	8
压缩技术	宽频动态	宽频动态	宽频动态	宽频动态
聆听程序	3	2	2	2
声反馈控制	有	有	有	有
降噪技术	有	有	有	有
方向性处理	有	有	有	有
电感	有	有	有	有

表 1 中文语音技术处理助听器和其它助听器的技术比较

在实验室中，具有中文言语增强方法的数字助听器，初步实验的结果表明，中文语音处理技术的运用，可帮助以汉语为母语的 patient 更好地理解语言，提高康复水平。在临床使用中，佩带 Intelligia 助听器的病人感觉效果很好，尤其在噪声环境中，增强了语音清晰度。从某种意义上讲，病人感到理解语言的能力得到提高。当然，我们必须意识到中文语音处理技术在全数码助听器中的运用仍处在早期研究阶段。笔者认为听力学科学家和助听器专家应从下列几方面作更深入的研究：

应对以英文和汉语为基础的语音处理技术作深入的对比研究，尤其是在噪音环境中，观察两种技术分别对两种语音不同处理的效果。最理想的实验条件应是利用有双语能力的受试者参加。

将中文语音处理技术和目前使用的非线性助听器验配法结合研究，观察以英文为基础制定的验配方法，是否在中文语音处理技术支持下，更有效地帮助以汉语为母语的 patient 在日常生活中提高言语理解能力。

中文语音处理技术目前是人机对话的研究热门之一，其算法复杂多样，我们应该更深入地研究具有中文特色的助听器技术算法，充分发挥数码芯片的巨大潜力。

将中文语音处理技术运用到听力器设备中去才刚刚开始，这是一项非常复杂、涉及许多尚未解决的技术问题的课题。但是，笔者认为只有开发出具有汉语语音

特色的助听器，才能更有效地帮助众多的以汉语为母语的听残者。