

Channel-Vocoder-Centric Modelling of Cochlear Implants: Strengths and Limitations



Fanhui Kong, Yefei Mo, Huali Zhou, Qinglin Meng, and Nengheng Zheng

Abstract Modern cochlear implants (CIs) generate electric current pulsatile stimuli from real-time incoming to stimulate residual auditory nerves of deaf ears. In this unique way, deaf people can (re)gain a sense of hearing and consequent speech communication abilities. The electric hearing mimics the normal acoustic hearing (NH), but with a different physical interface to the neural system, which limits the performance of CI devices. Simulating the electric hearing process of CI users through NH listeners is an important step in CI research and development. Many acoustic modelling methods have been developed for simulation purposes, e.g., to predict the performance of a novel sound coding strategy. Channel vocoders with noise or sine-wave carriers are mostly popular among the methods. The simulation works have accelerated the procedures of re-engineering and understanding of the electric hearing. This paper presents an overview of the literature on channel-vocoder simulation methods. Strengths, limitations, applications, and future works about acoustic vocoder simulation methods are introduced and discussed.

Keywords Cochlear implant · Auditory prosthesis · Speech perception · Hearing research · Pitch · Vocoder

F. Kong · H. Zhou · N. Zheng (✉)
Guangdong Key Laboratory of Intelligent Information Processing,
College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060,
Guangdong, People's Republic of China
e-mail: nhzheng@szu.edu.cn

Y. Mo · H. Zhou · Q. Meng (✉)
Acoustic Laboratory, School of Physics and Optoelectronics, South China University of
Technology, Guangzhou 510610, Guangdong, People's Republic of China
e-mail: mengqinglin@scut.edu.cn

Y. Mo
School of Medicine, South China University of Technology, Guangzhou 510006, Guangdong,
People's Republic of China

1 Introduction

Auditory perception relies on the normally functioning peripheral auditory coding within the inner ear (or cochlea). Thousands of hair cells, which lie between the basilar membrane and the tectorial membrane, convert vibrations of sounds coming through the oval window into action potentials according to the physical properties of the sound pressure waveform. The auditory nerve fibers synaptically connected to the hair cells are then electrically excited, and the sound signals are further encoded in both place and temporal patterns of spikes up to the brain [1]. Unfortunately, the cochleae may lose their function due to some reasons from genetic defects, noise exposure, drug toxicity, aging, etc. If an ear loses the normal acoustic hearing (NH) severely and cannot transmit enough speech information to the brain, a cochlear implant may be highly suggested by an audiologist.

Cochlear implants (CIs), by bypassing the cochlear hair cells and directly stimulating the auditory nerve fibers, have successfully helped more than 1 million deafened people all over the world to (re)gain hearing. A CI converts incoming sounds captured by one or multiple microphones into electric pulses and then stimulates the residual auditory nerve fibers through an electrode array which is surgically implanted in the scala tympani of the cochlea [2].

Most CI users can understand speech in favorable conditions. However, their sound perception abilities are still abnormal compared with their NH peers in many aspects, e.g. speech recognition in noise [3] and reverberation [4], music appreciation [5], articulatory features (F0 and formants) [6, 7], sound localization [8], and atypical pronunciation [9, 10]. Among individuals of any CI cohort, large performance variances are observed. What's more, the performance of a CI recipient cannot be precisely predicted before surgery, and the low-performance listeners usually cannot be trained to gain large improvement [11].

Technically, the main limitation comes from the electrode-to-neuron interface, i.e., the CI electrode array and its stimulation to the neurons. Information of different frequencies is coded in different stimulation places, i.e., different electrodes. A modern CI typically has 12 to 24 [12] electrodes, which are far less in number and larger in size than the hair cells in a healthy cochlea [13]. Compared with the natural frequency tuning of normal auditory nerve fibers, the tuning of the CI electrode stimulation is much broader. The broad frequency-tuning feature and the limited number of electrodes lead to a poor resolution of frequency information with the electric hearing. The electrode array mostly cannot reach the apical turn of the cochlea and a large mismatch exists in the mapping of all CI recipients. Furthermore, the temporal characteristics of the auditory nerve fibers responding to artificial electric pulse trains and natural acoustic sounds are significantly different [14], which may explain the temporal limits of electric hearing observed in many psychophysical studies. Healthy cochleae show phase-locking features to sinusoidal components with frequencies up to about 4 to 5 kHz. In contrast, the upper limit of temporal sensitivity of the electric hearing is only a few hundred hertz (see [15]).

The sound processing strategies also play an important role in the performance of CI users. Most modern clinically available CI signal processing strategies are based on temporal envelopes. Even though the physical resolutions of the CI interface are very coarse, there are many electric parameters (e.g., frequency mapping, electrode channel number, current spread, dynamic range, temporal envelope cut-off frequencies, and pulse shape) can be permuted to optimize the electric hearing performance. To this end, numerous efforts have been made to improve CI strategies.

To accelerate the procedure of the (re-)engineering of CI strategies and the fundamental research on CI understanding, several acoustic models have been proposed to simulate the electric hearing, e.g., in [16]. In that historical paper, it was stated that *“The main reason for formulating this acoustic model is to facilitate the development of speech coding schemes for use with multiple-channel cochlear implants. The model allows a normally hearing listener to gain insights into the capability of the implant and to assess alternative speech coding schemes first hand.”*

Acoustic CI models analyze a sound according to the core steps in a CI strategy and re-synthesize it into a new sound, which can be presented to NH subjects through headphones or speakers. The differences observed in NH listeners using acoustic CI models with different parameters may be used to predict the performance in CI listeners using strategies with corresponding different parameters. This is an ideal hypothesis, and the models are helpful research tools since CI listeners are much less accessible and more variable than NH listeners. Simulation experiments in NH listeners usually result in low-variance data and sometimes overestimate the mean actual CI performance. These are also not surprising, given that there are many physical differences between the neuron interface of the electric and acoustic hearing.

The listening tests with human subjects are usually time-consuming. Reasonable parameters should be carefully selected to get valuable results. However, in practice, because of the complicated multidisciplinary concepts, large variance of CI subjects, and research purposes, it is usually difficult to make the appropriate choice during the model designing especially for new researchers in the field.

This paper aims to review the strengths and limitations of the most conventional acoustic CI simulation methods, i.e., channel vocoders, and to outlook the future development of CI simulations. In other words, this work aims to provide an overview of the history, current status, and future directions of the channel-vocoder-centric CI simulation studies. Detailed arrangements are as follows. In Sect. 2, current CI strategies are briefly introduced. In Sect. 3, the channel vocoders and the physical parameters simulated in them are discussed.

2 CI Signal Processing Strategies: Interleavedly Sampling the Temporal Envelopes

A standard CI strategy framework is shown in Fig. 1(A). It was proposed and evaluated in the early 1990s by Wilson and colleagues and named as continuous

interleaved sampling (CIS) strategy [17]. The key features of CIS are 1) extracting temporal envelopes from multiple frequency channels, 2) stimulating each electrode place using a biphasic pulse train whose amplitude is modulated by the envelope extracted in feature 1, and 3) only one electrode place is stimulated at one time, i.e., there is no time overlapping between stimulations on multiple electrodes. Detailed implementations of each stage may be slightly different according to engineering choices.

The CIS strategy is integrated in almost all CI products, with some new features over the naive CIS. For example, not all the electrodes are stimulated in a processing frame in the advanced combination (ACE) strategy in Cochlear devices. Typically, only eight electrodes with maximum amplitudes (maxima) are dynamically selected from the 22 electrodes for stimulation in each processing frame. It is a typical model of the n -of- m strategies [2]. In the fine structure processing (FSP) strategy in MED-EL devices, the zero-crossing points were used for the timing of electric pulses at the lowest 2–4 channels, which is assumed to introduce some temporal fine structures. In the HiRes120 strategy in Advanced Bionics devices, a current steering (or virtual channel) technique, i.e., using simultaneous firing neighbor electrodes to induce

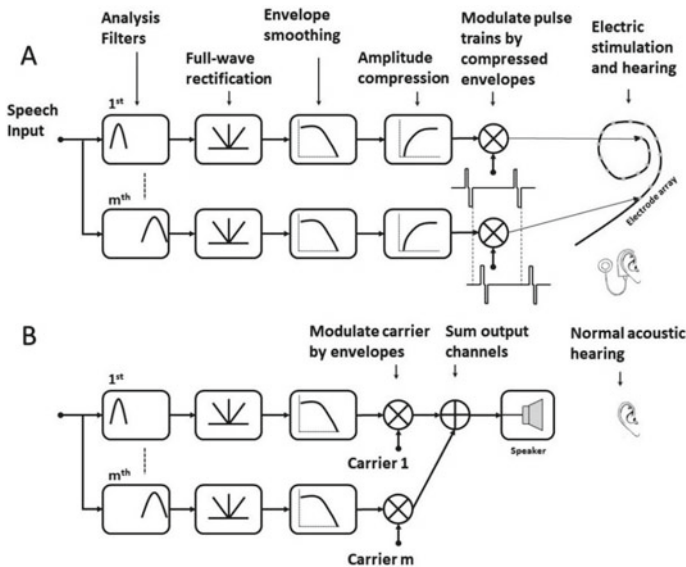


Fig. 1 Block diagrams of a typical CI strategy, i.e., continuous interleaved sampling (CIS) processing strategy (A), and of a channel vocoder modelling algorithm (B). **A:** The input speech sound is filtered through a bandpass filter bank. For the output from each channel, the following processing stages include rectification, envelope smoothing, amplitude compression, and pulse modulation. Then the amplitude-modulated pulse train is transmitted to the corresponding electrode in the cochlea. **B:** The channel vocoder can use the same procedure as CIS to get the temporal envelopes. Then the envelopes are used to amplitude-modulate sine-wave or noise carriers. The modulated carriers from all bands are superimposed to generate a sound signal to be presented to normal hearing ear(s) through a loudspeaker or headphones

intermediate pitch perception between physical electrodes, was used. 120 virtual channels are assumed to be created among the 16 physical electrodes [18].

Even though different variations (including ACE, FSP, and HiRes120) of the strategies have been advocated by different companies, they share the key features (i.e., temporal envelope-based interleaved sampling bi-phasic pulses) with CIS. There is no consistent evidence to support the significant advantages of any strategy over the others.

3 Channel Vocoders: The Algorithms and Applications

3.1 Algorithms of the Channel Vocoders

It is impossible to replicate the stimulation pattern delivered to auditory nerve fibers by CIs via acoustic stimulation in a normally hearing ear. Still, it can nevertheless be helpful for research and demonstration purposes to simulate CI processing using acoustic models. Since the 1990s, the most popular simulation method has been the channel vocoder, as shown in Fig. 1.B.

The analysis band-pass filters and the envelope extraction steps can be the same as the actual CI strategies. To re-synthesize a sound stimulus for NH listeners, the temporal envelopes were used to amplitude-modulate either band-limited noise carriers [19] or sine-wave carriers [20]. The amplitude-modulated results are summed up into one sound. In some cases, band limiting filters and power normalization were inserted between modulation and summing. In the algorithm, all parameters of all stages can be manipulated to simulate different aspects of the CI strategy. The vocoder-centric simulation has produced a cottage industry of auditory research [21]. The key physical parameters of CIs and their simulation in the channel vocoders are discussed as below.

Before going into the detailed parameters, it should be noted that brain plasticity should always be taken care of in vocoded speech perception experiments. CI listeners, at the time of experiments, usually have at least months of experience in listening to the artificial sound [22], whereas NH subjects are usually naive listeners to the vocoded speech. The hearing experience may influence their familiarity with the stimuli and thus affect their performance.

3.2 Frequency Allocation

The CI frequency allocation (i.e., which frequency range is allocated to which electrode) is based on the tonotopic organization of the cochlea, i.e., more basal electrodes correspond to higher frequency bands. However, the detailed allocations in CI products are quite arbitrary. For example, the frequency allocation of the ACE strategy

seems to be partially determined by the fast-Fourier-transform bins [23] and is not precisely following the physiological Green-wood frequency-place coding function [24]. Nonetheless, the CI frequency allocation in analysis filters as in Fig. 1(A) can be perfectly simulated in the analysis filter design of the channel vocoders as in Fig. 1(B). The analysis filter frequency allocation cannot guarantee either place-matching or harmonic relationships in the synthesis parts. The (center) frequencies of sine-wave carriers or band-limited noise carriers can be adjusted to match the physiological characteristic frequencies of the auditory neurons being stimulated by corresponding electrodes. The matching or mismatching degree can be directly simulated by changing the (center) frequencies, e.g., for electrode insert depth simulation [25] and bilateral mismatching simulation [26].

3.3 Spectral Channel Number

One spectral channel (band) usually corresponds to one electrode. The number of band-pass filters can be used to generate different numbers of spectral channels for both actual CIs and channel vocoders. Figure 2 shows examples of four and 16-channel vocoder simulated sentence speech with both sine-wave and noise carriers. While the channel number in actual CIs is limited by the physical electrode number, the channel number can be set in a much larger range for the simulations. Shannon et al. (1995) [19] has shown that temporal envelopes from four spectral channels can provide enough information for speech intelligibility. Dorman et al. 1998 [27] confirmed this finding again and showed that for recognition of speech in noisy conditions larger channel numbers are needed than in quiet conditions. This larger number need was confirmed by recent actual CI data [28]. [19] and [27] are the beginnings of the channel vocoder used as tools for CI simulation. Friesen et al. 2001 [29] found that a vocoder with about eight channels can simulate the speech recognition performance of the best CI results, so many works used vocoders with no more than eight channels to simulate the CI performance even the majority of CI devices have more than eight electrode channels.

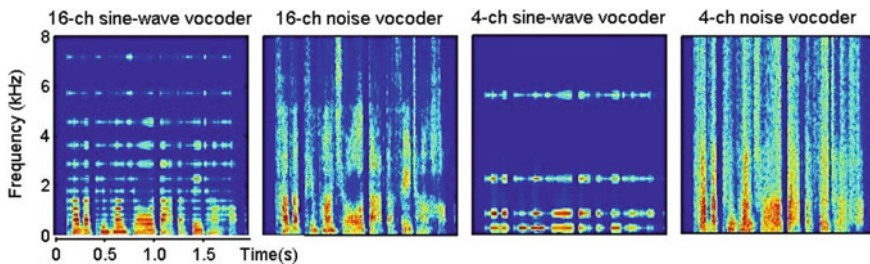


Fig. 2 Examples of spectrograms of channel-vocoded speech

3.4 *Current Spread*

The maximum electrode number is limited by the current spread of the electric stimulation on individual electrodes. Because of the large electrode size and the long neuron-electrode distance, the electrode current spread is too broad to provide a fine spectral resolution [13]. Different current spread degrees can be simulated by manipulating the bandwidth and slopes of the frequency responses of the band-pass filters. For example, in [30] carrier filter slopes of -24 or -6 dB/octave were used, and they found that a steeper slope representing less channel interactions may provide better speech intelligibility in noisy conditions. Another way to simulate the current spread is by adding weighted contributions from other bands' envelopes to current band' envelope, e.g., in [31]. Sometimes, bandlimited filters can be inserted between the carrier modulation and the output summing in Fig. 1.B to control the interaction to some extent.

3.5 *Temporal Envelope*

In [32], Rosen proposed a framework of the temporal envelope, periodicity, and fine structure to analyze the temporal structure of speech. In CI strategies, temporal envelopes from multiple channels are extracted. Cut-off frequencies of the low-pass filters (in Fig. 1) can be used to control the fluctuation rate of envelopes for both actual and simulated CI stimuli. Channel vocoders use continuous noise carriers or sin-wave carriers, while CIs use bi-phasic pulse train carriers. Noise carriers have been shown to introduce more intrinsic temporal fluctuation, which may interfere with the speech's temporal envelope to be transmitted [33]. Higher cut-off frequencies (e.g., 500 Hz) can include more periodicity information in the envelopes. With channel vocoders, the periodicity cues have been used in many studies to simulate CI pitch perception tasks, such as lexical tone recognition [34], voice gender discrimination [35], and stream segregation [36, 37]. However, the different physical characteristics of the carriers may prevent the vocoders from accurately modelling the periodicity-based CI pitch perception. What's more, a trade-off phenomenon in phoneme recognition was reported between the temporal fluctuation rate and the channel number [38].

The carrier effects should be carefully handled in both experimental design and discussions about implications on actual CIs. Another method of envelope extraction is using the Hilbert transform, i.e., calculating the magnitude of the analytic signal of the input band signal. The difference between the rectification-low-pass-filtering method and the Hilbert method are very tricky. The former is more physically meaningful, especially when we want to control the periodicity cues for our applications.

3.6 *Intensity and Dynamic Range*

NH listeners can hear sounds in a large intensity range of up to 120 dB SPL, with a resolution of 1–2 dB, whereas CI users have a much narrower dynamic range. The dynamic range of a CI user is defined as the difference between the current intensities at which a sound is just audible (the threshold) and perceived as uncomfortable. For example, in Cochlear’s implants, eight bits (i.e., 255 steps) [23] were used to quantify the current intensity, and most recipients are fitted with a range covering only a portion of the 255 steps. This means tens of dB SPL range (e.g., 35 to 85 dB SPL) will be compressed to only a several dB range. (Information about CI compression can be referred to [39]). The rationale has been explained by Loizou et al. [40] as that eight bits were enough for vowel and consonant recognition with 6-ch actual and simulated CIs. However, to the best of our knowledge, there’s no more detailed research on this intensity quantization in electric hearing. It is unknown whether higher intensity resolution can improve the intelligibility and quality of sounds in various conditions. This is a reasonable question, especially for the people who are headphone enthusiasts. What’s more, in most CI simulation studies, the original quantization resolution (e.g., 16 bits or 32 bits) of the input audio was kept unchanged through the whole vocoding procedure. It is also unknown whether the quantization resolution would influence the vocoded sound perception.

The dynamic range compression could be simulated by inserting a compression stage between the modulation and the summing stage in Fig. 1(B), and it was demonstrated to affect speech recognition negatively (e.g., in [41]), but it is often omitted in CI simulation studies.

3.7 *Carrier Waveform*

As mentioned above, the most conventional carriers are sine-wave and noise signals. They are both continuous signals. One essential defect of the conventional channel vocoders is the lack of pulsatile characteristics, while pulsatile stimulation is known to be one of the key reasons of CIS’s success. The continuous carrier waveforms of channel vocoders cannot be used to simulate the electric pulse trains, which was almost completely ignored in hundreds of CI simulation papers. One possible reason is that most current CI strategies use a high stimulation rate (≥ 900 pulse-per-second, pps), which is assumed to be much higher than the upper temporal pitch limits of the electric hearing. Hence, the pulsatile characteristics of electric pulse trains can be viewed as too fast to be discriminated. However, this defect is worth further investigation to provide a better simulation model for signal processing and phenomenological comparisons.

3.8 Short Summary

Frequency allocation, spectral channel number, current spread, and temporal envelope of CI stimuli can be conveniently simulated by manipulating the corresponding physical parameters in channel vocoders.

However, the intensity and dynamic range were simulated in few studies and the pulsatile features of CI stimuli were almost completely ignored [54]. Researchers have carried out many experimental studies to verify and get insights from the simulations. For a channel-vocoded sound, all of the physical consistencies or disparities can affect the explanation of the sound perception performance.

4 Channel-Vocoder Simulation vs. Actual CI Hearing

To examine the power of CI simulations, it is helpful to directly compare vocoded speech perception performance in normal hearing (NH) subjects and actual CI patients. Among the earliest publications on direct vocoder-vs-CI comparisons, Fu et al. 1998 [42] showed that using 4 channels, the vocoder simulations with noise carriers, and actual CIS strategy in CI patients can lead to similar performance as actual CI subjects using the same electrode channel numbers. Friesen et al. 2001 [29] showed that eight channels could lead to similar speech-in-noise recognition performance between the NH and CI groups, while higher channel numbers than eight can improve the performance of vocoder simulation but not for actual CI patients. Stickney et al. 2004 [43] compared 4-ch and 8-ch noise vocoders and five Nucleus CI users in a task of speech recognition in a speech masker. It is shown that 8-ch vocoder results are better than the CI subjects using 8-of-22 (or 8-of-24) ACE strategies. Fu et al. (2004) [35] showed that eight-channel sine-wave vocoders derived better speaker vowel recognition results than CI subjects and the envelope cutoff frequencies has no significant effect on the simulation results. Fu et al. (2005) [44] found that 4-ch noise band channel vocoders with -6 dB/oct bandpass filters lead to similar speech-in-Gated-Noise recognition. Iverson et al. (2006) [45] found actual CI subjects can use the formant movement for vowel recognition only as good as about 6-ch noise vocoders.

In many pure simulation studies (without actual CI tests), less than eight channels are used. This is far less than the current actual CI channel numbers (12–24). Their selection of the small channel number is mainly because that channel vocoders with those low channel numbers can lead to similar performance as actual CI listeners in that experiment tasks to be done, for example, in [46–48].

Actually, this often-used assumption does not fully make sense. It is a major limitation of the channel vocoders. The intrinsic aim of vocoder modelling of CIs is not only to generate a vocoder method who can derive similar scores with CIs in one or several tasks. Researchers are always willing to have a model which can model many phenomena in many perception tasks and most physical parameters of

actual CI stimuli. To make the vocoders with higher (but closer to actual CIs) channel numbers derive similar performance as CI listeners, channel interactions, frequency shifting, and dynamic range compression can be manipulated in the channel vocoders to some extent. In some works, alternative vocoder methods were designed to solve this problem and better mimic the electric stimulation or physiological procedure, e.g., in [54] and [55].

5 Sound Quality and Music Perception with Vcoded Sounds

When keywords of “((cochlear implant) AND (vocoder))” were searched on PubMed.gov on May 11, 2020, there were 306 results. When “Speech” was added to the keywords, 293 results were shown. This tells that most studies on CI vocoder simulation are working on speech sounds. All parameters of the channel vocoders can be enumerated to examine the physical parameters’ effects on simulated speech perception. The results and experiments are diverse and cannot be logically concluded here. Instead, we re-state here that channel-vocoded simulation should be carefully explained when implications for actual CI are produced. For example, in studies including both vocoded NH and actual CI cohorts, using the same channel number NH subjects’ results were usually much better than CI results, e.g. in [10, 29, 45].

When keywords of “((cochlear implant) AND (vocoder)) AND (Music)” were searched in PubMed.gov, there were only 22 results. This tells that music appreciation cannot be easily simulated by the channel vocoders. The quality of vocoded music is too different from that of original music for NH listeners to be appreciated, while CI subjects have no choice and long-time experience.

Another valuable group of CI subjects, i.e., single-sided deafness CI implantees, provide opportunities for direct quality comparison of sounds from vocoders and CIs. For example, in [49] and [50], different subjects prefer different vocoder configurations. We cannot consistently know which vocoder method can best predict the sound quality of CI hearing.

6 How to Simulate New Experimental Strategies?

So far, we mainly discussed the vocoder simulation of conventional CIS-like envelope-based strategies. Engineering researchers have proposed some temporal fine structure enhancement strategies, for example, in [51–53]. Conventional channel vocoders should be modified to include the temporal fine structure cues. It should be noted here that actual CI experiments will always be the golden standard to evaluate a new signal processing strategy, although the simulation experiment work may provide some implications.

7 Conclusion

The most popular acoustic modelling methods of CI hearing are the channel vocoders using sine-wave and noise carriers. Most critical physical parameters, including channel numbers, frequency allocations, current spread, temporal envelope fluctuation rate, and intensity and dynamic range can be simulated by the channel vocoders. They are powerful tools to simulate many aspects of cochlear implant hearing. However, the limitation is also significant. Many recent studies still use channel vocoders with no more than eight channels to simulate actual CI hearing with 12 to 24 channels, just because their performance is similar in some speech tasks. To make the vocoded hearing and CI hearing using the same channel numbers comparable to each other in performance, current spread and frequency shifting can be manipulated. Also, some alternative methods using novel carrier signals may provide more direct and quantitatively practical simulation to the cochlear implants. The channel-vocoder-centric modelling of cochlear implants is a useful method, but researchers should take care of the explanation of the vocoded results, especially in the quality judgments and novel temporal fine structure strategy evaluations.

Acknowledgements This work is jointly supported by the National Natural Science Foundation of China (11704129 and 61771320), Guangdong Basic and Applied Basic Research Foundation Grant (2020A1515010386), and the Science and Technology Program of Guangzhou (202102020944).

References

1. Moore BCJ (2013) An introduction to the psychology of hearing, 6th edn. Brill, Leiden
2. Zeng FG, Rebscher S, Harrison W, Sun X, Feng H (2008) Cochlear implants: system design, integration, and evaluation. *IEEE Rev Biomed Eng* 1:115–142
3. Zhou H, Wang N, Zheng N, Yu G, Meng Q (2020) A new approach for noise suppression in cochlear implants: a single-channel noise reduction algorithm. *Front Neurosci* 14:301
4. Kressner AA, Westermann A, Buchholz JM (2018) The impact of reverberation on speech intelligibility in cochlear implant recipients. *J Acoust Soc Am* 144:1113–1122
5. Nogueira W, Nagathil A, Martin R (2019) Making music more accessible for cochlear implant listeners: recent developments. *IEEE Signal Process Mag* 36:115–127
6. Meng Q, Zheng N, Mishra AP, Luo JD, Schnupp JW (2018) Weighting pitch contour and loudness contour in mandarin tone perception in cochlear implant listeners. In: *Interspeech*
7. Gaudrain E, Baskent D (2018) Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear Hear* 39:226–237
8. Jones H, Kan A, Litovsky RY (2014) Comparing sound localization deficits in bilateral cochlear-implant users and vocoder simulations with normal-hearing listeners. *Trends Hear*
9. Li Y, Zhang G, Kang HY, Liu S, Han D, Fu QJ (2011) Effects of speaking style on speech intelligibility for Mandarin-speaking cochlear implant users. *J Acoust Soc Am* 129:EL242–EL247
10. Meng Q et al (2019) Time-compression thresholds for Mandarin sentences in normal-hearing and cochlear implant listeners. *Hear Res* 374:58–68
11. Faulkner KF, Pisoni DB (2013) Some observations about cochlear implants: challenges and future directions. *Neurosci Disc* 1:9

12. Zeng FG et al (2015) Development and evaluation of the Neurotron 26-electrode cochlear implant system. *Hear Res* 322:188–199
13. Zeng FG (2017) Challenges in improving cochlear implant performance and accessibility. *IEEE Trans Biomed Eng* 64:1662–1664
14. Boulet J, White M, Bruce IC (2016) Temporal considerations for stimulating spiral ganglion neurons with cochlear implants. *J Assoc Res Otolaryngol* 17:1–17
15. Cosentino S, Carlyon RP, Deeks JM, Parkinson W, Bierer JA (2016) Rate discrimination, gap detection and ranking of temporal pitch in cochlear implant users. *J Assoc Res Otolaryngol* 17:371–382
16. Blamey PJ, Dowell RC, Tong YC, Clark GM (1984) An acoustic model of a multiple-channel cochlear implant. *J Acoust Soc Am* 76:97–103
17. Wilson BS (2015) Getting a decent (but sparse) signal to the brain for users of cochlear implants. *Hear Res* 322:34–38
18. Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM (1991) Better speech recognition with cochlear implants. *Nature* 352:236–238
19. Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. *Science* 270:303–304
20. Dorman MF, Loizou PC, Rainey D (1997) Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *J Acoust Soc Am* 102:2403–2411
21. Goupell MJ (2015) Pushing the envelope of auditory research with cochlear implants. *Acoust Today* 11:26–33
22. Reiss LA, Gantz BJ, Turner CW (2008) Cochlear implant speech processor frequency allocations may influence pitch perception. *Otology Neurotol* 29:160
23. Vandali AE, Whitford LA, Plant KL, Clark GM (2000) Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system. *Ear Hear* 21:608–624
24. Greenwood DD (1990) A cochlear frequency-position function for several species—29 years later. *J Acoust Soc Am* 87:2592–2605
25. Faulkner A, Rosen S, Stanton D (2003) Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth. *J Acoust Soc Am* 113:1073–1080
26. Xu K, Willis S, Gopen Q, Fu QJ (2020) Effects of spectral resolution and frequency mismatch on speech understanding and spatial release from masking in simulated bilateral cochlear implants. *Ear Hear* 41:1362
27. Dorman MF, Loizou PC, Fitzke J, Tu Z (1998) 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* 104:3583–3585
28. Croghan NBH, Duran SI, Smith ZM (2017) Re-examining the relationship between number of cochlear implant channels and maximal speech intelligibility. *J Acoust Soc Am* 142:EL537–EL543
29. Friesen LM, Shannon RV, Baskent D, Wang X (2001) Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am* 110:1150–1163
30. Fu QJ, Nogaki G (2005) Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *J Assoc Res Otolaryngol* 6:19–27
31. Oxenham AJ, Kreft HA (2014) Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing. *Trends Hear* 18 2014.
32. Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. *Phil Trans Royal Soc Lond Ser B Biol Sci* 336:367–373
33. Whitmal NA, Poissant SF, Freyman RL, Helfer KS (2007) Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience. *J Acoust Soc Am* 122:2376–2388
34. Yuan M, Lee T, Yuen KC, Soli SD, van Hasselt CA, Tong MC (2009) Cantonese tone recognition with enhanced temporal periodicity cues. *J Acoust Soc Am* 126:327–337

35. Fu QJ, Chinchilla S, Galvin JJ (2004) The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *J Assoc Res Otolaryngol* 5:253–260
36. Gaudrain E, Grimault N, Healy EW, Bera JC (2008) Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation. *J Acoust Soc Am* 124:3076–3087
37. Steinmetzger K, Rosen S (2018) The role of envelope periodicity in the perception of masked speech with simulated and real cochlear implants. *J Acoust Soc Am* 144:885–896
38. Xu L, Thompson CS, Pfingst BE (2005) Relative contributions of spectral and temporal cues for phoneme recognition. *J Acoust Soc Am* 117:3255–3267
39. Zeng F-G (2004) *Compression and cochlear implants*. Springer, Heidelberg
40. Loizou PC, Dorman M, Poroy O, Spahr T (2000) Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution. *J Acoust Soc Am* 108:2377–2387
41. Chen F, Zheng D, Tsao Y (2017) Effects of noise suppression and envelope dynamic range compression on the intelligibility of vocoded sentences for a tonal language. *J Acoust Soc Am* 142:1157–1166
42. Fu QJ, Shannon RV, Wang X (1998) Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing. *J. Acoust Soc Am* 104:3586–3596
43. Stickney GS, Zeng FG, Litovsky R, Assmann P (2004) Cochlear implant speech recognition with speech maskers. *J Acoust Soc Am* 116:1081–1091
44. Fu QJ, Nogaki G (2005) Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *J Assoc Res Otolaryngol* 6:19–27
45. Iverson P, Smith CA, Evans BG (2006) Vowel recognition via cochlear implants and noise vocoders: effects of formant movement and duration. *J Acoust Soc Am* 120:3998–4006
46. Luo X, Fu QJ (2004) Enhancing Chinese tone recognition by manipulating amplitude envelope: implications for cochlear implants. *J Acoust Soc Am* 116:3659–3667
47. Zaltz Y, Goldsworthy RL, Kishon-Rabin L, Eisenberg LS (2018) Voice discrimination by adults with cochlear implants: the benefits of early implantation for vocal-tract length perception. *J Assoc Res Otolaryngol* 19:193–209
48. Zaltz Y, Goldsworthy RL, Eisenberg LS, Kishon-Rabin L (2020) Children with normal hearing are efficient users of fundamental frequency and vocal tract length cues for voice discrimination. *Ear Hear* 4:182–193
49. Peters JP, et al (2018) The sound of a cochlear implant investigated in patients with single-sided deafness and a cochlear implant
50. Dorman MF, et al (2017) The sound quality of cochlear implants: studies with single-sided deaf patients
51. Nie K, Stickney G, Zeng FG (2005) Encoding frequency modulation to improve cochlear implant performance in noise. *IEEE Trans Biomed Eng* 52:64–73
52. Meng Q, Zheng N, Li X (2016) Mandarin speech-in-noise and tone recognition using vocoder simulations of the temporal limits encoder for cochlear implants. *J. Acoust Soc Am* 139:301–310
53. Goldsworthy RL (2019) Temporal envelope cues and simulations of cochlear implant signal processing. *Speech Communication*
54. Meng Q, Zhou H, Lu T, Zeng FG (2022) Gaussian-Enveloped Tones (GET): a vocoder that can simulate pulsatile stimulation in cochlear implants. medRxiv
55. Brochier T, Schlittenlacher J, Roberts I, Goehring T, Jiang C, Vickers D, Bance M (2022) From microphone to phoneme: an end-to-end computational neural model for predicting speech perception with cochlear implants. *IEEE Trans Biomed Eng*