

## 基于信号能量选取的电子耳蜗言语处理策略特征

陈又圣, 曹雪梅, 王庚

深圳信息职业技术学院信息与通信学院, 广东 深圳 518000

**【摘要】**基于信号能量选择的电子耳蜗言语处理策略的算法架构,分析该策略的信号处理过程,并比较波形、信号强度、语音频谱等特征,以获得该策略对语音信号处理的影响。此外,本文对信号序列每帧中选择的最大子带数量以及电子耳蜗滤波器组选择每个子带的概率进行深入统计。研究发现,基于信号能量选择的言语处理策略的特点是低频通道的选取概率较高,高频通道的选取概率较低。通过重点分析算法在信号特征参数传输上的匹配性和频带选择中的关键特征,为提出和改进新的电子耳蜗言语处理策略提供依据。

**【关键词】**电子耳蜗;频带选取;信号能量;言语处理策略

**【中图分类号】**R318;TH776

**【文献标志码】**A

**【文章编号】**1005-202X(2023)06-0736-06

## Characteristics of cochlear implant speech processing strategies based on signal energy selection

CHEN Yousheng, CAO Xuemei, WANG Geng

School of Information and Communication, Shenzhen Institute of Information Technology, Shenzhen 518000, China

**Abstract:** The study explores the algorithm architecture of cochlear implant speech processing strategies based on signal energy selection, analyzes the signal processing process, and compares several features such as waveform, signal intensity and speech spectrum for discussing the role of the strategy in speech signal processing. In addition, an in-depth statistical analysis is carried out on the maximum number of subbands selected in each frame of the signal sequence and the probability of each subband selected using the artificial cochlear filter bank. It's found that the speech processing strategies based on signal energy selection have the characteristics of high probability in low-frequency channels and low probability in high-frequency channels. The analysis on the matching of algorithms in signal feature parameter transmission and key features in frequency band selection provides a basis for putting forward and improving new speech processing strategies for cochlear implants.

**Keywords:** cochlear implant; frequency band selection; signal energy; speech processing strategy

### 前言

佩戴电子耳蜗是目前失聪患者恢复听力感知的唯一途径,是改善他们生活的装置<sup>[1-2]</sup>。电子耳蜗包括体内植入装置和体外机装置<sup>[3]</sup>,其中,外部语音处理器是关键部件之一。预先开发的言语处理策略嵌入到语音处理器中,以提取声音信号的关键特征参数,并将其传输到体内的电极进行电刺激。目前,电

子耳蜗在安静环境中的言语识别率已经很高,但在各种噪声环境和失配环境中的声音识别率仍然很低<sup>[4-7]</sup>。为了提升言语识别率,目前的方法包括:麦克风阵列前端语音增强<sup>[8-10]</sup>、双耳语音增强<sup>[11-13]</sup>、信号处理算法和策略的改进<sup>[14-17]</sup>、电场聚焦<sup>[18]</sup>、光学刺激耳蜗模式等<sup>[19-20]</sup>。言语处理策略是提高语音识别率的关键技术,也是当前发展的重点之一。随着技术的发展,电子耳蜗已经从早期的单通道言语处理策略改进为当前的多通道多电极言语处理策略。不同的电子耳蜗制造商在其产品中使用不同的言语处理策略。从算法架构本身的角度来看,当前主流的言语处理策略可以归结为基于信号能量选取的电子耳蜗言语处理策略。研究这类策略的信号处理流程和参数提取的特点,将有助于提出更有效的言语处理算法,在有限的电极条件下传输更有效的目标语音信息,并为提高电子耳蜗的言语识别率提供可能。

**【收稿日期】**2023-01-26

**【基金项目】**广东省普通高校特色创新类项目(2019GKTSCX094);深圳信息职业技术学院项目(SZIIT2020SK002);广东省普通高校青年创新人才项目(2020KQNCX206);深圳市科技计划项目(JCYJ20190808093001772)

**【作者简介】**陈又圣,博士,讲师,主要研究方向:语音增强、麦克风阵列波束形成、电子耳蜗言语处理、医疗仪器设计, E-mail: chenyoushengtsinghua@aliyun.com

1 电子耳蜗言语处理策略的分类

电子耳蜗言语处理策略的特点是使用有限数量的电极来表示复杂而丰富的声音信号。由于电极的

数量有限,原始信号的特征不能被完全提取和传输,并且信号只能以某种方式被提取和编码。根据言语处理策略提取的信号特征参数,可以根据图 1 对言语处理策略进行分类。

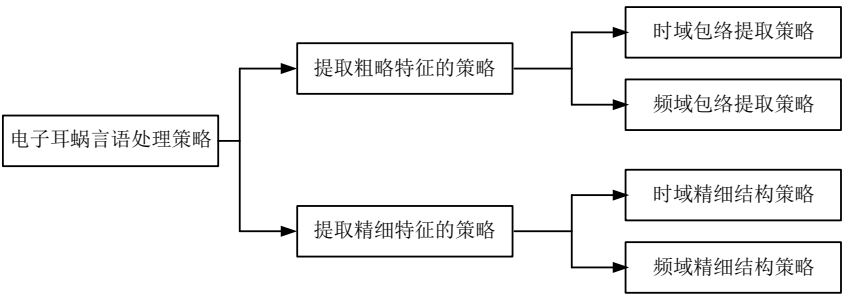


图 1 电子耳蜗言语处理策略的分类  
Figure 1 Classification of speech processing strategies for cochlear implants

从信号特征提取的精细程度来看,电子耳蜗言语处理策略包括提取粗略特征的策略和提取精细特征的策略。提取粗略特征的策略分为时域包络提取策略和频域包络提取策略两种,提取精细特征的策略包括时域精细结构策略和频域精细结构策略两种。在常规的言语处理策略中,连续交替取样(Continuous Interleaved Sampling, CIS)<sup>[21-22]</sup>、配对脉冲采样(Paired Pulsatile Sampler, PPS)、多脉冲采样(Multiple Pulsatile Sampler, MPS)、谱峰提取(Spectral PEAK, SPEAK)<sup>[23]</sup>、高级组合编码(Advanced Combined Encoder, ACE)<sup>[24]</sup>、n of m<sup>[25-26]</sup>等属于时域包络提取策略,F0/F2<sup>[27-28]</sup>、多峰(Multi-PEAK, MPEAK)等属于频域包络提取策略,频率幅度调制编码(Frequency Amplitude Modulation Encoding, FAME)、精细结构编码(Fine Structure Encoding Processing, FSP)<sup>[29]</sup>等属于时域精细结构策略,虚拟电极、高分辨率策略(HiRes Fidelity)<sup>[30]</sup>等属于频域精细结构策略。每个子类别的言语处理策略可以进一步细分。例如,在时域包络提取策略中,CIS、PPS、MPS等策略根据指定的规则传递频带参数,而SPEAK、ACE、n of m等策略基于信号的最大能量确定频带选择。在众多的电子耳蜗言语处理策略中,基于信号能量选择的SPEAK、ACE、n of m被广泛使用,是本文的重点。在SPEAK策略中,带通滤波器的数量通常为16~20个,并为相应的刺激电极选择6个或更多具有最高能量的频带。ACE策略与之类似,每个通道的刺激率更高,可以保留目标信号的更多时间细节。n of m策略是通用的,具有更多的带通滤波器,并且为电极刺激选择的频带数量也是可调整的。

2 基于信号能量选取的言语处理策略

基于信号能量选择的言语处理策略的核心是,每次用于电极刺激的频带不是预先确定的,而是根据当时每个频带的信号能量来确定的。信号处理流程如图 2 所示。

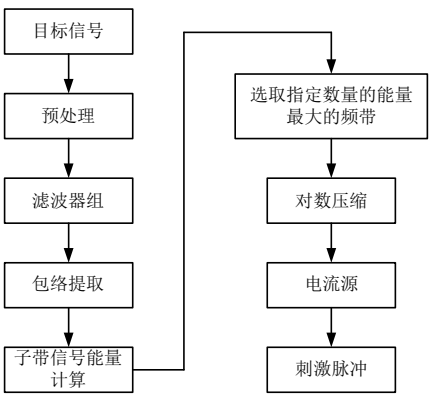


图 2 基于信号能量选取的言语处理策略实现流程  
Figure 2 Implementation process of speech processing strategies based on signal energy selection

对目标信号进行预处理,通过预加重来平衡信号的高频分量和低频分量,然后通过滤波器组将信号分成从低到高的频带。频带的数量与刺激电极的数量一致。通过半波整流、全波整流、希尔伯特变换等方法提取信号的时间包络。最后,基于每个子带的包络计算相应的子带信号能量,并选择具有最大能量的几个子带信息。对数压缩过程用于适应内部电动态范围。诸如刺激持续时间、刺激速率和刺激幅度等参数最终被传输到相应的电极,并且通过电流源产生所需的电刺激脉冲。

### 3 言语处理的信号特征及参数分析

在基于信号能量选取的言语处理策略中,不同的算法中频带的总数和所选频带的数量是不同的。在Nucleus系列的电子耳蜗中,总频带数通常为22和24。SPEAK策略选择的频带数通常为6~8,ACE策

略选择的频带数通常为8~12。以实际采集的中文单词“电子耳蜗”(女声)为例,采用了8 of 22策略(22个频道选择最大8个频带)。基于信号能量选取的言语处理策略在算法处理前后的波形比较如图3所示。

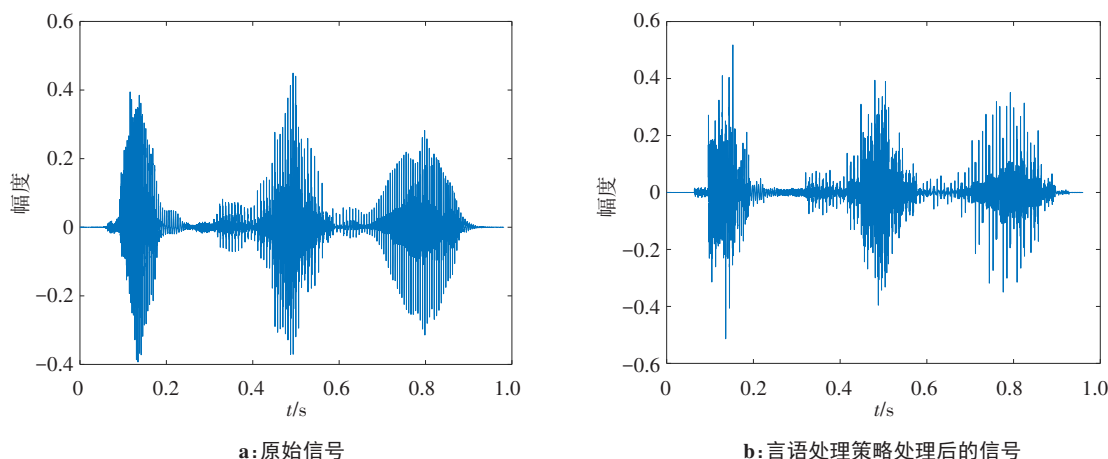


图3 电子耳蜗言语处理前后的波形对比

Figure 3 Comparison of waveforms before and after cochlear implant speech processing

图3a显示了原始信号的波形,包含丰富而连续的声音信息。图3b为通过22通道电子耳蜗使用能量选择算法获得的信号波形。22通道滤波器组的频带分别是:[152, 274]、[274, 396]、[396, 517]、[517, 639]、[639, 761]、[761, 883]、[883, 1005]、[1005, 1127]、[1127, 1310]、[1310, 1492]、[1492, 1675]、[1675, 1919]、[1919, 2224]、[2224, 2589]、[2589, 2955]、[2955, 3382]、[3382, 3869]、[3869, 4478]、[4478, 5149]、[5149, 5880]、[5880, 6794]、[6794, 7769] Hz。为了便于后面的分析,对这些子带分别编号为通道1~22。

由于只选择了部分频带,由言语处理策略处理的信号总能量将减少,因此有必要为处理后的信号提供增益以平衡信号能量。增益系数 $G$ 根据式(1)计算:

$$G = \frac{\sqrt{\sum |x(n)|^2}}{\sqrt{\sum |\hat{x}(n)|^2}} \quad (1)$$

其中, $x(n)$ 是原始信号序列, $\hat{x}(n)$ 是言语处理策略处理后的信号序列。

在执行信号能量平衡处理之后,信号处理前后的总能量保持恒定。从图3中的波形来看,语音的发音位置和停顿位置是一致的,但波形细节不同。电子耳蜗言语处理前后的信号强度对比如图4所示。

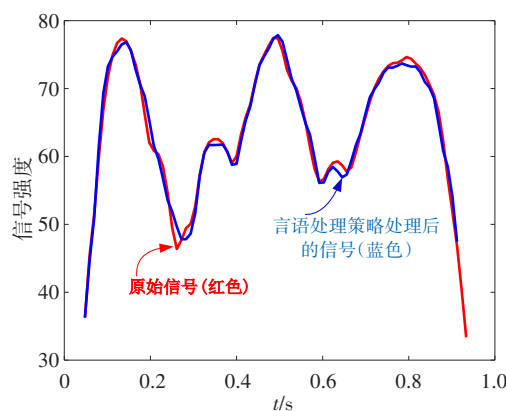


图4 电子耳蜗言语处理前后的信号强度对比

Figure 4 Comparison of signal intensity before and after cochlear implant speech processing

通过比较,可以看到信号处理前后的信号强度曲线高度一致,这表明基于信号能量选择的电子耳蜗言语处理策略具有保持整体能量变化的良好特征。

基于信号能量选取的言语处理策略筛选具有最大能量的频带并将其用于电极刺激。信号处理过程可以从频谱能量变化过程中实时看到,如图5所示。图5中的频谱图里每一帧对应512个采样点。当电子耳蜗使用有限的频带来传输相应的刺激电极时,原始信号(图5a)的能量被重新分配并集中在相应的22通道耳蜗滤波器组(图5b)每个子频带的中心频率



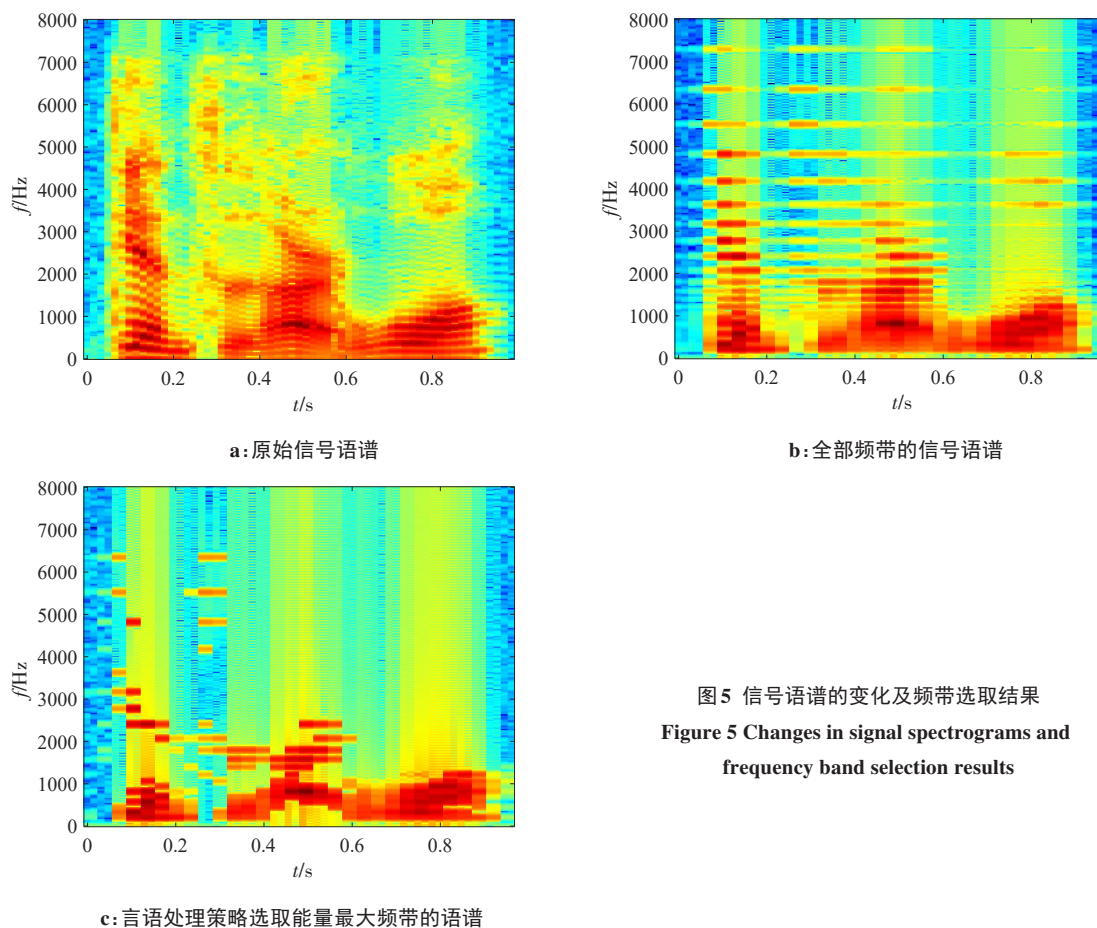


图5 信号语谱的变化及频带选取结果  
Figure 5 Changes in signal spectrograms and frequency band selection results

附近。图5中的颜色表示信号能量的强度。红色是信号能量集中的区域,蓝色是信号能量较弱的区域。在图5b中,每个子带在每个时间段都具有较高强度的相应信号能量,但在不同子带中的大小不同。因此,在图5c的每个时间段中,只有8个颜色最深的子带(深红色)被算法选择并传送到相应的刺激电极。

滤波器组的子带按照频率由低到高的顺序编号为1~22,可以分析和计数每个时间帧中选择的子带(在算法中使用Channel\_Num表示选择的频带数),结果如表1所示。在表1中,每一行有8个数字,代表所选取的按子带能量由大到小排序的子带编号。可以看到,小数值编号被选取的情形较多(小数值的编号对应电子耳蜗滤波器组的低频子带)。可以进一步精细地统计每一帧的频带编号的频数,获得不同子带被选取的概率。其中,子带被选取的概率 $P(i)$ 按式(2)进行计算:

$$P(i) = \frac{M_i}{N} \times 100\% \quad (2)$$

通过统计得到的各子带被选取的概率曲线如图6所示。从图6的子带被选取的概率曲线来看,第1~6通道被选取的概率较高,对应的是低频频带,第15~22通道被选取的概率较低,对应的是高频频带。进

一步改变字词来进行实验测试,也有类似的结果,即基于信号能量选取的言语处理策略将以较高的概率选取低频通道,较低的概率选取高频通道。

为了分析所选取的部分子带能力是否可以表征目标信号的信息,可进一步计算在每个时间帧里所选的部分子带的能量占全部子带能量的百分比。用 $j$ 表示帧编号,用 $C_j(n)$ 和 $A_j(n)$ 分别表示在第 $j$ 帧中选取部分子带和全部子带所对应的时域信号,按照512点的帧长, $n$ 的范围是1~512。则每一帧所选取的部分子带的能量占全部子带的能量的比值 $\eta(j)$ 按式(3)进行计算:

$$\eta(j) = \frac{\sqrt{\sum_{n=1}^{512} (C_j(n))^2}}{\sqrt{\sum_{n=1}^{512} (A_j(n))^2}} \times 100\% \quad (3)$$

通过计算得到每个时间帧的能量占比如表2所示。从表2可以看出,除了开始时几个静音帧的能量百分比约为80%之外,其他帧的能量比例大于90%。因此,尽管在电子耳蜗言语处理策略中仅选择了22个通道中的8个子带通道,但信号能量的匹配度非常

表 1 基于信号能量由大到小排序的 8 个被选取子带

Table 1 Eight selected subbands sorted from large to small based on signal energy by time frame

帧序号	频带能量排序							
	1	2	3	4	5	6	7	8
1	1	2	20	21	6	16	9	11
2	20	21	12	14	19	16	11	18
3	2	1	20	3	16	21	15	17
4	2	15	1	14	6	4	19	16
5	4	14	6	3	1	2	5	8
6	5	3	1	13	4	14	7	2
7	1	2	3	4	13	6	12	5
8	1	2	3	4	13	10	6	20
9	20	19	21	13	12	18	14	9
10	2	12	20	13	21	1	19	8
11	2	3	1	11	12	4	10	5
12	3	2	4	1	12	11	5	10
13	5	6	3	2	1	4	12	11
14	5	6	7	11	4	3	10	8
15	6	11	10	7	5	9	12	8
16	6	11	5	12	7	10	4	14
17	6	5	12	4	11	14	7	13
18	5	4	3	6	12	14	11	13
19	2	1	4	3	5	13	7	8
20	2	1	3	4	5	6	7	8
21	2	1	3	5	4	6	7	8
22	2	1	4	3	5	6	7	8
23	2	4	3	6	5	1	7	9
24	2	4	6	5	3	1	7	9
25	7	4	5	2	6	3	1	8
26	7	5	4	8	6	9	2	1
27	6	5	7	9	4	8	2	1
28	6	1	8	9	4	5	7	3
29	1	6	8	7	2	9	3	5
30	8	6	9	1	3	7	10	5

高,从信号能量的角度来看,可以最大限度地传输有效信息。

4 结 论

本文研究了目前广泛使用的基于信号能量选取的电子耳蜗言语处理策略的信号处理流程和信号特性。仿真实验表明,基于信号能量选取的算法具有较高的信号强度匹配度,并且低频通道是子频带选择中的主要通道,所选的部分频带的能量达到总频

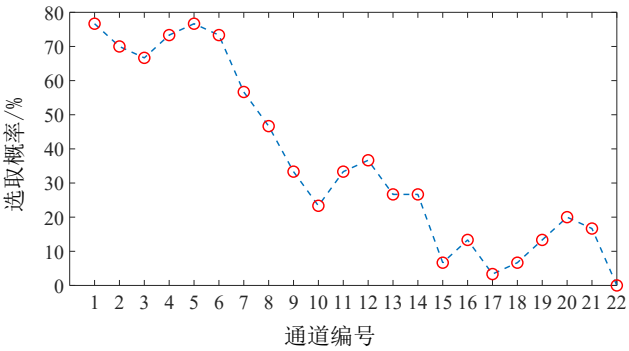


图 6 电子耳蜗滤波器组里各子带被选取的概率曲线

Figure 6 Probability curve of each subband selected using the cochlear filter bank

表 2 所选取的部分子带信号占全部信号能量的百分比

Table 2 Percentage of the energy of the selected subband signals to the total signal energy

帧序号	所选取的子带能量占全部子带能量的百分比/%
1	80.648 0
2	83.426 0
3	87.250 3
4	97.647 4
5	98.464 8
6	99.128 3
7	99.894 3
8	99.636 2
9	94.565 3
10	93.144 9
11	98.941 3
12	99.003 7
13	96.500 8
14	95.979 8
15	98.502 2
16	98.773 1
17	98.223 5
18	97.749 5
19	97.414 5
20	99.971 7
21	99.970 3
22	99.981 9
23	99.982 8
24	99.970 6
25	99.914 8
26	99.664 4
27	99.654 5
28	99.126 4
29	99.902 8
30	98.368 6

带信号能量的90%以上,可以更好地保持原信号的能量信息。通过本文的研究,可以获得信号频谱和信号强度的变化和匹配,并获得诸如电子耳蜗频带通道的选择和主要信号能量的提取百分比等关键参数,这有助于言语处理策略的进一步发展和改进。

## 【参考文献】

- [1] O'Donoghue G. Cochlear implants-science, serendipity, and success [J]. *N Engl J Med*, 2013, 369(13): 1190-1193.
- [2] Zeng FG. Trends in cochlear implant[J]. *Trends Amplif*, 2004, 8(1): 1-34.
- [3] Dorman M, Wilson B. The design and function of cochlear implants [J]. *Am Sci*, 2004, 92(5): 436-445.
- [4] Hast A, Schlücker L, Digeser F, et al. Speech perception of elderly cochlear implant users under different noise conditions [J]. *Otol Neurotol*, 2015, 36(10): 1638-1643.
- [5] Chung K, Zeng FG. Using hearing aid adaptive directional microphones to enhance cochlear implant performance [J]. *Hear Res*, 2009, 250: 27-37.
- [6] Chen YS, Gong Q. Broadband beamforming compensation algorithm in CI front-end acquisition[J]. *Biomed Eng Online*, 2013, 12: 18.
- [7] 陈又圣, 王健, 薛国伟, 等. 电子耳蜗近场信号采集震荡及失配研究[J]. *中国医学物理学杂志*, 2017, 34(7): 731-735.  
Chen YS, Wang J, Xue GW, et al. Vibration and mismatch of near-field signal acquisition for cochlear implant[J]. *Chinese Journal of Medical Physics*, 2017, 34(7): 731-735.
- [8] Li XX, Wang DW, et al. Robust adaptive beamforming using iterative variable loaded sample matrix inverse[J]. *IET Journals & Magazines*, 2018, 54(9): 546-548.
- [9] Yoo I, Kim H, Yang S, et al. Development of directional MEMS microphone single module for high directivity and SNR[J]. *IEEE Sens J*, 2021, 22(7): 6329-6335.
- [10] Tan K, Zhang XL, Wang DL, et al. Deep learning based real-time speech enhancement for dual-microphone mobile phones[J]. *IEEE/ACM Trans Audio Speech Lang Process*, 2021, 29: 1853-1863.
- [11] Lockwood ME, Jones DL, Bilger RC, et al. Performance of time-and frequency-domain binaural beamformers based on recorded signals from real rooms[J]. *J Acoust Soc Am*, 2004, 115(1): 379-391.
- [12] Kan A, Meng QL. The temporal limits encoder as a sound coding strategy for bilateral cochlear implants[J]. *IEEE/ACM Trans Audio Speech Lang Process*, 2021, 29: 265-273.
- [13] Zhang J, Zhang GH. A parametric unconstrained beamformer based binaural noise reduction for assistive hearing[J]. *IEEE/ACM Trans Audio Speech Lang Process*, 2022, 30: 292-304.
- [14] Hu Y, Loizou PC, Li N, et al. Use of a sigmoidal-shaped function for noise attenuation in cochlear implants[J]. *J Acoust Soc Am*, 2007, 122(4): 128-134.
- [15] Lai YH, Tsao Y, Lu XG, et al. Learning-based noise reduction approach to improve speech intelligibility for cochlear implant recipients[J]. *Ear Hear*, 2018, 3(4): 795-809.
- [16] Goehring T, Keshavarzi M, Carlyon RP, et al. Using recurrent neural networks to improve the perception of speech in nonstationary noise by people with cochlear implants[J]. *J Acoust Soc Am*, 2019, 146(1): 705.
- [17] Buyens W, Van Dijk B, Wouters J, et al. A harmonic/percussive sound separation based music pre-processing scheme for cochlear implant users[C]//21st European Signal Processing Conference (EUSIPCO 2013). 2013.
- [18] Padilla M, Stupak N, Landsberger DM. Pitch ranking with different virtual channel configurations in electrical hearing[J]. *Hear Res*, 2017, 348: 54-62.
- [19] Guan T, Yang M, Wei Z, et al. Simulation of the optical stimulation mechanism of cochlear nerves[J]. *Journal of Tsinghua University*, 2017, 57(10): 1102-1105.
- [20] Jiang B, Xia N, Wang X, et al. Auditory responses to short-wavelength infrared neural stimulation of the rat cochlear nucleus [C]//39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'17). IEEE, 2017: 1942-1945.
- [21] Wilson BS, Finley CC. Better speech recognition with cochlear implants[J]. *Nature*, 1991, 352(6332): 236-238.
- [22] 陈又圣, 王健, 薛国伟, 等. 电子耳蜗 CIS 言语处理策略参数特征研究[J]. *深圳信息职业技术学院学报*, 2017, 15(3): 12-18.  
Chen YS, Wang J, Xue GW, et al. Research on parameter feature of continuous interleaved sampling strategy for cochlear implant[J]. *Journal of Shenzhen Institute of Information Technology*, 2017, 15(3): 12-18.
- [23] Skinner MW, Holden LK, Whitford LA, et al. Speech recognition with the Nucleus 24 SPEAK, ACE, and CIS speechcoding strategies in newly implanted adults[J]. *Ear Hear*, 2002, 23(3): 207-223.
- [24] Seligman P, Mc Dermott H. Architecture of the Spectra 22 speech processor[J]. *Ann Otol Rhinol Laryngol Suppl*, 1995, 166: 139-141.
- [25] Stam L, Goverts ST, Smits C. Effect of cochlear implant n-of-m strategy on signal-to-noise ratio below which noise hinders speech recognition[J]. *J Acoust Soc Am*, 2019, 145(5): EL417.
- [26] 陈又圣, 王健, 陈伟芳, 等. 电子耳蜗 n of m 算法的特性[J]. *中国医学物理学杂志*, 2017, 34(3): 266-271.  
Chen YS, Wang J, Chen WF, et al. Feature of n of m algorithm for cochlear implant[J]. *Chinese Journal of Medical Physics*, 2017, 34(3): 266-271.
- [27] Loizou PC. Mimicking the human ear[J]. *IEEE Signal Process Mag*, 1998, 15(5): 101-130.
- [28] Zeng FG, Rebscher S, Harrison W, et al. Cochlear implants: system design, integration, and evaluation[J]. *IEEE Rev Biomed Eng*, 2008, 1: 115-142.
- [29] Müller J, Brill S, Hagen R, et al. Clinical trial results with the MED-EL fine structure processing coding strategy in experienced cochlear implant users[J]. *ORL J Otorhinolaryngol Relat Spec*, 2012, 74(4): 185-198.
- [30] Nogueira W, Litvak LM. Signal processing strategies for cochlear implants using current steering[J]. *Eurasip J Adv Signal Process*, 2009, 1: 1-21.

(编辑:薛泽玲)