

[编者按] 本刊开办海外园地栏目,旨在每期介绍一位在海外定居,学有成就的华裔学者,他们将介绍国外最新的临床、科研动态,同时欢迎对国内专业发展提出建设性意见,畅所欲言,批评争鸣,以诚恳求实的治学态度与国内同仁同道共勉。同时借本栏目广招天下饱学贤良之才,热情杂志建设之士,共办 Journal of Otology,为促进国际间合作搭建一个学术交流平台,有效快速地将中国不断发展的耳科学成果向世界展示,欢迎广大海外学者积极投稿。

本刊编辑部

言语识别中的时域及频域信息

徐立

School of Hearing, Speech and Language Sciences, Ohio University, Athens, Ohio, 45701, USA

【摘要】 本文对言语识别中的声学要素从时域和频域的角度进行探讨,旨在为人工耳蜗编码策略的改善提供理论依据。声码器技术被用于一系列的实验以确定时域和频域信息对言语识别和汉语四声识别的相互作用。频域信息是由声码器中的频道数来决定,而时域信息则是由声码器的低通滤波器的截止频率来决定。听力正常成人参加了各项感知试验。结果表明,时域和频域信息都对音素识别很重要。在安静环境下,辅音和元音识别率分别在 8 和 12 频道及 16 Hz 和 4 Hz 的低通截止频率时达到平台成绩。在噪声环境下,元音识别受益于增高的频道数。汉语四声的识别需要 256 Hz 的低通截止频率才达到平台成绩,这一频率比英语音素识别所需的时域信息高得多。声调识别率在本研究中最高频道数 12 时仍未见饱和。为了研究细微结构和时域包络对四声识别的相对重要性,我们用声嵌合技术将不同声调信号的时域包络和细微结构进行对换。感知实验结果表明,声调识别主要取决于细微结构,这一点与音乐感知的结果类似,而不象言语识别,后者主要依赖于时域包络信息。因此,增加人工耳蜗系统中有效的频道数将有助于尤其是噪声环境下的言语识别。将人工耳蜗刺激中提供更多的细微结构信息可能会提高患者声调识别的成绩。

【关键词】 人工耳蜗; 言语识别; 声调识别; 时域信息; 频域信息

【中图分类号】 R339.16, H018.4 **【文献标识码】** A **【文章编号】** 1672-2922(2006)04-0335-08

Temporal and spectral cues for speech recognition

XU Li

School of Hearing, Speech and Language Sciences, Ohio University, Athens, Ohio, 45701, USA

【Abstract】 The present study explores the temporal and spectral cues for speech recognition in an attempt to provide information for improving the speech processing strategies in cochlear implant systems. A noise-excited vocoder was used in a series of experiments to determine the relative contribution of temporal and spectral cues to phoneme recognition and lexical tone recognition. Spectral information was controlled by varying the number of channels and temporal information was controlled by varying the lowpass cutoff frequencies of the envelope extractors. Normal-hearing adult subjects participated in the perceptual tests. The results demonstrated that both temporal and spectral cues are important for phoneme recognition in quiet and in noise. The plateau performance for consonant and vowel recognition in quiet was reached when the number of channels was 8 and 12, respectively and the lowpass cutoff frequency was 16



作者简介 徐立,湖南长沙人。1985年毕业于衡阳医学院,1988年从首都医科大学研究生毕业后,留在北京同仁医院、北京市耳鼻咽喉科研究所,在刘铤教授的指导下进行耳科临床和研究工作。1991年应邀赴瑞士巴塞尔大学做访问学者,师从世界著名耳鼻咽喉科专家 C.R. Pfaltz 和 R. Probst 两位教授。1992-1999年跟随 J.C. Middlebrooks 教授在美国佛罗里达大学和密西根大学完成神经科学博士。之后在密西根大学 B.E. Pflingst 教授的人工耳蜗实验室做博士后研究。自 2001年起在美国俄亥俄大学任教,现为听力学副教授,并任首都医科大学耳鼻咽喉科客座教授。在国外多个科学杂志上发表论文 20 余篇,其中包括在 Science 和 Nature 杂志上各一篇。

基金项目 美国 NIH (F32-DC00470, R01-DC03808, R03-DC006161.) 和俄亥俄大学研究基金。

and 4 Hz, respectively. In noise conditions, vowel recognition benefited from increased spectral resolution. For Mandarin-Chinese tone recognition, the lowpass cutoff frequency required for asymptotic performance was 256 Hz, much higher than that required for English phoneme recognition. Tone recognition performance had not yet reached plateau when 12 channels, the highest in this study, were used. To study the relative importance of fine structure and temporal envelope in lexical tone recognition, a separate experiment using the auditory chimera technique was carried out. The perceptual results demonstrated that tone recognition relies more on the fine structure as does melody perception rather than on the temporal envelope as does English speech perception. Therefore, to improve speech recognition, especially in noise, efforts should be concentrated on providing more effective channels in the cochlear implant systems. Lexical tone recognition could benefit from fine structure information presented in the cochlear implant stimulations.

【Key Words】 Cochlear implant; Speech perception; Tone perception; Temporal cues; Spectral cues

前 言

言语信号中含有丰富的信息。这些信息可认为是冗余的 (redundant),因为在相当程度上削弱这些信息时并不影响言语识别。人工耳蜗系统便是一个很好的例子,几乎所有的人工耳蜗系统都是将声信号进行带通滤波,然后将不同频率的信号送到耳蜗不同的部位,通常这种信号是以调幅脉冲波的方式来刺激听神经,因而,调幅波的调制频率携带着重要的时域信息,而频域信息则体现在耳蜗内电极刺激之部位^[1]。

众所周知,人工耳蜗患者获取的频域信息比听力正常人少得多。Friesen^[2]等(2001)发现多数人工耳蜗患者最多获得7至8个频率通道的信息,而听力正常人至少使用20至30多个通道的信息。但是,人工耳蜗患者获取时域信息的能力可能与听力正常人相当^[3,4]。当然,通过电刺激传送给耳聋患者的时域信息很可能与听力正常人听到的时域信息有质的不同,因为电刺激会引起听神经高度同步的放电^[5,6]。尽管如此,人工耳蜗患者的言语识别可以达到相当好的水平,当然在许多方面仍有待改进,如在噪声环境下的言语识别的方面便是一大难题。Skinner^[7]等(2002)对62名人工耳蜗患者的言语识别进行了研究,发现1/3的受试者在+10 dB信噪比的噪声下句子的识别率不到75%,而听力正常人在这种信噪比的噪声下句子识别率是接近于100%。因此,为了进一步改善人工耳蜗患者言语识别的能力,我们有必要探讨频域和时域信息对言语识别的相对作用以找到弥补各自不足的方法。

以往有不少研究探讨了频谱分辨率(spectral resolution)对言语识别的影响。人们用各种方法将言语信号的频谱分辨率削弱以达到频谱模糊化(spectral smearing)的效果,在这种削弱言语信号频谱分

辨率的情况下,言语识别受到不同程度的影响^[8-13]。然而,Baer和Moore^[11,12]也发现,频谱模糊化使得模拟的听觉滤波器的带宽为正常的6倍时,无噪声情况下的言语识别几乎不受影响。可见言语识别对频谱分辨率的要求并不高。Hill^[14]等(1968)用声码器(vocoder)技术(有关声码器技术的描述见下文)探讨了频谱信息通道数对言语识别的影响。他发现6到8个频道可达到较好(~70%准确率)的言语识别。Shannon^[15]等(1995)在对经过较多训练的受试者测试后发现,只需4个频道的信息便足以达到良好(>85%的准确率)的言语识别。近年的文献一致认为,要达到良好的言语识别只需要为数不多的频道信息,根据言语材料的不同,这一频道数为4~8^[16-19]。儿童受试者所需的频道数要高一些^[20]。

另外一些研究则将注意力集中在时域信息对言语识别的作用上。Rosen^[21](1992)将言语的时域信息分为三类,即包络(envelope)、周期性(periodicity)和细微结构(fine structure),三者的频率范围分别为2-50 Hz、50-500 Hz及500-10000 Hz。由于较高频的细微结构主要由部位编码(place code)而决定,因而在谈到时域信息时主要指低频的包络和周期性信息。Drullman^[22,23]等的时域包络模糊化(temporal envelope smearing)的研究及Shannon^[15,24]等的声码器的研究均指出,言语识别只依赖于低于16到20 Hz的时域信息,也就是说,包络中的低频的调幅信息对言语识别起决定性的作用。

我们近年来对言语识别中时域及频域信息的相对作用进行了一系列的研究^[8,19,25,26],从辅音到元音、到汉语声调的识别,从安静环境到噪声环境,逐项进行了探讨,下面,我们总结我们在这些方面做的研究,并结合文献中最近的数据进行讨论。首先,我们有必要将这些研究的关键方法——声码器技术进行介绍。

人工耳蜗的声学模拟——声码器技术

声码器技术早在上世纪 30 年代便由 Dudley^[27] 提出,用于言语的分析及合成。在沉寂多年后,Shannon^[15] 等(1995)将声码器作为模拟人工耳蜗刺激的一种声学的方法,倡导其研究应用。近十年来,不少研究室纷纷采用这一技术来研究人工耳蜗刺激情况下,言语识别或音乐识别可能达到的成绩。声码器的信号处理与人工耳蜗言语处理方案大同小异,只是最后一步是合成为声信号,而非电刺激,这样听力正常人可以听到,从而达到模拟人工耳蜗刺激的效果,人们普遍认为这种方法代表着人工耳蜗患者最佳成绩的状况^[2,24,28-35]。而且,人们可用这一技术来模拟一些在人工耳蜗患者身上无法调控的参数,如电极的数目或电极插入的深度^[36-38]。

声码器的信号处理流程图见图 1。首先,言语信号通过一组带通滤波器的滤波,带通滤波器的数目可用于控制频域信息的多少,我们多用 1 至 16 个带通滤波器,以模拟 1 至 16 个人工耳蜗通道的情形。各带通滤波器的中心频率及截止频率可通过 Greenwood^[39] 的公式推算,该公式即: $F=165.4(10^{0.06x}-1)$,其中 F 为频率, x 为离蜗顶的距离(mm)。下一步便是通过半波整流 (half-wave rectification) 及低通滤波的方法将各通道的时域包络进行提取,低通滤波器的截止频率决定时域信息的多少,我们一般用从 1 至 512 Hz 以倍频程递增的多个频率,即 1, 2, 4, 8, ..., 256, 512 Hz。然后将一个与言语信号时程相等的白噪声进行带通滤波,这一带通滤波器组与第一步中对言语信号进行滤波的

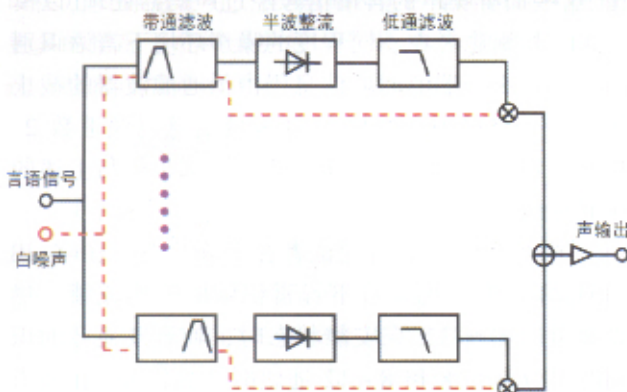


图 1 声码器信号处理流程图。言语信号经过带通滤波分成 N 个频道。各频道的时域包络在整流和低通滤波后被提取。低通滤波器的截止频率的高低决定时域包络信息的多少。白噪声经过同一带通滤波器组过滤后被提取的言语信号的包络所调制。叠加各通道的被调制的噪声后传给耳机或扬声器播放。

带通滤波器组实为一致。过滤后的噪声用提取的包络进行调制。最后,调制后的各通道的噪声被相加起来,这一合成的信号便携带有数个通道的时域包络信息,送往耳机或扬声器,听力正常人便可听到人工耳蜗的声学模拟。

音素识别中的时域和频域信息

如前所述,人们对言语识别中的时域信息或频域信息分别做了不少研究,而对两者同时进行研究的工作则未见报道。我们在这里将两者同时进行调控,以研究两者在言语识别中的相互作用。

受试者为 7 名听力正常母语为英语的成人。言语测试材料为辅音和元音两部分,辅音取自 Shannon^[40] 等(1999)的 20 个辅音,即 ba, cha, da, fa, ga, ja, ka, la, ma, na, pa, ra, sa, sha, ta, tha, va, wa, ya, za。发音录自男女各 1 名。元音取自 Hillenbrand^[41] 等(1995)的 12 个元音,即 had, hawed, hayed, head, heard, heed, hid, hod, hoed, hood, hud, who'd。发音录男女各 2 名。

采用上述的声码器技术,信号处理在 MATLAB 环境下完成。决定频域信息量的频道数为 1、2、3、4、6、8、12、16,决定时域信息量的低通滤波器的截止频率为 1、2、4、8、16、32、64、128、256、512 Hz。所有辅音和元音材料经过声码器处理产生的 80 种 (8 个频道数 \times 10 个低通截止频率)频域和时域信息量的组合。这 80 种组合则以随机顺序通过扬声器输出给听力正常受试者,声级选择受试者最舒适级。受试者通过计算机图形介面选择所听到的辅音或元音。

图 2 示这组受试者在不同频道数或不同低通截止频率时音素识别的均值。总的来说,音素识别率随着频道数的增加而提高。辅音识别率的提高主要在频道数从 1 增加至 6 或 8,而进一步增加频道数则未见辅音识别率的提高(图 2 左上)。元音识别率的提高见于频道数 1 至 8 或 12,而其识别率在 12 和 16 道之间无统计学差异 ($p>0.05$) (图 2 右上)。音素识别亦随声码器的低通截止频率的增高而提高。辅音识别率除在单通道情况下,一般均在 16 Hz 的低通截止频率处达到饱和,而在单通道时辅音识别率随着低通截止频率从 1 至 512 Hz 都在缓慢而持续地提高(图 2 左下)。元音识别率在单通道情况下不随低通截止频率的变化而变化,在多通道时,元音识别率大多在 4 Hz 时便达到饱和(图 2

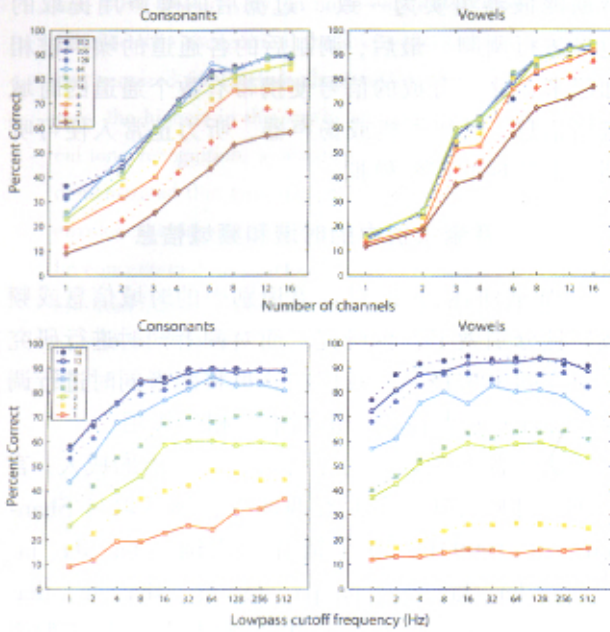


图2 人工耳蜗声学模拟下的音素识别。左、右两侧分别为辅音和元音识别的成绩。上排为音素识别与频道数的关系,不同曲线代表使用不同的低通滤波截止频率。下排为音素识别与低通滤波器截止频率的关系,不同曲线代表使用不同的频道数。

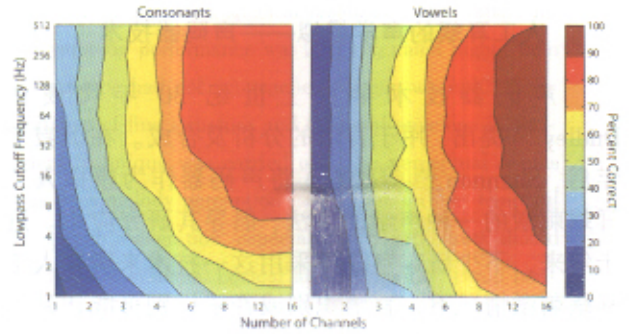


图3 音素识别成绩的轮廓图。左、右两侧分别为辅音和元音识别。纵、横轴分别为低通滤波截止频率和频道数。不同颜色代表不同的音素识别率,如右侧色标所标示。

存在这种频域和时域信息相互弥补的现象。

此项研究表明,频域信息和时域信息对音素识别都很重要,频域信息的作用尤为突出,在一定的频道数和低通截止频率范围内,时域和频域信息可相互弥补。为达到音素识别的最佳成绩,人们应考虑更多地传送时域和频域信息。

噪声下音素识别中的时域和频域信息

上面报道的是在安静状态时音素识别所需的时域和频域信息的研究。在噪声环境,言语识别是否需要更多的时域或频域信息?两者相互作用的模式是否与安静状态言语识别时有所不同?这便是下面将报道的研究旨在探讨的问题。实验设计与上面报道的研究相似,另7名英语为母语的受试者参加了本项实验,辅音及元音材料仍为上述的20个辅音和12个元音。进行声码器处理前,言语信号与一定量的言语谱噪声相混合以达到+6 dB和0 dB的信噪比。这些加有噪声的言语信号经过声码器处理,以模拟人工耳蜗患者在不同程度的噪声环境下言语识别的情形。声码器的时域信息仍由低通滤波器的截止频率1至512 Hz来决定,其频域信息以频道数2、4、6、8、12、16、24、32为定。其它测试步骤与上述的研究相似。

图4总结这7名受试者在安静状态及在+6 dB和0 dB信噪比的噪声下音素识别的平均成绩。尽管频道增加到32,在安静状态时,辅音及元音的识别率仍分别在8和8~12通道时达到饱和。在噪声情况下,音素识别率明显下降,辅音识别仍在8通道时达到饱和,频域信息的增加不能改善辅音识别。元音识别率在噪声情况下需要更多的频域信息才能达到饱和,在+6 dB信噪比时,达到饱和所需的频道

右下)。

从图2还可观察到,通道数(即频域信息)较低通截止频率(即时域信息)对音素识别起着更显著的作用。图2上排两小图中,不同低通截止频率产生的辅音和元音识别率变化范围平均为33.8和18.3百分点。而图2下排两小图中,不同频道数产生的辅音和元音识别率变化范围则平均为58.6和75.9百分点。这一频道数起决定作用的现象在图3中表达得更为清楚。图3为综合图2中的数据来表达时域信息(纵轴)与频域信息(横轴)的相互作用。图中,辅音(左)和元音(右)识别的百分率用彩色代替,并将同一水平的识别率以轮廓的方式围在一起,这种描绘方式也称为轮廓图(contour plot)。从图3可见,辅音在低通截止频率16 Hz以上或元音在低通截止频率4 Hz以上,其识别仅依赖于频道数。另一重要的观察为,当频道数为1至10或12而低通截止频率为1至16或32 Hz时,辅音识别在频域和时域信息两者之间可以获得相互弥补(trade-off)的作用,比方说,如果要达到50%至60%的辅音识别率,在3个通道时需要16 Hz以上的低通截止频率,而在8至12个通道时,达到同等的识别率只需1 Hz的低通截止频率。元音识别中,在频道数4及低通截止频率4 Hz的范围内也

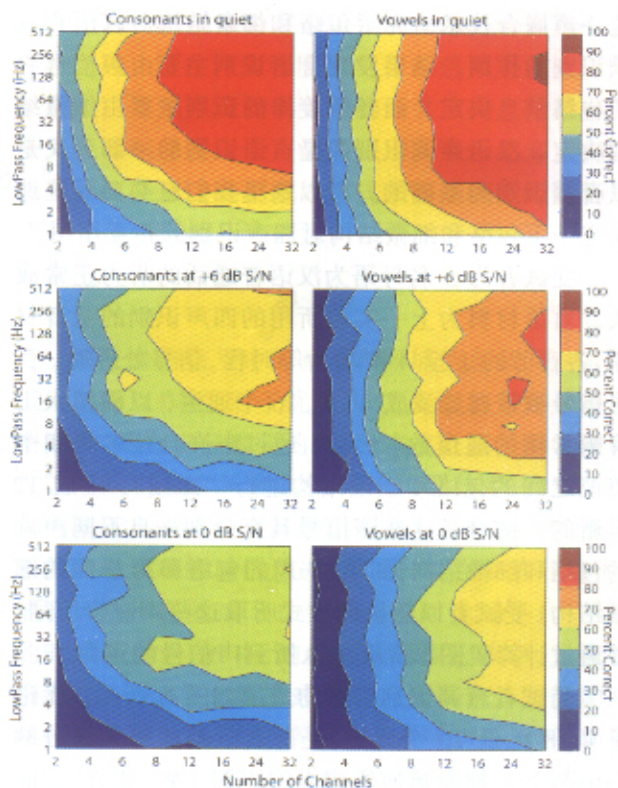


图 4 音素识别成绩的轮廓图。上、中、下三排分别为安静环境、+6 dB 和 0 dB 信噪比环境下音素识别的成绩。除注意频道数与图 3 中有不同外,其余注释见图 3。

数为 12 左右,而在噪声量更大的情况下(即在 0 dB 信噪比时)达到饱和的频道数为 16 左右。从时域信息的角度来看,无论是安静状态还是不同信噪比的噪声环境,时域信息对音素识别的贡献基本上无差异,也就是说,辅音识别需要 16 Hz 的时域信息,而元音识别只需要 4 Hz 的时域信息,在噪声环境下,受试者无法利用更多的时域信息来帮助言语识别。以往几项研究表明,在噪声环境下言语识别在频道更多的情况下成绩会更好^[28,30],从本项研究可见,这一改善可能是由元音识别的提高而产生的。

汉语四声识别中的时域和频域信息

汉语包括其多种方言属于声调语言 (tone language) 的一种,而汉语普通话为所有语种中使用人数最多的语言。以往对汉语四声识别的声学依据的研究不多,较经典的见于梁之安^[42] (1963) 的一项报道。在该研究中,他将汉语言语信号通过 2.4 kHz 的高通滤波,以除去信号中的基频 (F_0) 信息,结果听力正常受试者四声识别不受影响。他的另一项实验用耳语做为声刺激,结果发现受试者四声识别仍有 60 ~ 70% 的准确率,耳语信号中既无 F_0 也无谐音成

份,但时域包络信息尚存,故四声识别可能利用言语信息中的时域包络信息来得到补偿。Whalen 和 Xu^[43] (1992) 证实言语信号中的振幅轮廓 (amplitude contour) (即时域包络) 含有声调的信息。

从理论上讲,当代人工耳蜗的言语处理器并未考虑声调的问题, F_0 不能有效地体现于人工耳蜗的电刺激模式中^[44]。但是,人工耳蜗可以传递数个频道的信息,而每个频道内含有的时域包络信息也可得到较好的体现。在这种状态下,声调识别究竟能达到什么水平? 言语信号中频域和时域信息对声调识别的贡献是什么样的关系?

本研究实验设计与上面报道的研究相似,4 名母语为汉语普通话的听力正常成人参加了本项实验。言语材料为 10 个音节的 1~4 声,如 m (妈)、má (麻)、m (马)、mà (骂)。这 40 个字分别用 1 男声和 1 女声录制。在汉语普通话中,各声调的时程通常不等,如第 3 声通常最长而第 4 声通常最短^[18,43]。为避免时程对声调识别的影响,我们的言语材料中只选用录制时 1~4 声为等时程的样本。这些四声识别的言语信号经过上述的声码器处理,以模拟人工耳蜗刺激的情形。时域信息仍由低通滤波器的截止频率 1~512 Hz 决定,频域信息由通道数 1、2、3、4、6、8、10、12 来决定。声刺激由扬声器传给受试者,受试者通过计算机图形界面以四选一的方式选择听到的声调。随机成绩为 25% (即 1/4)。

四名受试者声调识别的平均成绩由图 5 所示。比较图 5 与图 3 中的轮廓,我们可见频域和时域信息在四声识别中与在辅音和元音识别中有明显不同的作用。首先,四声识别的成绩普遍较低,最佳成绩在 12 个通道及低通截止频率 256 Hz 时才达到 75% 的准确率。其二,声调识别率依赖于频道数,当频道数为 12 时,尚未见声调识别率达到饱和的倾向。其三,声调识别对时域信息的需求远远多于辅音和元音对之的要求,声调识别受益于高达 256 Hz 的低通截止频率的时域信息。最后,时域和频域信息在相当广的范围内有相互弥补的作用,这一范围是 1 至 12 个频道与 1 至 256 Hz 的低通截止频率。

对人工耳蜗患者声调识别的报道均指出目前人工耳蜗系统不能充分提供声调信息,患者声调识别率由随机水平至 80% 准确率不等^[45-52]。我们的一项初步研究还发现人工耳蜗患儿四声的发声不准确^[53]。Liu^[54] 等 (2004) 发现人工耳蜗电极数目从 20 降到 6 时,声调识别率有下降。Wei^[55] 等 (2004) 的声

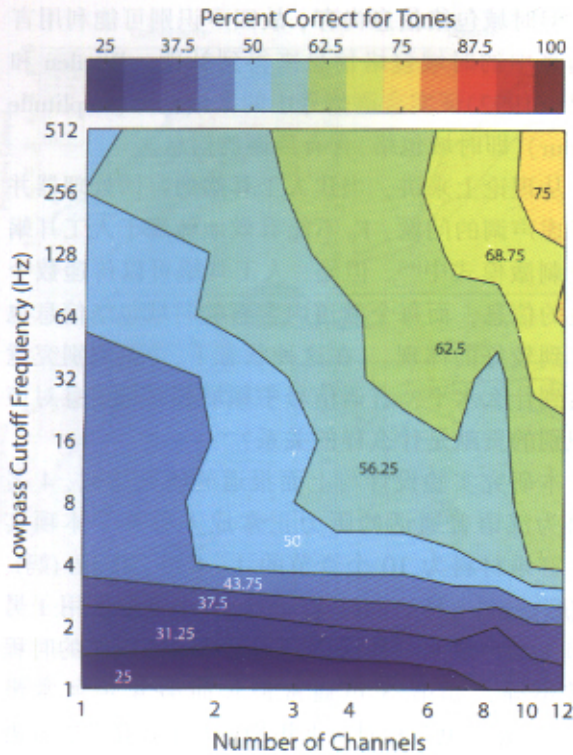


图5 汉语四声识别成绩的轮廓图。不同颜色代表不同的声识别率,如轮廓图上方的色标所示。其余注释见图3。

调识别实验中使用 1、4、7、10、14 和 20 个电极,结果发现人工耳蜗患者的四声识别率由电极数 1 增加到 7 时逐步增加,而使用 7 至 20 个电极时,四声识别率处于 70% 的平台水平。他们还观察到,患者对电脉冲速率的分辨能力(实际上是一种时域信息的接受能力)与使用 20 个电极时声调识别率相关。因此,我们的结论,即频域和时域信息对声调识别都很重要,在人工耳蜗患者得到证实。

时域包络和细微结构在声调识别中的作用

言语信号的波形从数学上可分解为缓慢变化的包络和快速变化的细微结构。在此我们对这两者在声调识别中的作用进行探讨。信号处理的方法与上述的声码器技术截然不同,它是一种所谓的“声嵌合”(auditory chimera)技术。这一技术的基本要点是采用 Hilbert 转换对两个声信号(A和B)的包络和细微结构分别进行提取,然后将A的包络与B的细微结构进行“嫁接”,形成一个新的声信号,也可将A的细微结构与B的包络进行“嫁接”,形成另一个新的声信号。受试者听取这两种新合成的声信号以判断听到的是A或是B,研究者从而判断是包络还是细微结构决定着受试者的识别。Smith^[6]等(2002)首先用

这一声嵌合技术研究了包络和细微结构对言语和音乐识别的作用。结果发现言语识别主要由声信号中的包络信息决定,而音乐旋律的识别主要由细微结构决定。汉语声调识别即是言语识别的一部分又是以音调识别为基础的,所以这里我们有必要系统地研究一下包络和细微结构对四声识别的相互作用。

受试者为 5 名母语为汉语普通话的听力正常成人。言语材料为上一实验所用的四声识别的录音材料,各音节的 1 至 4 声均为等时程。信号处理时,言语信号被带通滤波成 4、8、16 个通道,以模拟人工耳蜗常用的通道数,然后,通过声嵌合技术将四个声调之间的包络和细微结构进行“嫁接”,生成 12 组新的声信号,这些声信号具有分别来自不同声调的包络和细微结构,如第 1 声的包络和第 2 声的细微结构。受试者以随机的方式听取这些声嵌合信号,并通过计算机图形介面输入听到声信号的声调。

受试者声调识别的平均成绩如图 6 所示。无论是 4、8 还是 16 个通道,受试者听到的绝大部分的声调都是与细微结构所代表的声调一致,比方说,如果声刺激含有第 1 声的包络和第 2 声的细微结构,大部分时间受试者听到的是第 2 声,而只有很少部分的时间听到的是第 1 声,当然还有极少部分的时间受试者听到即非第 1 声又非第 2 声。因此,我们声调识别的结果与 Smith^[6]等(2002)的音乐旋律识别的结果相似,而与英语言语识别的结果相悖。说

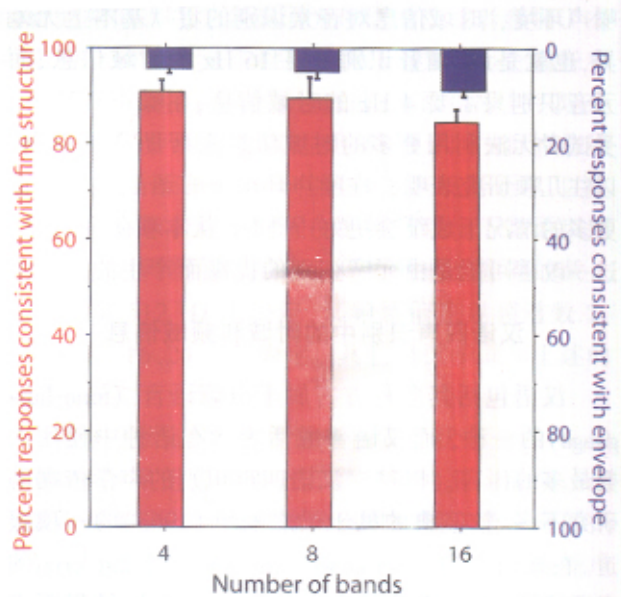


图6 声嵌合实验的汉语四声识别的平均成绩及其标准差。横轴代表 4、8、16 个频率通道。当四声的判断与细微结构一致时,其百分率用红色柱表示并以左侧纵轴为准。当四声的判断与时域包络一致时,其百分率用蓝色柱表示并以右侧纵轴为准。

明汉语四声识别虽是言语识别的一部分,但具有其独特之处,它在许多程度上与音乐感知更接近。如果在人工耳蜗系统提供更多的细微结构,这将有助于人工耳蜗患者汉语声调的识别及音乐感知。当然,患者如何通过电刺激的方式来接受细微结构仍是研究人员面临的一大难题。

总 结

研究言语信号中的时域和频域信息及其对言语识别的相互作用,对于优化人工耳蜗的言语处理方案来说,具有十分重要的意义。本文介绍了用于模拟人工耳蜗刺激的声学技术,即声码器技术。在这一技术的基础上,我们对言语识别在安静和噪声环境下所需的时域和频域信息进行了研究。在安静环境下,辅音识别率在 6 或 8 个频率通道时达到平台,所需的时域信息为 16 Hz 的包络低通截止频率,元音识别率在 8 到 12 个频率通道时达到平台,所需的时域信息只有 4 Hz 的包络频率。在噪声环境下,辅音识别率不因频道数多于 8 或时域信息大于 16 Hz 而有提高,元音识别率对时域信息的依赖仍为 4 Hz,但频域信息的增多有助于元音识别的提高。汉语普通话四声识别在人工耳蜗声学模拟的实验中显得明显不足,而且当频道数达 12 时尚未见平台平级,所需的时域信息明显高于辅音和元音的识别,包络低通截止频率高达 256 Hz 时仍对声调识别有益。无论是对辅音和元音识别还是对声调识别,时域信息和频域信息都有一种相互弥补的关系。我们用声嵌合技术的研究表明,声调识别与音乐感知相似,主要由声信号中细微结构而决定。我们正将这些声码器和声嵌合技术的研究扩展到感言神经性聋患者,以探讨言语信号中时域和频域信息在耳蜗病理情况下对言语识别和声调识别的相互作用。

志谢 研究得到美国 NIH 和俄亥俄大学研究基金的支持。

参 考 文 献

- 徐立. 人工耳蜗的工作原理 [M]// 韩德民. 人工耳蜗. 北京: 人民卫生出版社, 2003: 7- 38.
- Friessen LM, Shannon RV, Baskent D, et al. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am*, 2001, 110: 1150- 1163.
- Shannon RV. Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. *Hear Res*, 1983, 11: 157- 189.
- Shannon RV. Temporal modulation transfer functions in patients with cochlear implants. *J Acoust Soc Am*, 1992, 91: 2156- 2164.
- van den Honert C, Stypulkowski PH. Physiological properties of the electrically stimulated auditory nerve. II. Single fiber recordings. *Hear Res*, 1984, 14(3): 225- 243.
- Rubinstein JT, Hong R. Signal coding in cochlear implants: Exploiting stochastic effects of electrical stimulation. *Ann Otol Rhinol Laryngol*, 2003, 112: 14- 19.
- Skinner MW, Arndt PL, Staller SJ. Nucleus 24 advanced encoder conversion study: Performance vs preference. *Ear Hear*, 2002, 23: 2S- 25S.
- Villchur E. Electronic models to simulate the effect of sensory distortions on speech perception by the deaf. *J Acoust Soc Am*, 1977, 62: 665- 674.
- ter Keurs M, Festen JM, Plomp R. Effect of spectral envelope smearing on speech reception. I. *J Acoust Soc Am*, 1992, 91: 2872- 2880.
- ter Keurs M, Festen JM, Plomp R. Effect of spectral envelope smearing on speech reception. II. *J Acoust Soc Am*, 1993, 93: 1547- 1552.
- Baer T, Moore BCJ. Effects of spectral smearing on the intelligibility of sentences in noise. *J Acoust Soc Am*, 1993, 94: 1229- 1241.
- Baer T, Moore BCJ. Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. *J Acoust Soc Am*, 1994, 95: 2277- 2280.
- Boothroyd A, Mulhearn B, Gong J, et al. Effects of spectral smearing on phoneme and word recognition. *J Acoust Soc Am*, 1996, 100: 1807- 1818.
- Hill FJ, McRae LP, McClellan RP. Speech recognition as a function of channel capacity in a discrete set of channels. *J Acoust Soc Am*, 1968, 44: 13- 18.
- Shannon RV, Zeng FG, Kamath V, et al. Speech recognition with primarily temporal cues. *Science*, 1995, 270: 303- 304.
- Dorman MF, Loizou PC, Rainey D. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise outputs. *J Acoust Soc Am*, 1997, 102: 2403- 2411.
- Loizou PC, Dorman M, Tu Z. On the number of channels needed to understand speech. *J Acoust Soc Am*, 1999, 106: 2097- 2103.
- Xu L, Tsai Y, Pflugst BE. Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *J Acoust Soc Am*, 2002, 112: 247- 258.
- Xu L, Thompson C, Pflugst BE. Relative contributions of spectral and temporal cues for phoneme recognition. *J Acoust Soc Am*, 2005, 117: 3255- 3267.
- Eisenberg LS, Shannon RV, Martinez AS, et al. Speech recognition with reduced spectral cues as a function of age. *J Acoust Soc Am*, 2000, 107: 2704- 2710.
- Rosen S. Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philos Trans R Soc London SerB* 1992, 336: 367- 373.
- Drullman R, Festen JM, Plomp R. Effect of temporal envelope

- smearing on speech perception. *J Acoust Soc Am*, 1994, 95: 1053-1064.
- 23 Drullman R, Festen JM, Plomp R. Effect of reducing slow temporal modulations on speech reception. *J Acoust Soc Am*, 1994, 95: 2670-2680.
- 24 Fu Q-J, Shannon RV. Effect of stimulation rate on phoneme recognition by Nucleus-22 cochlear implant listeners. *J Acoust Soc Am*, 2000, 107: 589-597.
- 25 Xu L, Pflingst BE. Relative importance of the temporal envelope and fine structure in tone perception. *J Acoust Soc Am*, 2003, 114: 3024-3027.
- 26 Xu L, Zheng Y. Spectral and temporal cues for phoneme recognition in noise. American Auditory Society Annual Meeting, Scottsdale, AZ, 2006.
- 27 Dudley H. Remaking speech. *J Acoust Soc Am*, 1939, 11:169-177.
- 28 Dorman MF, Loizou PC, Fitzke J, et al. The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels. *J Acoust Soc Am*, 1998, 104: 3583-3585.
- 29 Dorman MF, Loizou PC, Fitzke J, et al. Recognition of monosyllabic words by cochlear implant patients and by normal-hearing subjects listening to words processed through cochlear implant signal processing strategies. *Ann. Otol. Rhinol. Laryngol.* 2000, Suppl 109: 64-66.
- 30 Fu Q-J, Shannon RV, Wang X. Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. *J Acoust Soc Am*, 1998, 104: 3586-3596.
- 31 Fu Q-J, Shannon RV. Effects of electrode location and spacing on phoneme recognition with the Nucleus-22 cochlear implant. *Ear Hear*, 1999, 20: 321-331.
- 32 Loizou PC, Dorman M, Poroy O, et al. Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution. *J Acoust Soc Am*, 2000, 108: 2377-2387.
- 33 Henry BA, Turner CW. The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *J Acoust Soc Am*, 2003, 113: 2861-2873.
- 34 Turner CW, Gantz BJ, Vidal C, et al. Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing. *J Acoust Soc Am*, 2004, 115: 1729-1735.
- 35 Kong Y-Y, Cruz R, Jones JA, et al. Music perception with temporal cues in acoustic and electric hearing. *Ear Hear*, 2004, 25: 173-185.
- 36 Shannon RV, Zeng F-G, Wygonski J. Speech recognition with altered spectral distribution of envelope cues. *J Acoust Soc Am*, 1998, 104: 2467-2476.
- 37 Dorman MF, Loizou PC, Rainey D. Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *J Acoust Soc Am*, 1997b, 102: 2993-2996.
- 38 Rosen S, Faulkner A, Wilkinson L. Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *J Acoust Soc Am*, 1999, 106: 3629-3636.
- 39 Greenwood DD. A cochlear frequency-position function for several species-29 years later. *J Acoust Soc Am*, 1990, 87: 2592-2605.
- 40 Shannon RV, Jansvold A, Padilla M, et al. Consonant recordings for speech testing. *J Acoust Soc Am*, 1999, 106: L71-74.
- 41 Hillenbrand J, Getty LA, Clark MJ, et al. Acoustic characteristics of American English vowels. *J Acoust Soc Am*, 1995, 97: 3099-3111.
- 42 梁之安. 汉语普通话中声调的听觉辨认依据. *生理学报*, 1963, 26(2): 85-91.
- 43 Whalen DH, Xu Y. Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 1992, 49: 25-47.
- 44 Moore BCJ. Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants. *Otol Neurotol*, 2003, 24: 243-54.
- 45 Zeng FG. Cochlear implants in China. *Audiology*, 1995, 34: 61-75.
- 46 Huang TS, Wang NM, Liu SY. Tone perception of Mandarin-speaking postlingually deaf implantees using the nucleus 22-channel cochlear mini system. *Ann Otol Rhinol Laryngol*, 1995, Suppl 166: 294-298.
- 47 Huang TS, Wang NM, Liu SY. Nucleus 22-channel cochlear mini-system implantations in Mandarin-speaking patients. *Am J Otol*, 1996, 17: 46-52.
- 48 Sun JC, Skinner MW, Liu SY, et al. Optimization of speech processor fitting strategies for Chinese-speaking cochlear implantees. *Laryngoscope*, 1998, 108: 560-568.
- 49 Wei WJ, Wong R, Hui Y, et al. Chinese tonal language rehabilitation following cochlear implantation in children. *Acta Otolaryngol*, 2000, 120: 218-221.
- 50 Lee KYS, van Hasselt CA, Chiu SN, et al. Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children. *Int J Ped Otorhinolaryngol*, 2002, 63: 137-47.
- 51 Ciocca V, Francis AL, Aisha R, et al. The perception of Cantonese lexical tones by early-deafened cochlear implantees. *J Acoust Soc Am*, 2002, 111: 2250-2256.
- 52 Wu JL, Lin CY, Yang HM, et al. Effect of age at cochlear implantation on open-set word recognition in Mandarinspeaking deaf children. *Int J Ped Otorhinolaryngol*, 2006, 70: 207-211.
- 53 Xu L, Li Y, Hao J, et al. Tone production in Mandarin-speaking children with cochlear implants: A preliminary study. *Acta Otolaryngol*, 2004, 124:363-367.
- 54 Liu T-C, Chen HP, Lin HC. Effects of limiting the number of active electrodes on Mandarin tone perception in young children using cochlear implants. *Acta Otolaryngol*, 2004, 124:1149-1154.
- 55 Wei CG, Cao KL, Zeng FG. Mandarin tone recognition in cochlear-implant subjects. *Hear Res*, 2004, 197: 87-95.
- 56 Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 2002, 416: 87-90.

(收稿日期 2006-9-20)