Applied Acoustics 208 (2023) 109386

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Technical note

Pulsatile Gaussian-Enveloped Tones (GET) for cochlear-implant simulation

Qinglin Meng [a,*], Huali Zhou [b,a], Thomas Lu [c], Fan-Gang Zeng [c,*]

[a] Acoustics Laboratory, School of Physics and Optoelectronics, South China University of Technology, Guangzhou 510641, Guangdong, China
[b] Guangdong Key Laboratory of Intelligent Information Processing, College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, Guangdong, China
[c] Center for Hearing Research, Department of Otolaryngology -Head and Neck Surgery, University of California, Irvine 92697, CA, USA

ARTICLE INFO

Article history:
Received 8 December 2022
Received in revised form 22 March 2023
Accepted 10 April 2023

Keywords:
Cochlear implant
Channel vocoder
Gabor atom
Speech perception

ABSTRACT

Acoustic simulation of cochlear implants (CIs) allows studies of not only perceptual performance in normal-hearing listeners but also relative contribution of spectral and temporal cues to speech recognition. Different from conventional simulations using continuous sinusoidal or noise carriers, this study employs pulsatile Gaussian-enveloped tones (GETs) to simulate several key features in modern CIs. Subject to the time–frequency uncertainty principle, the GET has a well-defined tradeoff between its duration and bandwidth. In theory, the key pulsatile features of CI electric stimuli could be simulated in the analysis part of a pulsatile acoustic model. We hypothesized that a GET vocoder could transmit speech information similar to actual CIs. In an experiment, 900-Hz/channel pulse trains were directly mapped to 900-Hz GET trains to simulate the maxima selection and amplitude compression of a widely-used *n*-of-*m* processing strategy, or the Advanced Combination Encoder. The simulated and actual implant performance of speech-in-noise recognition was similar in terms of the overall trend, absolute mean scores, and standard deviations. The results suggest that the pulsatile GET can serve as an alternative to simulating modern CIs in terms of simulating pulsatile and several other key features of CI stimuli and, more importantly, producing similar speech perception patterns. Investigation is needed to further evaluate the performance of the GET in different acoustic and electric configurations.

1. Introduction

Vocoders as a means of speech synthesis have a long and rich history. At the 1939 New York World's Fair, Homer Dudley of Bell Labs demonstrated his vocoder invention that could "remake speech" automatically and instantaneously (18-ms delay) by controlling energy in 10 frequency bands (from 0 to 3000 Hz) that contained either buzz-like tone or hiss-like noise carriers [1]. He later realized that the vocoder could be used in synthesizing speech, and transformed in various ways to study the relative contributions of fundamental parameters in speech synthesis and recognition. He found that good intelligibility can be achieved by controlling "*only low syllabic frequencies of the order of 10 cycles per second*", whereas the emotional content of speech can be controlled by altering the frequency of the buzzing tones.

Cochlear implants (CIs) are medical devices that help patients with severe-to-profound hearing loss regain hearing abilities. The development of various CI signal processing strategies is closely related to speech synthesis vocoders. The early multi-channel CIs followed Dudley's original vocoder idea closely by extracting and delivering speech fundamental frequency (F0) in the form of electric pulse rate and one or two formants (F2 or F1/F2) in the form of electrode position. The speech understanding of the early CIs was relatively low ($< 50\%$ correct for sentence recognition in quiet), due not only to crude F0 and formant extraction methods (i.e., zero-crossing) at that time, but, more importantly, to the complicated CI-related conditions, for example, individual variability in electrode insertion angle or depth, cochlear vs. ganglion cell tonotopic organization, current spread, and nerve survival. These interactions make accurate F0 and formant representation difficult if not impossible even if both F0 and formants can be exactly extracted by today's algorithms. As a result, contemporary CIs have abandoned the F0 and formant extraction method but adopted

Abbreviations: CI, cochlear implant; ITD, interaural time difference; GET, Gaussian-Enveloped Tones; ACE, advanced combination encoder; SRT, speech reception threshold; TCT, time-compression threshold; SSN, speech-shaped noise; PTA, pure-tone audiometry.

* Corresponding authors.
E-mail addresses: mengqinglin@scut.edu.cn (Q. Meng), fzeng@uci.edu (F.-G. Zeng).

https://doi.org/10.1016/j.apacoust.2023.109386
0003-682X/

speech processing strategies that extract band-specific temporal envelopes from 8–24 frequency bands. The envelopes are used to amplitude modulate a continuous, but fixed, high-rate (at least two to four times the highest envelope frequency) pulss then delivered to a corresponding electrode in an interleaved fashion in which two electrodes never fire simultaneously [2]. These advances in multi-channel CIs have produced 70–80% correct sentence recognition in quiet, which is sufficient for an average user to carry on a conversation without lipreading [3].

Acoustic simulations of CIs have been developed and widely used for at least three reasons. First, acoustic simulations minimize the effect of large CI individual variability (e.g., cognitive differences, demographic variables, and electrode-neuron interface), which may confound or mask the relative importance of speech processing parameters, e.g., [4]. Second, acoustic simulations allow the evaluation of relative contributions of different cues to auditory and speech perception, e.g., [5–7]. Third, acoustic simulations allow a normal-hearing listener to appreciate the quality of CI processing and the degree of difficulty facing a typical CI user.

Traditionally, acoustic simulations of CIs have used either noise- [8] or sinusoid-excited [9] vocoders. In these vocoders, the noise or sinusoid simulates the electric pulse train, while the number of frequency bands and their overlaps simulate the limited number of electrodes and their current spread, e.g., [8]. A significant drawback of these traditional vocoder models is the lack of simulation of the pulsatile nature of CI electric stimulation. Several studies have attempted to develop acoustic models that simulate pulsatile electric stimulation, such as filtered noise bursts [10,11], filtered harmonic complex tones [12], and pulse-spread harmonic complexes [13,14]. However, there are limitations to those methods in simulating some important features in modern CIs. For example, these vocoders cannot simulate the discrete nature of pulsatile stimulation on a pulse-by-pulse basis. It is also difficult for vocoders using continuous carriers to simulate some CI speech processing strategies, e.g., *n*-of-*m*, in which the *n* low-energy bands of all *m* bands are abandoned to produce temporally separated envelopes.

Here we identified the Gabor atom [15], also known as the Gaussian-enveloped tone (GET), as a means of simulating the essential features of modern CI processing as discussed above. The GET has been used in many psychoacoustic studies on a wide range of auditory phenomena in normal-hearing or hearing-impaired listeners, e.g., temporal gap detection [16,17], intensity discrimination [18–21], simultaneous and non-simultaneous masking [22,23], interaural timing difference (ITD) [24], and cortical encoding of pulsatile stimulation [25–27]. More recently, GET train has been used to simulate some basic tasks on binaural hearing with CIs, e.g., sound localization [28,29], lateralization [30], binaural masking level differences [31], temporal weighting of ITD and interaural level difference (ILD) [32], effects of electrode place mismatch on binaural cues [33,34], and effects of temporal quantization on ITD discrimination [35].

In signal processing, due to the time–frequency uncertainty principle (also referred to as the Gabor limit), the duration and bandwidth of a signal cannot be independently controlled, and their product is no lower than a limit, which is reachable only by GETs (or say Gabor atoms) [36,15]. This is an important reason why most of the above-mentioned psychoacoustic studies use GETs as stimuli.

However, the performance of GET-based vocoders in simulating speech perception with CIs has not been investigated by other labs. The use of GET trains as a replacement for the carriers typically employed in conventional channel vocoders was first proposed in 2007 in [37], marking the initial introduction of this technique. In our recent pilot studies [38–40], a GET vocoder was implemented directly by mapping individual electric pulses from a clinical *n*-of-*m* strategy with 900-pps pulse rate into an acoustic GET. In this way, any CI electrodogram (not limited to the selected strategy) can be directly transformed into a vocoded sound. Such direct transformation can simulate not only pulsatile timing cues but also many other features of CI electric stimuli (e.g., amplitude compression and maxima selection).

This study detailedly introduces the theory of this novel GET vocoder and further demonstrates its potential for simulating CI speech perception. The pulsatile GET vocoder can replicate the temporal (pulsatile), intensity (compressed and quantized), and spectral (maxima-selected) features of an actual CI strategy. We recommend that the GET vocoder could be an alternative vocoder model to simulate speech perception with CIs. Nevertheless, the uncertainty principle imposes unavoidable physical constraints on the time–frequency tradeoff, which might limit the performance of the pulsatile simulation and should be carefully controlled.

2. GET Theory and Vocoder Algorithms

2.1. GET Theory

A Gaussian function is symmetrical in the time domain:

$$g_{\mathrm{env}}(t) = a\mathrm{e}^{-\frac{\pi(t-t_0)^2}{2\sigma^2}} \tag{1}$$

where a determines the function's maximum amplitude, t_0 the maximum amplitude's temporal position, and σ the effective duration or $D = \sqrt{2}\sigma$, at which the amplitude is −6.82 dB down from the maximum amplitude [18]. Its Fourier transform is:

$$G_{\mathrm{env}}(f) = \sqrt{2}a\sigma \cdot \mathrm{e}^{-2\pi(\sigma f)^2} \cdot \mathrm{e}^{-j2\pi f t_0}. \tag{2}$$

The shape of its amplitude spectrum is also a Gaussian function with an effective bandwidth between the −6.82 dB down cutoff frequencies.

The effective duration D and the effective bandwidth B can be traded:

$$D \cdot B = 1 \tag{3}$$

meaning that increasing the duration will narrow the bandwidth and vice versa.

Acoustic simulation of a single electric pulse in a frequency channel can be generated by multiplying the above Gaussian function by a sinusoidal carrier:

$$s(t) = g_{\mathrm{env}}(t) \cdot \sin(2\pi f_c t + \phi_0) = a\mathrm{e}^{-\frac{\pi(t-t_0)\cdot 2}{2\sigma^2}} \cdot \sin(2\pi f_c t + \phi_0) \tag{4}$$

where $s(t)$ has the same effective duration and effective bandwidth as $g_{\mathrm{env}}(t)$ except for changing the center frequency from 0 to f_c, and ϕ_0 is an initial phase.

Fig. 1 illustrates both waveform (c) and spectrum (d) of a unit-amplitude Gaussian-enveloped tone (i.e., $a = 1$ in (4)). It is used to simulate a single electric pulse (a) at the third site of the electrode array (b). The carrier frequency f_c is 5 kHz. The timing and stimulation place of individual electric pulse could be simulated by the timing and the carrier frequency of the Gaussian-enveloped tone respectively. The −6.82 dB cutoff point (corresponding to $D = \sqrt{2}\sigma$) with an amplitude of 0.456 in Fig. 1 was derived by substituting $t = t_1 = \frac{D}{2} + t_0 = \frac{\sqrt{2}}{2}\sigma + t_0$ into (1), i.e.,

$$g_{\mathrm{env}}(t_1) = \mathrm{e}^{-\frac{\pi(\frac{\sqrt{2}}{2}\sigma)^2}{2\sigma^2}} = \mathrm{e}^{-\frac{\pi}{4}} \approx 0.456. \tag{5}$$

Using the GET defined by (4), the change of amplitude and timing of an electric pulse can be simulated by manipulating a and t_0 respectively. Acoustic simulation of a continuous electric pulse train can

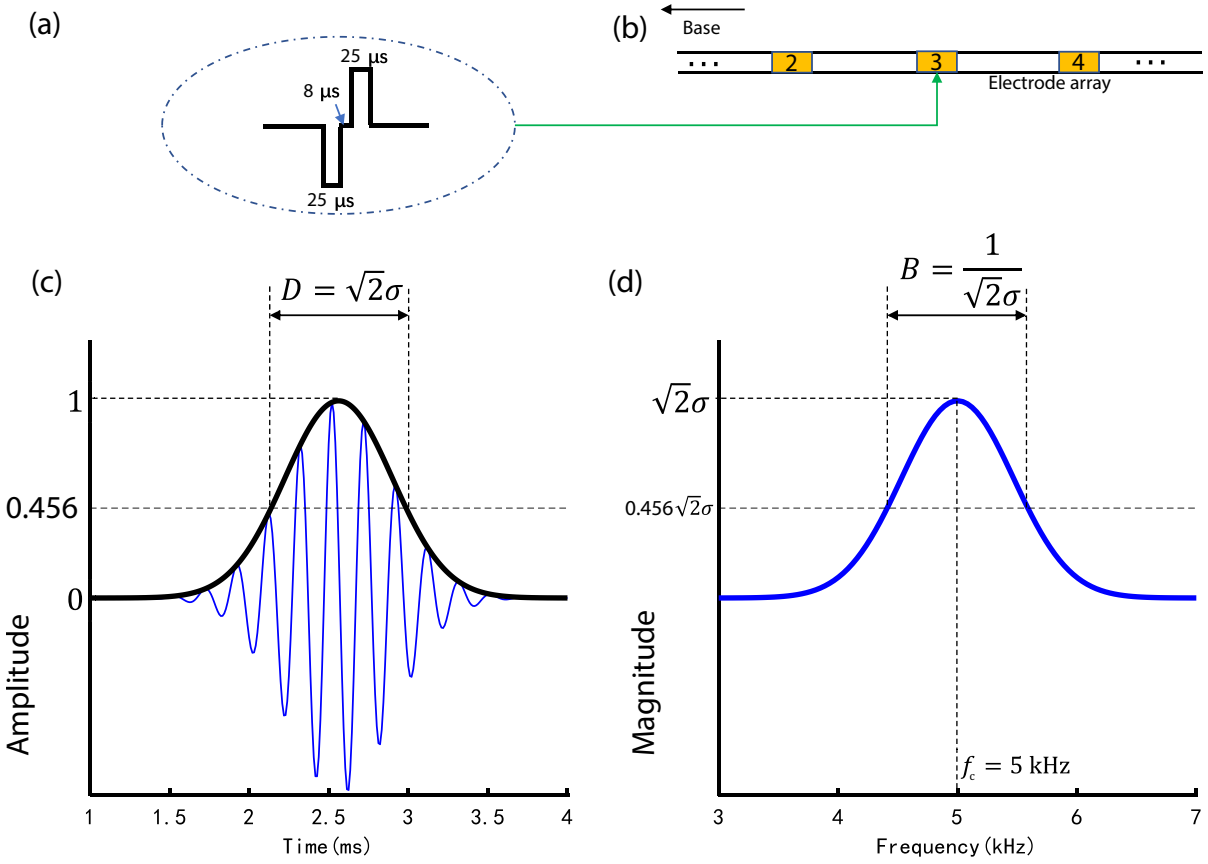

Fig. 1. A biphasic electric pulse (a) stimulating at the third electrode site (b) and a corresponding unit-amplitude single pulse with Gaussian-shaped envelope (black line) in both the time (c) and frequency (d) domains. The carrier frequency is 5 kHz (the blue waveform in the left panel and the frequency with maximum amplitude in the right panel). The σ equals to $3/f_c = 0.6$ ms in (1), producing an effective duration of 0.85 ms and an effective bandwidth of 1.2 kHz. At the time domain, (a) and (c) are depicted, while (b) and (d) are depicted at the frequency (or place) domain.

be constructed by periodically repeating $s(t)$ or convolution of the electric pulse train and a GET.

Different from the CI electric pulses with constant duration at the order of tens of microseconds, the GET duration should be much longer to contain at least several (l) periods (e.g., $l = 2, 3$, or 4) of the tone carrier. Therefore, the carrier period or frequency will determine the lower limits of the GET duration. The three lines in the two panels of Fig. 2 illustrate the dependent relationship between the GET duration (bandwidth), pulse rate, and carrier frequency, when $\sigma = l/f_c = 2/f_c, 3/f_c$, or $4/f_c$, respectively. The GET effective bandwidth equals in value to the maximum pulse rate that can be transmitted without obvious temporal interaction between neighboring GETs. Here the GET duration threshold for the "obvious temporal interaction" was defined as the effective duration of GET, i.e., $D = \sqrt{2}\sigma$. Increasing the duration (i.e., larger σ) can decrease the bandwidth with the maximum rate decreasing correspondingly.

At frequency bands with high carrier frequencies above 2.5 kHz ($f_c = \frac{l}{\sigma} = l\sqrt{2}B = \sqrt{2} \cdot 900l \approx 2546, 3818$, and 5091 Hz for l = 2, 3, and 4, respectively), a conventional pulse rate of 900 pps could be simulated without obvious temporal interaction between neighboring GETs. For carrier frequencies within the middle-frequency range around 2 kHz, the 900 pps is still possible to simulate, but neighboring GETs have moderate temporal interaction. The amplitude of the crossing point of neighboring GETs at a 2 kHz carrier would be

$$20\log_{10} e^{-\frac{\pi(t-t_0)^2}{2\sigma^2}} = 20\log_{10} e^{-\frac{\pi(\frac{1}{2\times 900})^2}{2(\frac{l}{f_c})^2}} = 20\log_{10} e^{-\frac{\pi}{2}(\frac{10}{9l})^2} \quad (6)$$

whose values are $-4.21, -1.87$, and -1.05 dB (relative to the maximum amplitude) for l = 2, 3, and 4, respectively. For a low-frequency carrier, the pulsatile feature for simulation of individual electric pulses cannot be guaranteed due to temporal interactions between neighboring GETs.

The temporal envelopes delivered in electric speech stimuli are often temporally separated across channels in many CI strategies, as speech contains natural gaps within each channel of signal between syllables, and frame-wise low power bands are temporarily abandoned resulting from the maxima selection for n-of-m strategies. Additionally, envelope energies lower than the threshold level (or T level) are not represented in electric stimuli (i.e., no stimulation) in some strategies. For the temporally separated electric stimuli within each channel, GET carriers can better represent temporal separation features as well as CI compression (limited electric dynamic range), both of which are often omitted in conventional noise and sine-wave vocoders. The temporal separation features may be simulated in all channels, and the low carrier frequency limit $f_{c,low}$ is mainly determined by the duration d_{gap} of each gap in the pulse trains:

$$f_{c,low} = \frac{l}{\sigma} = \frac{\sqrt{2}l}{D_{max}} = \frac{\sqrt{2}l}{d_{gap}} \quad (7)$$

where D_{max} is the maximum possible GET duration, which equals the gap duration.

Current (or spectral) spread was acknowledged to be an important issue influencing the frequency resolution of CIs [41]. For a single GET (defined by (4)), its bandwidth is determined by its duration due to the time–frequency uncertainty principle. There-

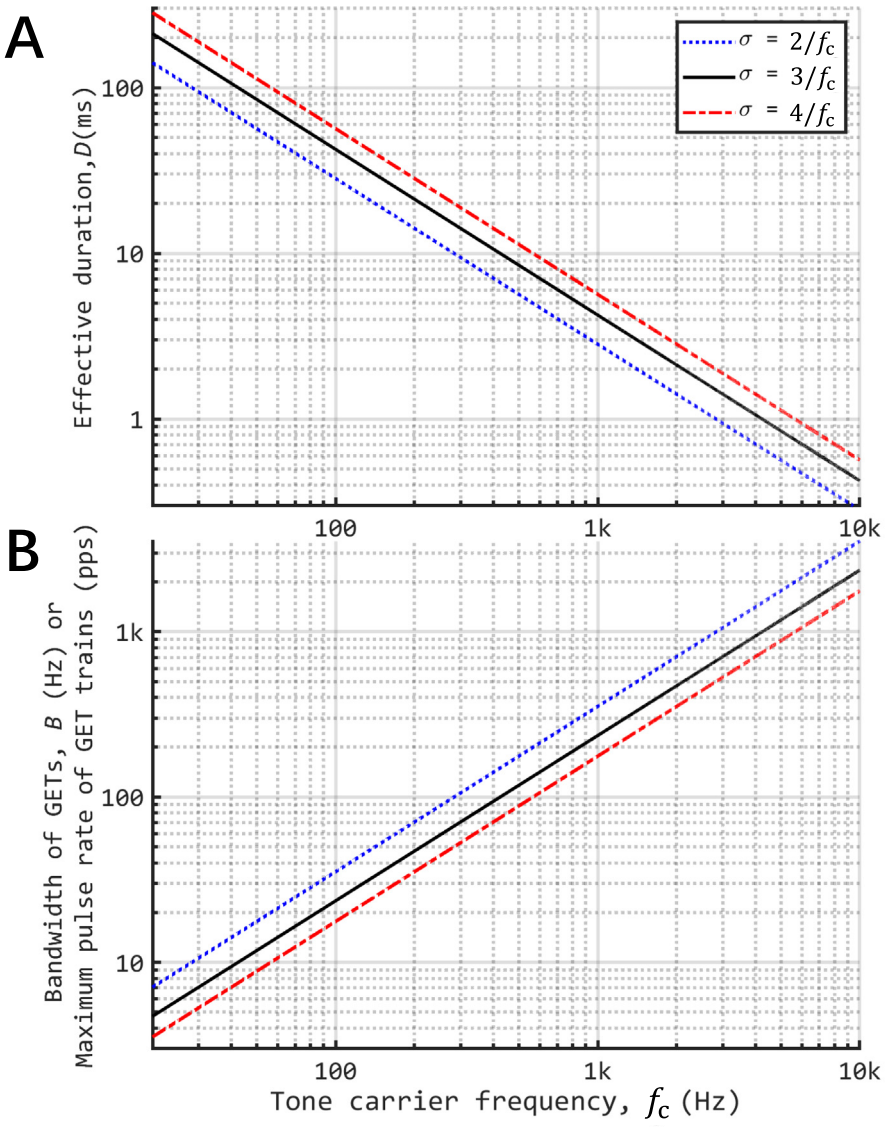

Fig. 2. The relationship between the tone carrier frequency and the effective duration $D = \sqrt{2}\sigma$ (see Panel A) or effective bandwidth $B = 1/D$ (see Panel B) of Gaussian-enveloped tones (GETs). All axes are logarithmically scaled. The σ was assumed to be $2/f_c$, $3/f_c$, or $4/f_c$ to demonstrate the effects of different duration of GETs. For certain combinations of f_c and σ, the maximum GET rate that can be transmitted with no temporal interaction between neighboring GETs is $1/D$, which equals in value to the effective bandwidth in Panel B.

fore, it is possible to simulate CI current spread by manipulating the GET duration, meaning the pulsatile timing feature and the current spread cannot be independently manipulated.

In short, the GETs can simulate and manipulate five important parameters of CI processing or stimulation: (1) pulse rate by changing the period of pulse generation, (2) temporal envelope (including its compression and quantization) by changing the amplitude of individual GETs in a pulse train within a channel, (3) spectral envelope by changing the GET amplitude across channels, (4) place of excitation by changing the carrier tone frequency, and (5) spread of excitation by changing the effective bandwidth in GETs. Fig. 1 provides an GET simulation example of a single electric pulse, demonstrating the ability to adjust both the amplitude and timing of the electric pulse (a) and GET (b) while also changing the stimulation site or carrier frequency to alter the location of the stimulation. As the figure reveals, GET has a significantly longer duration than electric pulses, and a balance must be struck between duration and bandwidth. By skillfully manipulating these five parameters through adjustments in amplitude, timing, duration, and carrier frequency, it is possible to create an acoustic simulation of contemporary CIs using pulsatile electric stimulation.The limitations from the dependent relationships between duration, bandwidth, and carrier frequency of GETs are discussed above and should be taken into consideration during algorithm design and experiments of CI simulations with GETs.

2.2. Vocoder Algorithm Framworks

Fig. 3A shows the conventional acoustic simulation of CI using either noise [8] or sine-wave vocoders [9]. The output filters can be used to control the current spread, but no temporal separation feature (e.g., pulsatile timing and temporally separated envelope) can be simulated.

The first GET vocoder was proposed by [37] (see Fig. 3B) and subsequently used in a sound localization study [28]. As a naïve implementation, this approach replaces the conventional continuous carriers with pulsatile GET carriers. The GET vocoder in Fig. 3C was recently proposed by [38]. Compared to the naïve implemen-

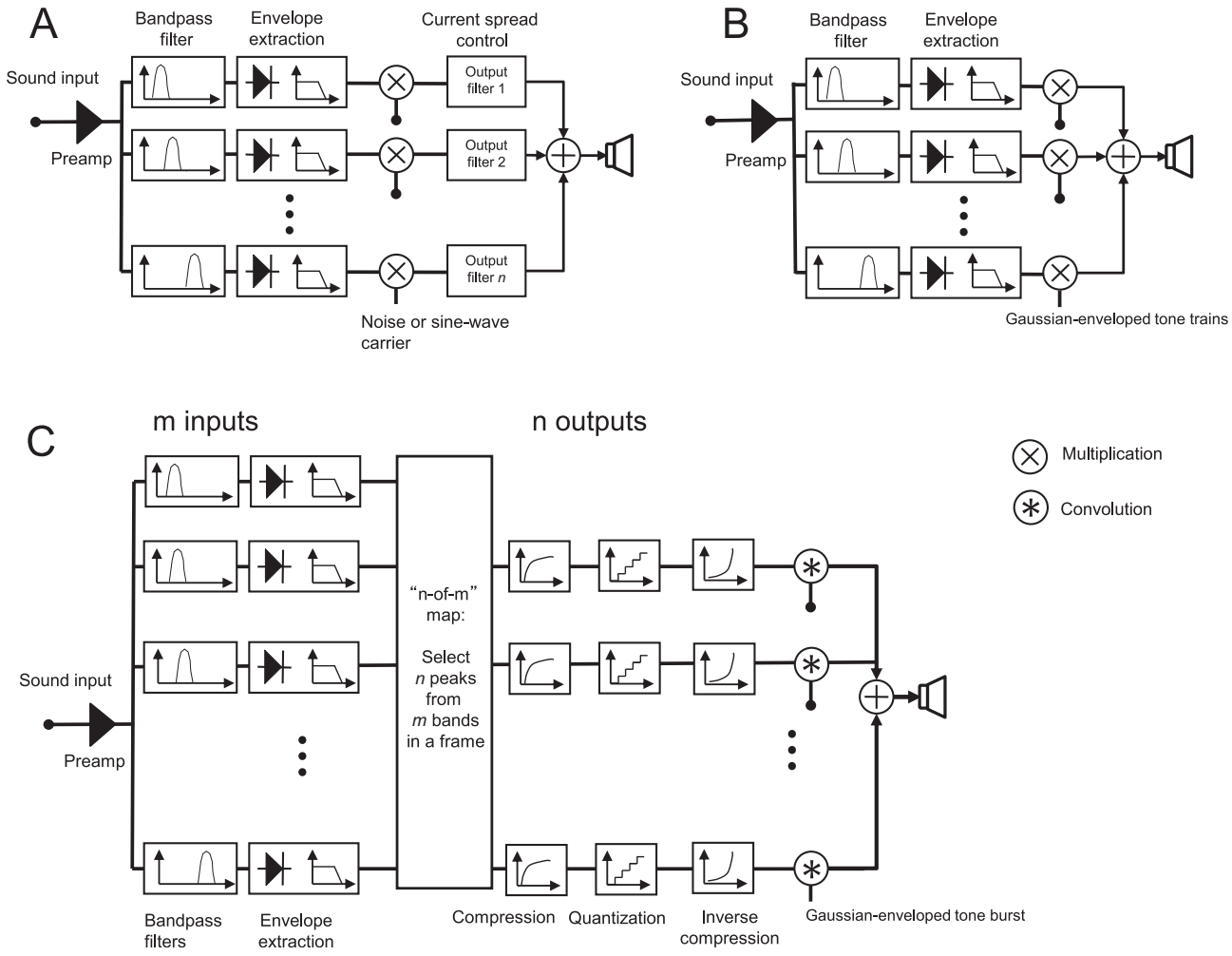

Fig. 3. Block diagrams of conventional channel vocoder (A), the first (B) and second (C) types of GET vocoders. The pulsatile vocoders are using GETs as carriers or using a single GET as an impulse response (in current study). The front-end pre-emphasis, bandpass filter, and envelope extraction can be implemented either in the temporal or spectral domain.

tation of the first type, the new GET vocoder hypothesized that a direct mapping from individual CI electric pulses to individual GET acoustic pulses could transmit similar speech information in both stimuli of CI and GET simulation. The implementation framework of the second GET vocoder considers a common feature of temporal-frame-based *n*-of-*m* selection in some CI processing strategies. The *n*-of-*m* selection means *n* maximum envelope values are selected out of the envelope values from the *m* input channels within a given time window. In this framework, the amplitude compression and quantization widely used in modern CIs can also be simulated. In this study, sentence intelligibility tests were carried out to demonstrate the feasibility of GET simulation on speech perception with the advanced combination encoder (ACE) strategy, which is a typical *n*-of-*m* strategy and has a default pulse rate of 900 pps per channel. More details about the implementations of the GET vocoder are provided in the following section. A MATLAB program is also open for free use in academic research.

3. ACE-GET Implementation and Experiment

3.1. Vocoder Theory: Direct mapping from electric pulses to GETs

In theory, the GETs are applicable for directly transferring any pulsatile CI electrodogram to a pulsatile vocoded sound. To be more illustrative, Fig. 4A demonstrates a 10-channel electrodogram (note: single vertical lines were used to represent electric pulses so that the amplitude and timing of the electric pulse can be represented, while the phase and gap durations in the common bi-phasic electric pulses were not considered in this study). To generate a GET vocoder, the 10 channels were converted into frequency bands spanning over 10 equally divided parts of the basilar membrane between characteristic frequencies of 150 and 8000 Hz [42]. The cutoff frequencies are 150, 271, 439, 672, 994, 1439, 2057, 2911, 4094, 5732, and 8000 Hz. Then, a band-specific GET was generated in this demonstration by setting the parameters in (1) as $a = 1, t_0 = 0$, and $\sigma = \frac{2}{f_c}$, where f_c denotes the center frequency of the specific band. As a result, the band-specific GET had a −6.82 dB duration of

$$D = \sqrt{2}\sigma = \frac{2\sqrt{2}}{f_c} \tag{8}$$

and a −6.82 dB bandwidth of

$$B = \frac{1}{D} = \frac{\sqrt{2}}{4} f_c \tag{9}$$

Then the acoustic GET train at the *k*-th channel in Fig. 4B is derived by

$$p_{a,k}(t) = (p_{e,k}(t) * e^{-\frac{\pi t^2}{2\sigma^2}}) \cdot \sin(2\pi f_c t + \phi_0) \tag{10}$$

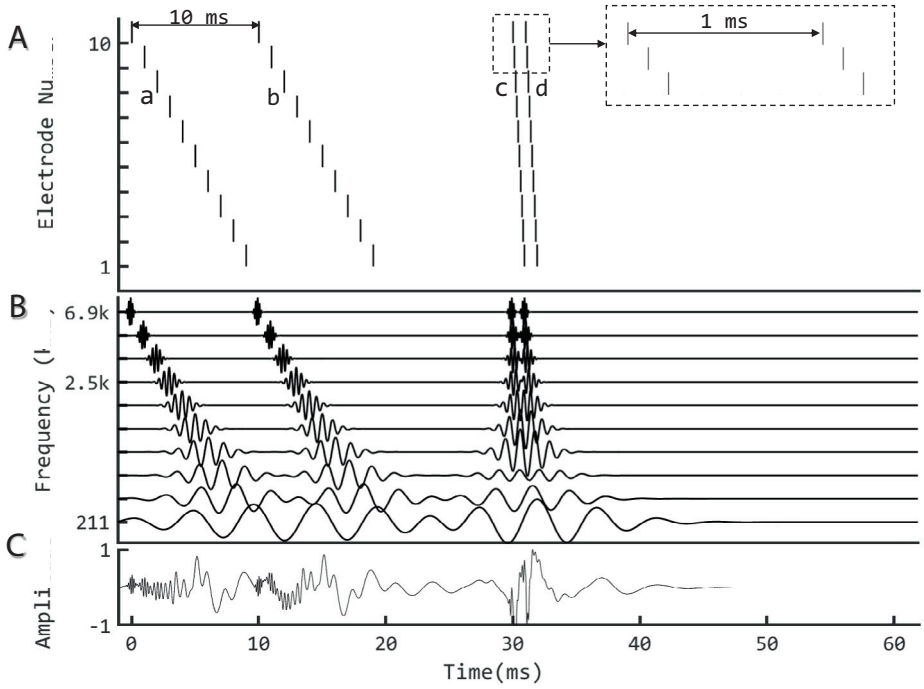

Fig. 4. Mapping a CI electrodogram to a sound using the second type GET vocoder. A. An artificial 10-channel CI electrodogram, including two pulse sweeps with a 10-ms difference between a and b, as well as two additional sweeps with a 1-ms difference between c and d, corresponding to stimulation rates of 100 pps and 1000 pps, respectively. B. GETs mimicking the electric pulse trains. C. The final GET waveform resulting from the sum of ten band-specific GET trains in B.

where $p_{e,k}(t)$ and $p_{a,k}(t)$ denote the electric and acoustic pulse trains in Fig. 4A and 4B, respectively, "$*$" denotes a convolution, σ and f_c are band-dependent parameters as defined above, and ϕ_0 is an initial phase that could be arbitrarily defined and was uniformly randomized between 0 and 2π here.

Fig. 4B shows the 10-channel GET trains, which have temporally separated waveforms for high-frequency channels, but overlapping waveforms for low-frequency channels. Fig. 4C shows the overall waveform summed from the 10 bands.

According to the theoretical analysis of GET simulation, pulsatile features for individual electric pulses cannot be guaranteed in the low-frequency channels, but the temporal-separation feature between groups of pulses may be simulated to some extent. For example, in Fig. 4B, at the lowest frequency channel, the 12-ms gap between b and c sweeps could have a counterpart, i.e., a shallow amplitude-modulation dip, in the waveform.

3.2. Experiment method: Simulation of the n-of-m strategy ACE

By employing the previously described method, it is possible to convert any electrodograms into vocoded sounds. This includes the commonly used *n*-of-*m* strategy, such as the ACE strategy, which currently serves as the default strategy in Nucleus cochlear implants [43]. The specific vocoder is named ACE-GET. Following the preliminary results which showed comparable acute data between the ACE-GET vocoder and actual CI users [39], in this paper a battery of speech recognition tasks was carried out to further explore the potential of ACE-GET vocoder on simulation of speech perception with CIs.

In the clinical fitting of CIs with the ACE strategy, the dynamic range should be measured behaviorally electrode-by-electrode and is also limited and variable among users. In the ACE-GET vocoders, the dynamic range could be easily manipulated either in the compression stage of the ACE encoding or in the inverse compression stage of the GET synthesizing. The latter method was used in this study, and two dynamic ranges corresponding to two ACE-GET vocoders were tested. It was hypothesized that the vocoder with a higher dynamic range would simulate the top CI participants while the vocoder with a lower dynamic range would simulate the average performance of CI participants. The combination of n = 8 and m = 22 is one default option in the clinical fitting of ACE and was simulated in this experiment.

In detail, two 22-channel ACE-GET vocoders (denoted by GETlargeDR and GETsmallDR) were compared with two 22-channel sine-carrier conventional vocoders (125 Hz and 250 Hz envelope cutoff, denoted by Sin250 and Sin125, respectively) with minimum channel overlapping as shown in Fig. 3A. The hypotheses for the parameter selection of the four vocoders are discussed later.

Detailed implementation methods of the *vocoders*: First, the default setting of the ACE software integrated in the CCi-Mobile software [44] was used to convert input sounds into electrodograms. An inverse-mapping function was used to transfer the electric current of each electric pulse in the electrodogram to an envelope power. The implementation of GET involved simplifying each electric pulse in the electrodogram as a single line. The height of this line was determined by the constant current level of the anodic phase of the corresponding bi-phasic pulse. This line is referred to as a single sample of the electric pulse. Single-sample pulse trains from each band were "convolved" with a Gaussian function with $\sigma = 3/f_c$. σ is an arbitrary parameter (e.g., 2 in Section 3.1 and 3 here). Larger σ results in longer duration and narrower bandwidth. In the specific implementation of the experiment, the convolution step was replaced by simply comparing any overlapping sampling points from two GETs and preserving the larger point as the final sample value.[1] The output was used to

[1] In the theory and framework analysis in Section 2, a convolution calculation was recommended, but in our experiment, we only preserved the largest point to show better pulsatile waveform than the cumulative effect of a convolution. Based on the findings of our pilot test, either there is no discernible difference or only a negligible fluctuation in the speech intelligibility of the output speech between the two ways.

multiply a sinusoidal carrier with a frequency of f_c at the center of the corresponding band and an arbitrary initial phase (a random initial phase in this study). The average power of each band was kept unchanged. Finally, the modulated signals were summed to produce the vocoded stimulus.

The difference between GETlargeDR and GETsmallDR was only between their inverse (i.e., electric-to-acoustic) mapping functions, which are (11) and (12), respectively:

$$L_a = \frac{1}{\alpha}((1+\alpha)^{L_e} - 1) \tag{11}$$

and

$$L_a = \frac{1}{2.72\alpha}(e^{L_e}(1+\alpha) - 1) \tag{12}$$

in which, the L_a denotes the recovered acoustic level, L_e denotes the electric current level defined by the electrodogram from the ACE strategy based on a specific patient's fitting map, and α is a constant 416.0 which is a default value in the ACE program provided by CCi-Mobile software [44]. In the present study, the threshold levels and most comfortable levels are constantly defined as 100 and 255 CU (current unit), i.e., 100 CU $< L_e <$ 255 CU. In this case, based on (11) and (12), the recovered acoustic level ranges were 32.7 dB and 5.3 dB for GETlargeDR and GETsmallDR, respectively. The output stimuli level was controlled at a comfortable level around 65 dBA ($L_{AF,eq}$, or acoustic time-equivalent A-weighted sound pressure level integrated at 125 ms intervals). Eq. (11) is directly based on the default setting of the acoustic-to-electric compression function in ACE. It was hypothesized that GETlargeDR could simulate the best performance of CI listeners with the corresponding ACE strategy and GETsmallDR would significantly degrade the performance because of the much narrower range. Otherwise, the implementation details of the vocoder were the same as in [38].

In the two sine vocoders, the frequency spacing for cutoffs for the analysis filters was defined in the range of [80,7999] Hz according to a Greenwood map [42]. Specifically, the cutoff frequencies were 80, 122, 172, 230, 298, 379, 473, 583, 712, 864, 1042, 1250, 1494, 1781, 2117, 2512, 2974, 3516, 4152, 4898, 5772, 6797, and 7999 Hz. The filtered signals were full-wave rectified and low-pass filtered (6th order Butterworth; 125 Hz for Sin125 and 250 Hz for Sin250) to extract the envelope for each channel. A sine wave with a frequency centered at the corresponding analysis band was used as the carrier, which was then multiplied by the corresponding envelope. The final vocoded stimuli were generated by a summation of the modulated carriers. In previous studies, it was found that speech intelligibility was better with a higher cutoff frequency in the envelope extraction [45]. Therefore, Sin250 was expected to be better than Sin125.

In Fig. 5, a Mandarin sentence was used to demonstrate the vocoded speech using the four vocoders, i.e., GETlargeDR, GETsmallDR, Sin250, and Sin125. It shows that the GET vocoders resemble the ACE-electrodogram more than the sine vocoders. The temporal separation between groups of pulses can also be found in the band signals of GET vocoded speech. Because the GET vocoders directly use the information of the ACE electrodogram, it was hypothesized that speech intelligibility would be worse, but closer to actual CI results, with the GET vocoders than with the sine vocoders. The reason is that, in the selected sine vocoders, no current spread was additionally included, and no amplitude compression and maxima selection was applied. The right row depicts an enlarged view of a particular area in the figures from the middle row. According to the right row, the GET vocoded signals were able

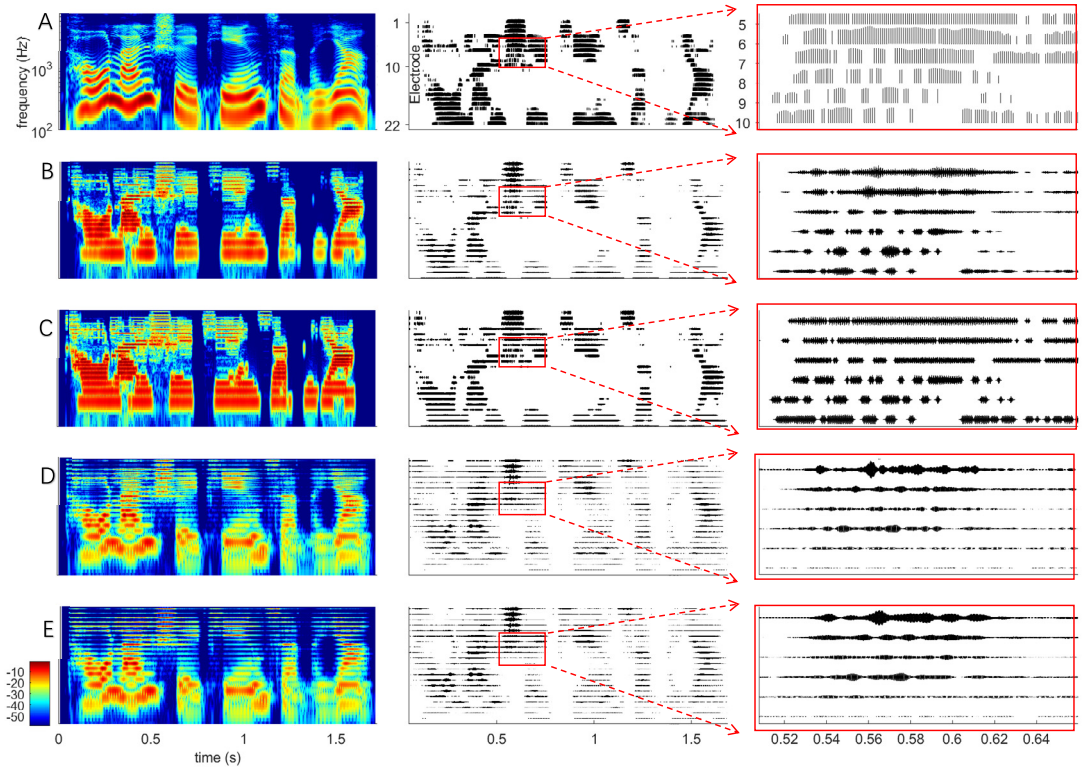

Fig. 5. Speech stimulus demonstrations for the ACE-GET simulation experiment. Left: Spectrogram; middle: band-specific signal; right: a closer view of the signals contained within the boxes. A. Spectrogram and ACE electrodogram of a clear sentence of speech. B-E. Spectrogram and band-specific waveforms of vocoded speech using two GET vocoders (GETlargeDR, and GETsmallDR) and two conventional sine-wave vocoders (Sin250 and Sin125), respectively.

Table 1
Detailed Information of the 7 CI Particpicants in the 3rd task (i.e., Speech in Reverberation Test).

ID	Gender	Age (yr)	CI Experience (yr)	CI Processor	Strategy	Number of Channels	Etiology
C14	F	41	12	CP810	ACE	22	Drug induced
C23	M	31	11	CP810	ACE	22	Sudden deafness
C30	M	13	10	Freedom	ACE	22	LVAS
M5	M	18	15	OPUS-2	FS4-p	12	Virus infection
C16	F	25	2	Freedom	ACE	22	Unknown
M17	F	18	5	OPUS-2	FSP series	12	Genetic
M16	F	17	7	OPUS-2	FSP series	12	Unknown

to preserve some of the pulsatile characteristics in the CI electrodogram, while these features were mostly lost in the sine-wave vocoded signals.

3.3. Experiment method: Participants and Tasks

Two groups of normal-hearing participants (ten in each group, ages 18–29, and native Mandarin speakers) were tested in a soundproof room. An automated pure-tone audiometry (PTA) method [46] was used to confirm that both ears of the individuals had normal hearing thresholds (i.e., equal to or less than 25 dB HL from 125 Hz to 8 kHz)[2]. Group 1 used Sin250 and GETlargeDR, and Group 2 used Sin125 and GETsmallDR. This design was worse than a pure within subject one, but this was a compromise considering the pandemic influence and the purpose (i.e., demonstration but not full validation) of the experiment. One loudspeaker positioned one meter in front of the listener was used to present the vocoded sounds.

To demonstrate GET performance in speech perception in multiple tasks, three open-set Mandarin Chinese recognition tasks were tested, i.e., time-compression threshold, sentence-in-noise recognition, sentence-in-reverberation recognition. The results for the three tasks with the two vocoders in these normal-hearing participants were compared with actual CI results from our previous experiments in [47] as well as newly collected data in this work. These experiments were conducted following procedures approved by the Medical Ethics Committee of Shenzhen University, China. The procedures of the first and second tests followed Experiment 2 of [47] strictly, in which ten CI subjects (9/10 adults) with various hearing histories were tested. Detailed information about the three experiments is as follows:

(1) Time-compression thresholds (TCTs), i.e., accelerated sentence speeds at which 50% of words could be recognized correctly, were measured using the Mandarin speech perception corpus [48]. The sentence was compressed using a synchronized overlap-add, fixed synthesis algorithm which could change the duration with the spectral distribution kept unchanged. An adaptive procedure was used to change the compression ratio in each trial [47].
(2) Speech reception thresholds (SRTs) in speech-shaped noise (SSN) and five-talker Mandarin babble noise, i.e., signal-to-noise ratio (SNR) at which 50% of words could be recognized correctly, were measured using the Mandarin hearing in noise test (MHINT) corpus [49]. An adaptive procedure was used to change the SNR in each trial [47].
(3) Recognition of speech in reverberation was measured using a Mandarin BKB-like sentence corpus [50], whose quiet sentences were convolved with simulated room impulse responses (RIRs). The RIRs were generated using a MATLAB function with its default setting, except the reverberation times (T60) were set as 0, 0.3, 0.6, and 0.9 s (https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator). For each T60, one sentence list was used. Seven CI participants with various hearing histories were also tested for comparison (See Table 1).

We had three subject groups, two of which were normal-hearing listeners each using two different vocoders. A mixed model was used to assess the repeated measures within subjects as well as independent measures between subjects. Paired-sample *t*-test and two-sample *t*-test were used to examine the statistical significance of the means' difference for within-subject comparisons and between-subject comparisons, respectively. For each task, the five CI processing conditions, i.e., Sin250, Sin125, GETlargeDR, GETsmallDR, and CI, were pair-wisely examined to yield 10 pairs of comparison. Bonferroni corrections were used to adjust the p values, and the final significance was examined using the α level of 0.05.

3.4. Results

The results with the four 22-channel vocoders, i.e., GETlargeDR, GETsmallDR, Sin250, and Sin125 are shown in Fig. 6. To provide a benchmark for the simulated results, outcomes from real patients with CI were also incorporated.

For the TCT test (Fig. 6A), a significant decreasing trend was found from Sin250 (mean = 16.1 syllables/s), Sin125 (13.9), GETlargeDR (12.3), GETsmallDR (9.4), to actual CI (6.8) results (Bonferroni adjusted $p < 0.05$), while their standard deviations are comparable within the range from 1.0 to 1.2 syllables/s.

For the SRT test (Fig. 6B; the results are expressed in terms of the SNR measured in decibels (dB)), there was no significant difference (adjusted $p > 0.05$) between Sin250 (means: −4.7 dB in SSN and −0.1 dB in babble noise) and Sin125 (means: −4.8 dB in SSN and −0.1 dB in babble noise) and between GETsmallDR (means: 5.6 dB in SSN and 10 dB in babble noise) and actual CIs (means: 6.5 dB in SSN and 8.8 dB in Babble noise). The mean results with GETlargeDR (means: −1.5 dB in SSN and 4.5 dB in babble noise) were significantly lower (adjusted $p < 0.05$) than those with Sin250 and Sin125, and significantly higher (adjusted $p < 0.05$) than those with GETsmallDR and CIs. The mean SRTs in babble noise were always significantly lower than those in SSN for all four vocoder conditions (adjusted $p < 0.05$). For CI users, mean SRTs in the two noise types did not show a significant difference (adjusted $p > 0.05$).

For the reverberant speech recognition test (Fig. 6C), all vocoders and the actual CI condition showed a significant trend of decreased recognition scores when the reverberation time increased. However, the normal-hearing listeners using sine vocoder simulations were much less sensitive to reverberation than the CI users. It is shown that even with T60 = 0.9 s, the sine vocoders still derived > 94% means, which were much higher than CI partic-

[2] PTA is considered a gold standard method for confirming hearing loss, as it provides important diagnostic information on the degree, type, and configuration of hearing loss. However, the relationship between PTA and speech perception abilities is not straightforward, and PTA alone may not accurately predict an individual's ability to understand speech in challenging listening situations.

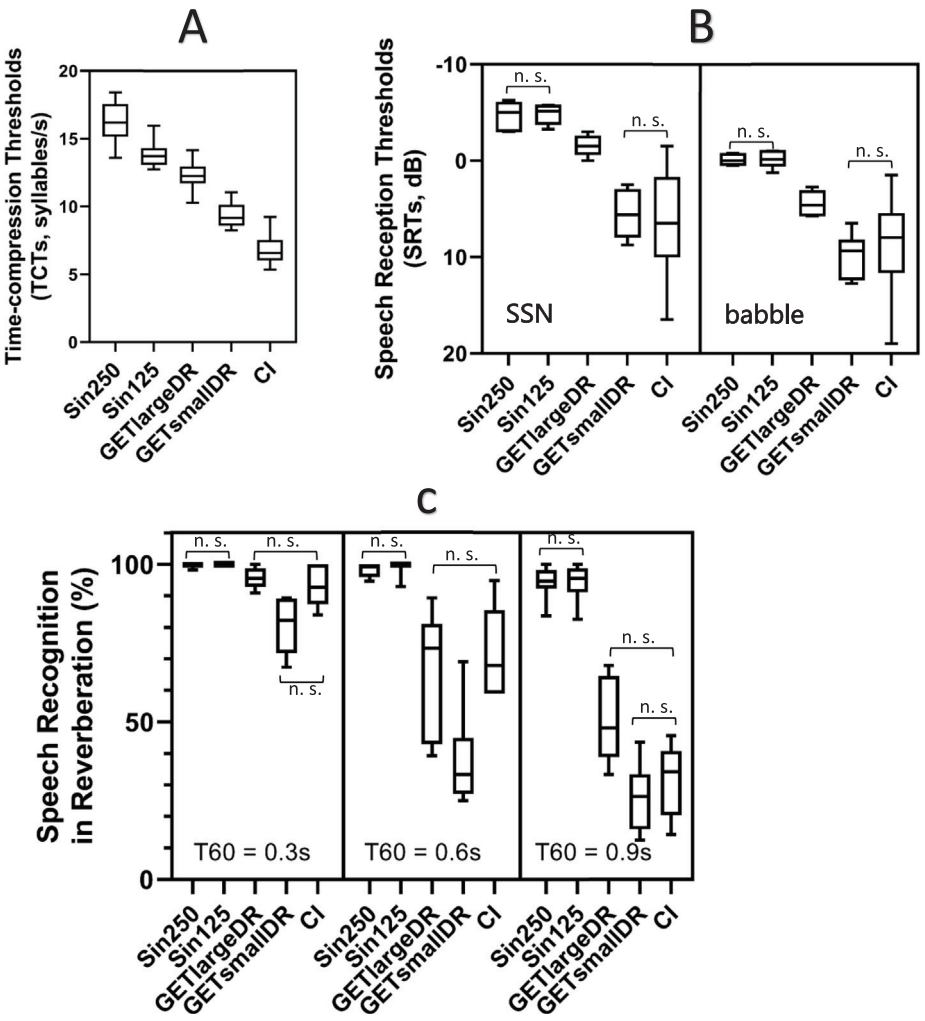

Fig. 6. Results from three speech recognition tasks with two 22-channel sine-wave vocoders (Sin250: 250 Hz cut-off envelope; Sin125: 125 Hz cut-off envelope) and two GET vocoders (GETlargeDR and GETsmallDR; their difference is only in the intensity dynamic range, i.e., 32.7 dB and 5.3 dB for GETlargeDR and GETsmallDR respectively) compared with the results of some CI subjects. There were two groups of normal-hearing participants, each with ten participants. One group used Sin250 and GETlargeDR, and the other group used Sin125 and GETsmallDR. A. Time-compression threshold results. B. Speech reception threshold results of a speech in noise recognition experiment (SSN and babble noise). C. Speech recognition scores in reverberation with T60 = 0.3, 0.6, and 0.9s. Pairwise comparisons with Bonferroni corrections were examined. In each box, "n. s." denotes the non-significant difference ($p > 0.05$), otherwise, there was a significant difference.

ipants' 32%. The GETlargeDR and GETsmallDR derived significantly lower scores than the sine vocoders did (adjusted $p < 0.05$). Under the T60 = 0.3 s and 0.9 s conditions, there was no significant mean score difference between either GET vocoder and CI (adjusted $p > 0.05$), while GETsmallDR derived significantly lower mean scores than GETlargeDR did (adjusted $p > 0.05$). However, the mean results with CI were closer to GETlargeDR at T60 = 0.3 s and to GETsmallDR at T60 = 0.9 s. Under the T60 = 0.6 s condition, there was no significant mean score difference between GETlargeDR and CI, while GETsmallDR derived significantly higher mean scores than GETlargeDR and CI.

In all three tasks, GET vocoders resulted in more similar performance to actual CI performance than sine vocoders did. In fact, the sine vocoders overestimated CI performance in all tasks. Sin250 performed slightly better than Sin125 in mean results but did not show a significant difference. In the time-compression task, all vocoders produced better than CI performance, with GETsmallDR being the closest (Fig. 6A). In the SRT-in-noise test, GETsmallDR and CI produced comparable performance (Fig. 6B). In the reverberation task, GETlargeDR had similar-to-CI performance in all T60 conditions and GETsmallDR in the T60 = 0.3 and 0.9s conditions (Fig. 6C).

4. Discussion

The GET vocoder is proposed to be an alternative model for the simulation of CI strategies. Although directly mapping individual electric pulses into individual acoustic pulses may seem like an intuitive idea in the simulation of cochlear implants (CIs), there have been relatively few studies exploring this approach. In conventional CI simulations, such as those using sine-wave and noise carrier channel vocoders [8,9], temporal envelopes from a limited number of frequency bands are used to modulate the amplitude of continuous carriers, producing a sound intended to convey similar information to a CI strategy. However, naive channel vocoders often lead to overestimation of actual CI performance [51], which is why signal degradation is typically introduced through the use of fewer channels than in reality [52], greater current spread between channels [41,53], or a frequency-to-place mismatch [54]. While the latter two are based on hypotheses regarding the potential influence of current spread and frequency-to-place mismatch on CI performance, the decision to use a smaller number of channels is based purely on the conventional assumption that experiment scores for real and simulated CIs should be comparable. In a specific experiment, speech can be degraded in various feature domains

to produce similar scores between actual and simulated CIs, as speech signals exhibit significant redundancy.

In this study, based on the "intuitive" idea, a GET-based vocoder was theoretically analyzed, and evaluated for its performance in CI speech perception simulation.

4.1. GETs and electric pulses

The GET can be used to simulate a "perceivable" atom of sound, which can be traced back to [15]. More recently, it has been used in many psychoacoustic studies. The GET vocoder model can be a phenomenological one, in which each GET corresponds to an electrical pulse. The amplitude of the GET is scaled proportionally to the pulse current level. Moreover, the analysis part of GET vocoders can simulate main features in CIs, including the place of stimulation, pulse time, temporal envelope, spectral envelope and spectral interaction, and intensity quantization and maxima-selection, by corresponding features of the acoustic pulses. In the synthesis end of GET vocoders subject to the time–frequency uncertainty principle, the pulsatile features may be partially smooth outed at low-frequency channels. In the future, this limitation could potentially be mitigated by utilizing higher carrier frequencies, which would enable the simulation of frequency-to-place (tonotopic) mismatch more effectively.

The principle of time–frequency uncertainty states that the product of a signal's duration and bandwidth cannot be made extremely compact. Specifically, when it comes to acoustic signals, the minimum duration of a signal should encompass several periods of the carrier frequency, which is dependent on the analysis band or characteristic frequency of the targeted stimulation site. In the implementation of our GET vocoder, the GET duration was set to be a function of the carrier frequency. The GET duration was shorter for the higher frequency band. On the other hand, electric pulses have a duration in the range of tens of microseconds, with the frequency or site of stimulation determined by the position of the electrode. The physical constraint on acoustic GET cannot be avoided and must be thoroughly considered during the application of the vocoder. Nevertheless, many key features could be better simulated in GET vocoders than conventional channel vocoders using continuous carriers in theory. In this study, the ACE strategy was simulated by GET. Upon comparing the GET-vocoded spectrogram and ACE-electrodogram, we observed that they share similarities, indicating that comparable information related to intelligibility is being conveyed.

4.2. Speech perception with GET vocoders

[37] proposed a naive impementation of GET vocoder by simply replacing the carriers in conventional channel vocoders by GET trains. In that implementation (i.e., Figure 3B),B a non-interleaved sampling 100-pps GET carrier was generated to study the effects of spread of excitation by controlling the GET duration according to the time–frequency uncertainty principle.

In this study, we further examined another GET vocoder proposed by us in [38] recently. It directly mapped individual electric pulses in a CI electrodogram to individual GETs to simulate the ACE strategy (Fig.3C). This direct mapping allows simulation of all processing steps including the *n*-of-*m* maxima selection to amplitude compression and quantization. Compared with the conventional sine-wave vocoder, not only did the GET vocoder better resemble the ACE electrodogram, but more importantly the GET vocoder, which has a small dynamic range close to that of actual CIs (referred to as GETsmallDR), produced a mean and range of speech in noise recognition performance similar to that of actual CI users. Together with the other pilot applications of the GET vocoder in [39] and [40], this work established several pieces of evidence to support the GET vocoder as an alternative to the conventional vocoders.

GETsmallDR produced both better and worse results than actual CI users in the TCT and speech in reverberation tasks, respectively. The reason for this discrepancy is not well understood. The TCT results tested using unreal fast speech suggested that normal-hearing listeners had better temporal processing with GET than with CIs. However, real CI users may have more experience with degraded speech in daily life, which could account for their superior performance in perceiving reverberated speech compared to GET simulations. Furthermore, the larger dynamic range of GET simulations yielded better CI performance than the smaller dynamic range, as illustrated in Fig. 6. The three tasks illustrated the complex association between GET simulation and actual CI outcomes. Furthermore, it is important to note that the training of vocoded sound in normal-hearing listeners is restricted, and this factor could have a substantial impact on the outcomes. Therefore, this training limitation should be taken into account in future studies.

Future studies are also needed to establish and evaluate individualized CI simulation, in which both the mean and error patterns of phonemic recognition are used to judge the validity and quality of the simulation model [55]. A vocoder showed smaller real-simulation discrepancy in multiple tasks than other vocoders would be identified as an optimal vocoder.

4.3. GET vocoders as a general model

The GET vocoder is perhaps a more general vocoder model as it can closely approximate conventional noise (using noise carriers instead of sine waves) and sine-wave vocoders by summing many GETs occurring at high rates or long GET duration and using high-fidelity intensity (or envelope) information. This means that the conventional vocoders can be treated as special cases of GET vocoders. Further experiment studies (e.g., in phoneme confusion patterns) are warranted to examine the performance of GET simulation systematically. The MATLAB source code of the GET vocoder for the ACE strategy is provided for academic research purposes (https://github.com/BetterCI/GETVocoder). Based on this code, more variants could be generated by manipulating the vocoder parameters, e.g., spectral spread, stimulation place or frequency shifting, and carrier types.

5. Conclusions

(1) The time–frequency uncertainty principle imposes constraints on using GETs for CI simulation. Using a shorter duration for GET in a specific band will lead to a broader bandwidth, mimicking a greater current distribution and reducing temporal interference among neighboring GETs. The opposite is also true.
(2) Many features of modern CIs including pulsatile timing, current spread, *n*-of-*m* maxima selection, and dynamic compression could be implemented in GET vocoders and then used to derive similar sentence-in-noise recognition performance to actual CI users. It is necessary to conduct additional experiments to assess the effectiveness of GET in various speech-related tasks such as phoneme recognition and reverberant speech perception, taking into account the aspect of training.
(3) A GET vocoder framework for arbitrary CI strategy and a package of source code (using ACE as an example) are provided to serve as a general-purpose research tool to generate vocoded sounds (including speech) based on the direct pulse-to-pulse mapping.

CRediT authorship contribution statement

Qinglin Meng: Conceptualization, Methodology, Software, Investigation, Writing - original draft, Writing - review & editing. **Huali Zhou:** Software, Methodology, Writing - original draft. **Thomas Lu:** Writing – original draft. **Fan-Gang Zeng:** Conceptualization, Writing – review & editing.

Data availability

Data will be made available on request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We thank all the participants in these experiments. Fanhui Kong and Yulong Xiao helped collect the data in experiment. Thanks to Drew Cappotto for proof-reading a previous version of this article.

This research was supported by NIH R01 DC15587 (F.G.Z.), National Natural Science Foundation of China (11704129 and 61771320), Guangdong Basic and Applied Basic Research Foundation Grant (2020A1515010386), and Science and Technology Program of Guangzhou (202102020944) (Q.M.). Language polishing assisted by OpenAI's ChatGPT in the Discussion of the paper during revision.

References

[1] Dudley H. Remaking speech. J Acoust Soc Am 1939;11:169–77.
[2] Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM. Better speech recognition with cochlear implants. Nature 1991;352:236–8.
[3] Zeng F-G, Rebscher S, Harrison W, Sun X, Feng H. Cochlear implants: system design, integration, and evaluation. IEEE Rev Biomed Eng 2008;1:115–42.
[4] Skinner MW, Holden LK, Whitford LA, Plant KL, Psarros C, Holden TA. Speech recognition with the nucleus 24 speak, ace, and cis speech coding strategies in newly implanted adults. Ear Hearing 2002;23:207–23.
[5] Meng Q, Zheng N, Li X. Loudness contour can influence mandarin tone recognition: Vocoder simulation and cochlear implants. IEEE Trans Neural Systems Rehab Eng 2017;25:641–9.
[6] Singh S, Kong Y-Y, Zeng F-G. Cochlear implant melody recognition as a function of melody frequency range, harmonicity, and number of electrodes. Ear Hearing 2009;30:160.
[7] Xu L, Thompson CS, Pfingst BE. Relative contributions of spectral and temporal cues for phoneme recognition. J Acoust Soc Am 2005;117:3255–67.
[8] Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science 1995;270:303–4.
[9] 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-band outputs. J Acoust Soc Am 1997;102:2403–11.
[10] Blamey P, Dowell R, Tong Y, Clark G. An acoustic model of a multiple-channel cochlear implant. J Acoust Soc Am 1984;76:97–103.
[11] Blamey P, Dowell R, Tong Y, Brown A, Luscombe S, Clark G. Speech processing studies using an acoustic model of a multiple-channel cochlear implant. J Acoust Soc Am 1984;76:104–10.
[12] Deeks JM, Carlyon RP. Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation. J Acoust Soc Am 2004;115:1736–46.
[13] Hilkhuysen G, Macherey O. Optimizing pulse-spreading harmonic complexes to minimize intrinsic modulations after auditory filtering. J Acoust Soc Am 2014;136:1281–94.
[14] Mesnildrey Q, Hilkhuysen G, Macherey O. Pulse-spreading harmonic complex as an alternative carrier for vocoder simulations of cochlear implants. J Acoust Soc Am 2016;139:986–91.
[15] Gabor D. Acoustical quanta and the theory of hearing. Nature 1947;159:591–4.
[16] Schneider BA, Pichora-Fuller MK, Kowalchuk D, Lamb M. Gap detection and the precedence effect in young and old adults. J Acoust Soc Am 1994;95:980–91.
[17] Trehub SE, Schneider BA, Henderson JL. Gap detection in infants, children, and adults. J Acoust Soc Am 1995;98:2532–41.
[18] Baer T, Moore BC, Glasberg BR. Detection and intensity discrimination of gaussian-shaped tone pulses as a function of duration. J Acoust Soc Am 1999;106:1907–16.
[19] Baer T, Moore BC, Marriage J. Detection and intensity discrimination of brief tones as a function of duration by hearing-impaired listeners. Hearing Res 2001;159:74–84.
[20] Nizami L, Reimer JF, Jesteadt W. The intensity-difference limen for gaussian-enveloped stimuli as a function of level: tones and broadband noise. J Acoust Soc Am 2001;110:2505–15.
[21] van Schijndel NH, Houtgast T, Festen JM. Intensity discrimination of gaussian-windowed tones: Indications for the shape of the auditory frequency-time window. J Acoust Soc Am 1999;105:3425–35.
[22] Laback B, Zimmermann I, Majdak P, Baumgartner W-D, Pok S-M. Effects of envelope shape on interaural envelope delay sensitivity in acoustic and electric hearing. J Acoust Soc Am 2011;130:1515–29.
[23] Laback B, Necciari T, Balazs P, Savel S, Ystad S. Simultaneous masking additivity for short gaussian-shaped tones: Spectral effects. J Acoust Soc Am 2013;134:1160–71.
[24] Buell TN, Hafter ER. Discrimination of interaural differences of time in the envelopes of high-frequency signals: Integration times. J Acoust Soc Am 1988;84:2063–6.
[25] Johnson LA, Della Santina CC, Wang X. Representations of time-varying cochlear implant stimulation in auditory cortex of awake marmosets (callithrix jacchus). J Neurosci 2017;37:7008–22.
[26] Lu T, Wang X. Temporal discharge patterns evoked by rapid sequences of wide-and narrowband clicks in the primary auditory cortex of cat. J Neurophysiol 2000;84:236–46.
[27] Lu T, Liang L, Wang X. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nature Neurosci 2001;4:1131–8.
[28] Goupell MJ, Majdak P, Laback B. Median-plane sound localization as a function of the number of spectral channels using a channel vocoder. J Acoust Soc Am 2010;127:990–1001.
[29] Jones H, Kan A, Litovsky RY. Comparing sound localization deficits in bilateral cochlear-implant users and vocoder simulations with normal-hearing listeners. Trends Hearing 2014;18. 2331216514554574.
[30] Ehlers E, Kan A, Winn MB, Stoelb C, Litovsky RY. Binaural hearing in children using gaussian enveloped and transposed tones. J Acoust Soc Am 2016;139:1724–33.
[31] Lu T, Litovsky R, Zeng F-G. Binaural masking level differences in actual and simulated bilateral cochlear implant listeners. J Acoust Soc Am 2010;127:1479–90.
[32] Brown AD, Stecker GC. Temporal weighting of interaural time and level differences in high-rate click trains. J Acoust Soc Am 2010;128:332–41.
[33] Goupell MJ, Stoelb C, Kan A, Litovsky RY. Effect of mismatched place-of-stimulation on the salience of binaural cues in conditions that simulate bilateral cochlear-implant listening. J Acoust Soc Am 2013;133:2272–87.
[34] Kan A, Stoelb C, Litovsky RY, Goupell MJ. Effect of mismatched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users. J Acoust Soc Am 2013;134:2923–36.
[35] Dieudonné B, Van Wilderode M, Francart T. Temporal quantization deteriorates the discrimination of interaural time differences. J Acoust Soc Am 2020;148:815–28.
[36] Feichtinger HG, Strohmer T. Gabor analysis and algorithms: Theory and applications. Springer Science & Business Media; 2012.
[37] T. Lu, J. Carroll, F. Zeng, On acoustic simulations of cochlear implants, in: Conference on Implantable Auditory Prostheses, Lake Tahoe, CA, 2007.
[38] Meng Q, Yu G, Wan Y, Kong F, Wang X, Zheng N. Effects of vocoder processing on speech perception in reverberant classrooms. In: 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE; 2018. p. 761–5.
[39] Kong F, Wang X, Teng X, Zheng N, Yu G, Meng Q, Reverberant speech recognition with actual cochlear implants: verifying a pulsatile vocoder simulation method. In: the 23rd Intenational congress on Acoustics, ICA, 2019, pp. 3109–3112.
[40] Mo Y, Zhou H, Kong F, Liu Z, Liu X, Huang H, Huang Y, Zheng N, Meng Q, Wu P. Effects of number of maxima and electrical dynamic range on speech-in-noise perception with an?n-of-m? cochlear-implant strategy. Biomed Signal Processing Control 2023;79:104169.
[41] Mehta AH, Lu H, Oxenham AJ. The perception of multiple simultaneous pitches as a function of number of spectral channels and spectral spread in a noise-excited envelope vocoder. J Assoc Res Otolaryngol 2020;21:61–72.
[42] Greenwood DD. A cochlear frequency-position function for several species?29 years later. J Acoust Soc Am 1990;87:2592–605.
[43] Vandali AE, Whitford LA, Plant KL, Clark GM, et al. Speech perception as a function of electrical stimulation rate: using the nucleus 24 cochlear implant system. Ear Hearing 2000;21:608–24.
[44] Ghosh R, Hansen JH. Bimodal cochlear implant processing based on assisted hearing algorithms with cci-mobile: an open-source research platform. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2022. p. 4265–8.
[45] Souza P, Rosen S. Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech. J Acoust Soc Am 2009;126:792–805.
[46] Guo Z, Yu G, Zhou H, Wang X, Lu Y, Meng Q. Utilizing true wireless stereo earbuds in automated pure-tone audiometry. Trends in Hearing 2021;25. 23312165211057367.

[47] Meng Q, Wang X, Cai Y, Kong F, Buck AN, Yu G, Zheng N, Schnupp JW. Time-compression thresholds for mandarin sentences in normal-hearing and cochlear implant listeners. Hearing Res 2019;374:58–68.
[48] Fu Q-J, Zhu M, Wang X. Development and validation of the mandarin speech perception test. J Acoust Soc Am 2011;129:EL267–73.
[49] Wong LL, Soli SD, Liu S, Han N, Huang M-W. Development of the mandarin hearing in noise test (mhint). Ear and hearing 2007;28:70S–4S.
[50] Xi X, Ching TY, Ji F, Zhao Y, Li J-N, Seymour J, Hong M-D, Chen A-T, Dillon H. Development of a corpus of mandarin sentences in babble with homogeneity optimized via psychometric evaluation. Int J Audiol 2012;51:399–404.
[51] Kong F, Mo Y, Zhou H, Meng Q, Zheng N, Channel-vocoder-centric modelling of cochlear implants: Strengths and limitations. In: Proceedings of the 9th Conference on Sound and Music Technology: Revised Selected Papers from CMST, Springer, 2022, pp. 137–149.
[52] Winn MB, Nelson PB. Cochlear implants. In: Oxford Research Encyclopedia of Linguistics.
[53] Oxenham AJ, Kreft HA, Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing, Trend. Hear. 18 (2014) 2331216514553783. 10.1177/2331216514553783, pMID: 25315376.
[54] Zhou N, Xu L, Lee C-Y. The effects of frequency-place shift on consonant confusion in cochlear implant simulations. J Acoust Soc Am 2010;128:401–9.
[55] Brochier T, Schlittenlacher J, Roberts I, Goehring T, Jiang C, Vickers D, Bance M. From microphone to phoneme: an end-to-end computational neural model for predicting speech perception with cochlear implants. IEEE Trans Biomed Eng 2022;69:3300–12.