medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Gaussian-Enveloped Tones (GET): a vocoder that can simulate pulsatile stimulation in cochlear implants

Qinglin Meng,[a] and Huali Zhou[c]

Acoustics Laboratory, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, Guangdong 510641, China

Thomas Lu, and Fan-Gang Zeng[b]

Center for Hearing Research, Department of Otolaryngology –Head and Neck Surgery, University of California, Irvine, California 92697, USA

Running title: Acoustic simulations of cochlear implants

This paper is submitted for consideration for a special issue on Reconsidering Classic Ideas in Speech Communication on the Journal of the Acoustical Society of America

Submission date: Dec. 19, 2021

[a] Electronic mail: mengqinglin@scut.edu.cn. ORCID: 0000-0003-0544-1967.
[b] Electronic mail: fzeng@uci.edu. ORCID: 0000-0002-4325-2780
[c]Also at: College of Electronics and Information Engineering, Shenzhen University, Shenzhen, Guangdong, 518060, China

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

ABSTRACT

Acoustic simulations of cochlear implants allow comprehensive evaluation of not only perceptual performance under impoverished listening conditions but also relative contributions of classical spectral and temporal cues to speech recognition. Conventional simulations use continuous sinusoidal or noise carriers, lacking the vital pulsatile characteristics in a typical cochlear-implant processing strategy. The present study employed Gaussian-enveloped tones (GETs) as a discrete carrier to simulate the electric pulse train in modern cochlear implants. Two types of GET vocoders were implemented and evaluated in normal-hearing listeners to compare their performance to actual cochlear-implant performance. In the first implementation, GETs with different durations were used to simulate electric current interaction across channels that produced vowel and consonant recognition similar to the actual cochlear-implant result. In the second implementation, a direct mapping from electric pulses to GETs was developed to simulate a widely-used clinical *n*-of-*m* strategy in cochlear implants. The GET processing simulated the actual implant speech in noise perception in terms of the overall trend, the absolute mean scores, and their standard deviations. The present results demonstrated that the pulsatile GET vocoders simulate the cochlear implant performance more accurately than the conventional sinusoid or noise vocoders.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

I. 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 can "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 (Dudley, 1939). He realized then that "not only can the speech be remade to simulate the original but it can be changed from the original in a variety of ways" to study experimentally the relative contributions of the 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. The relationship of vocoders to speech production has been firmly established by Gunner Fant as the source-filter model (Fant, 1970), whereas that to speech perception remains relatively unclear but is highly relevant to the success and limitation of modern multi-channel cochlear implants or CIs (Zeng, 2017). The present study has developed a novel vocoder that improves not only acoustic simulation of CIs but also understanding of the relationship between vocoders and speech perception.

The early multi-channel CIs followed closely Dudley's original vocoder idea 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 (Tong *et al.*, 1980; Skinner *et al.*, 1991). The speech understanding of the early CIs was relatively low (<50% correct for sentence recognition in quiet) because of crude F0 and formant extraction methods (i.e., zero crossing) at that time and poor correspondence between sound frequency and electrode position, which remains a significant issue in CI today. Contemporary CIs have all adopted

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

speech processing strategies that extract band-specific temporal envelope from 8-20 frequency bands and use the envelope to amplitude modulate a continuous, but fixed high-rate (at least 2 × the highest envelope frequency) pulse train, which is then delivered to a corresponding electrode in an interleaved fashion without any two electrodes being stimulated simultaneously (Wilson, 1991; Skinner, 2002). In one popular implementation called the *n*-of-*m* strategy, the number of frequency bands (*m*) is greater than the number of stimulating electrodes (*n*), which typically corresponds to the *n* frequency bands with the maximal energy (e.g., Zeng *et al*., 2008). 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.

Acoustic simulations of CIs have been developed and widely used (Svirsky *et al*., 2021) for at least three reasons. First, acoustic simulations minimize the effect of large CI individual variability, which may confound or mask the relative importance of speech processing parameters (e.g., Skinner 2002). Second, acoustic simulations allow evaluation of relative contributions of different cues to auditory and speech perception (e.g., Singh *et al.*, 2009; Xu *et al.*, 2005). 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- (Shannon *et al.*, 1995) or sinusoid-excited (Dorman *et al.*, 1997) vocoders, in which the noise or sinusoid simulates the continuous electric pulse train while the number of frequency bands and their overlaps simulates the limited number of electrodes and their current spread and other characteristics (e.g., Shannon *et al*., 1998). A significant drawback of these traditional vocoder models is the lack of simulation of the pulsatile nature of electric stimulation. Several studies have attempted to develop acoustic models that simulate pulsatile electric stimulation, such as filtered noise bursts (Blamey *et al.*, 1984a, 1984b),

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

filtered harmonic complex tones (Deeks and Carlyon, 2004), and pulse-spread harmonic complexes (Hilkhuysen and Macherey, 2014; Mesnildrey *et al.*, 2016). However, these methods are limited in their ability to simulate three important features in modern CIs. First, these vocoders cannot simulate the discrete nature of pulsatile stimulation on a pulse-by-pulse basis. Second, they do not allow independent manipulation of the overlap between spectral and temporal representation. Third, it is difficult for vocoders with continuous carriers to simulate some CI speech processing strategies, e.g., *n*-of-*m*, in which the low-energy bands are abandoned to produce disrupted temporal envelopes.

Here we identified the Gabor atom (Gabor, 1947), also known as the Gaussian-enveloped tone (GET), as a means of simulating the three features of modern CI processing. The GET has been used to study a wide range of auditory phenomena including interaural timing difference using temporal envelope cues (Buell and Hafter, 1988; Bernstein and Trahiotis, 2002), intensity discrimination (van Schijndel *et al.*, 1999), and cortical encoding of acoustic and electric pulsatile stimulation (Johnson et al., 2017). Because Gaussian envelope is preserved in both time and frequency domains, the GET can be used to simulate and control accurately the discrete pulses and their current spread, producing reasonable CI simulation in binaural masking release and binaural image (Lu *et al.*, 2007; Lu *et al.*, 2010; Goupell *et al.*, 2013; Kan *et al.*, 2013). Here we first analyzed how to use the GET to simulate pulsatile stimulation in CIs and demonstrated its unique advantage in simultaneous and accurate manipulation and control of the spectral spread and temporal resolution. We also simulated the pulsatile timing, amplitude compression and quantization in a typical *n*-of-*m* strategy. Speech performance of the GET vocoder was obtained and compared to that obtained in the actual CI users and traditional noise- or sinusoid-excited vocoders.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

II. VOCODERS AND EXPERIMENT METHODS

Fig. 1A shows the traditional acoustic simulation of CI using either noise (Shannon *et al.*, 1995) or sinewave vocoders (Dorman *et al.*, 1997). The output filters can be used to control the current spread, but no temporal separation feature can be simulated. Fig. 1B shows the proposed pulsatile acoustic simulation using GET vocoders (the first type; see Section II.B; used in Experiment 1). The effective duration of the Gaussian envelope can be used to control the current spread. Meanwhile, the temporal separations between electric pulses can be simulated by the same separations between the center times of the Gaussian-enveloped tone pulses, although the spectral spread and temporal separations are not independent of each other. Fig. 1C (the second type; see Section II.C; used in Experiment 2) shows a variation of Fig. 1B, which 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. The amplitude compression and quantization which are widely used in modern CIs can also be simulated in Fig. 1C. For all the three methods, their front-end processing can share the same blocks of bandpass filters and envelope extraction, e.g., in a traditional temporal envelope based continuous interleaved sampling (CIS) CI strategy (Wilson, et al., 1991), as the figure shows.

In the following, Section II.A analyzes the theory of using the Gabor atom or Gaussian enveloped tone for the simulation of electric pulses; Section II.B and C introduces the detailed experiment methods to validate the two types of GET vocoders (see Fig.1B and 1C) respectively in comparison with both actual CI listeners and conventional continuous-carrier vocoders (see Fig. 1A).

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

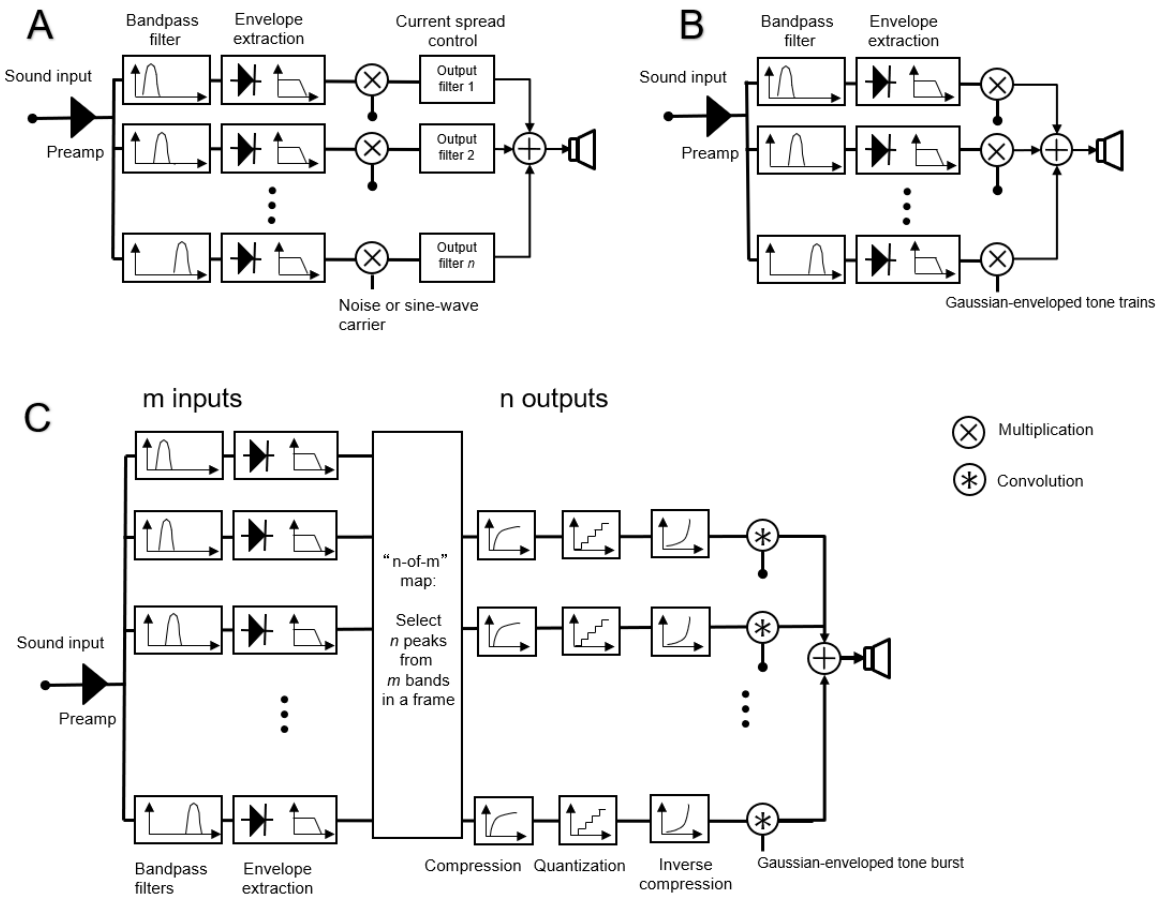

FIG. 1. Block diagrams of standard channel vocoder (A) and the proposed pulsatile simulation vocoders with (B) or without (C) considering the *n*-of-*m* feature. The pulsatile vocoders are using GETs as carriers or using single GET as an impulse response. The front-end pre-amplification, bandpass filter, and envelope extraction can be implemented either in temporal or spectral domain.

A. Simulation of pulsatile stimulation

A Gaussian function is symmetrical in the time domain:

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

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. Its Fourier transform is:

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

The shape of its amplitude spectrum, $\sqrt{2}a\sigma \cdot e^{-2\pi(\sigma f)^2}$, is also a Gaussian function with an effective bandwidth being $B = \frac{1}{\sqrt{2}\sigma}$ 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 to a sinusoidal carrier:

$$s(t) = g_{env}(t) \cdot \sin(2\pi f_c t + \varphi_0) = ae^{-\frac{\pi(t-t_0)^2}{2\sigma^2}} \cdot \sin(2\pi f_c t + \varphi_0), \tag{4}$$

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

Fig. 2 illustrates both waveform (A) and spectrum (B) of a unit-amplitude Gaussian-enveloped single pulse. The carrier frequency is 5 kHz. Acoustic simulation of a continuous

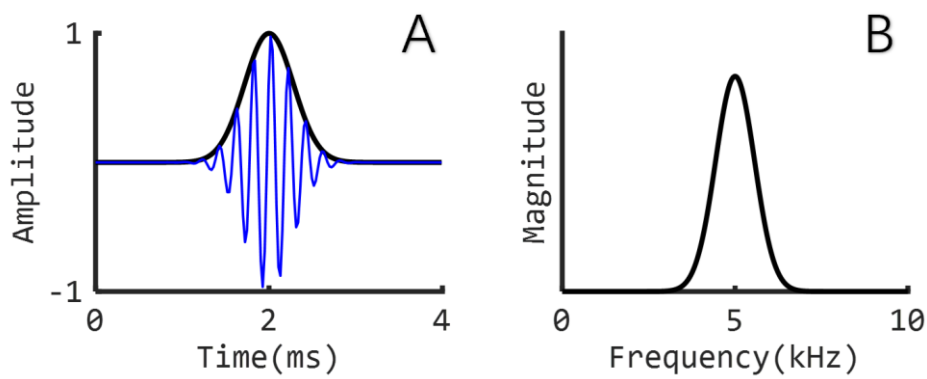

FIG. 2. (Color online) A unit-amplitude single pulse with Gaussian-shaped envelope (black line) in both the time (left panel) and frequency (right panel) domains. The carrier frequency is 5kHz (the blue waveform in the left panel and the frequency with maximum amplitude in the right panel). The σ is 0.5 ms in Eq. (1), producing an effective duration of 0.7 ms and an effective bandwidth of 1.4 kHz.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

electric pulse train can be constructed by periodically repeating $s(t)$, with the period determining the pulse rate of the simulated pulse train.

Different from the CI electric pulses with constant duration at the order of tens of microseconds, the GET duration should be much longer that containing at least several periods of the tone carrier. Therefore, the carrier period or frequency will determine the lower limits of the GET duration. Fig. 3 illustrates the interaction of the GET duration (bandwidth), pulse rate, and carrier frequency. The GET effective bandwidth equals in value to the maximum pulse rate that can be transmitted without obvious temporal interaction between neighboring GETs. Increasing the duration (i.e., larger σ) can decrease the bandwidth, but at the same time, the maximum rate is also decreased.

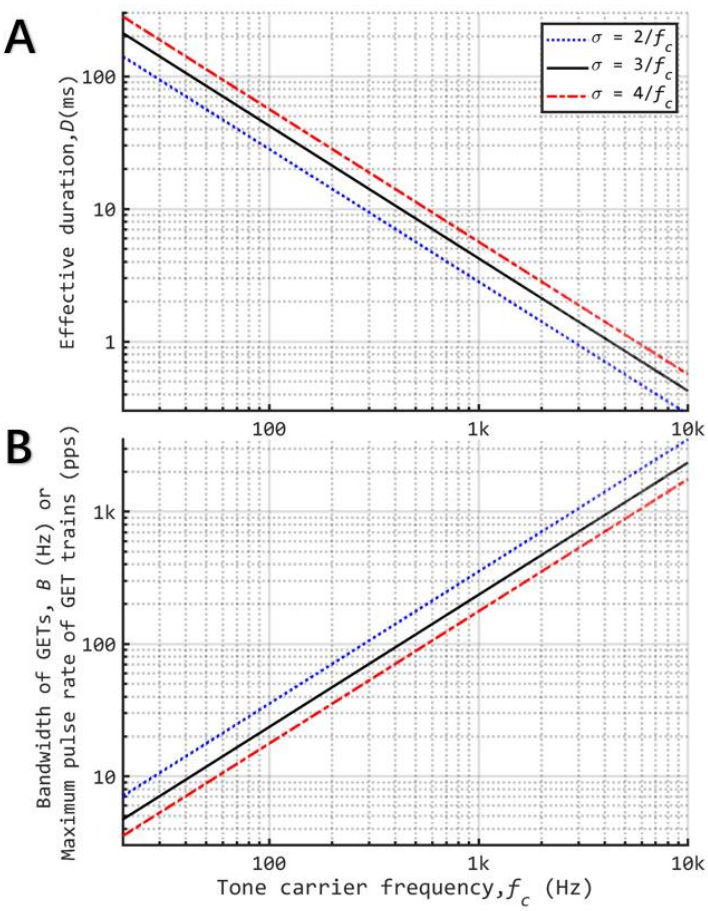

FIG 3. (Color online) The relationship between the tone carrier frequency and the effective duration $D = \sqrt{2}\sigma$ (see Panel **A**) or effective bandwidth $B = 1/D$ (see Panel **B**) of Gaussian-enveloped tones (GETs). All axes are logarithmic scaling. The σ was assumed to be $2/f_c$, $3/f_c$,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

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

It can be found that 1) a conventional pulse rate of 900 pps could only be simulated with carrier frequencies above ~3 kHz, 2) for carrier frequencies within the middle frequency range around 2 kHz, the 900 pps is still possible to be simulated but neighboring GETs have moderate temporal interaction, 3) for low frequency carrier, the pulsatile feature cannot be guaranteed. One approach (not used in current experiments) may partially reduce the temporal interactions, like upward shifting the carrier frequency or reducing the pulse rate.

In short, the GETs can simulate and manipulate five important parameters of CI processing or stimulation: (1) the pulse rate by changing the period of the pulse generation, (2) the temporal envelope (including its compression and quantization) by changing the amplitude of individual GETs in a pulse train within a channel, (3) the spectral envelope by changing the GET amplitude across channels, (4) the place of excitation by changing the carrier tone frequency, and (5) the spread of excitation by changing the effective bandwidth in GETs. The precise manipulation of these five important parameters allows acoustic simulation of modern CIs using pulsatile electric stimulation. The limitations from the dependent relationships between duration, bandwidth, and carrier frequency of the GETs were discussed as above and should be taken care of during algorithm design and experiments of CI simulations with GETs.

B. Experiment 1: Simulation of Current Spread

There is a significant difference in simulating the spread of excitation between the standard vocoder and the GET implementation. In the standard simulation, the spread of excitation is manipulated by changing the filter type and the bandwidth of the synthesis filters at the vocoder output stage. For the GETs, the spread of excitation is manipulated by increasing or decreasing the Gaussian tone duration, which produces a corresponding change in narrowing or widening the spectral bandwidth for each pulse.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

In Experiment 1, five vocoders were used: three standard vocoders, *sine-wave*, *noise-adjacent*, and *noise-overlap* (Fig. 1A), and two proposed vocoders incorporating the GET simulation, *GET-adjacent* and *GET-overlap* (Fig. 1B).

Analysis processing of all five vocoders: the analysis filter banks consist of *N* band-pass filters (4th order Butterworth). The frequency spacing for cutoffs for the filter bank was defined in the range of [80, 7999] Hz according to a Greenwood map (Greenwood, 1990). The filtered signals were half-wave rectified and lowpass filtered (50 Hz 4th order Butterworth) to extract the envelope for each channel.

Synthesis processing for the three standard vocoders: For the *sine-wave* vocoder, a sine wave with a frequency centered at the corresponding analysis filtering band is used as the carrier. For the *noise-adjacent vocoder*, bandpass noise carriers were generated by passing white noise through filters which were the same as the analysis filters. The *noise-adjacent* vocoder provides upper-bound performance that has a minimum of simulated electrode interaction. For the *noise-overlap* vocoder, low-pass filters (4th order Butterworth) were used to pass white noises for generating lowpass noise carriers. The cutoff frequencies of the low-pass filters were the same as the upper cutoff frequencies of the analysis filters. The signal carriers in each band were corresponding lowpass noise. Lowpass filters were chosen to maximize interaction between channels and provide a lower bound of performance with simple manipulation. For the two noise vocoders, after modulating each channel of filtered noise with the channel envelope, the output was filtered again to band-limit each channel. The band-limiting filters are the same as those used for the noise carrier generation. The final vocoded signal was synthesized by summing up all channels. A reduced number of bands was simulated by subdividing the frequency space

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

equally by the number of bands. For example, the 2-band vocoder will have two channels covering the frequency space.

Synthesis processing for the two pulsatile GET vocoders: For the GET vocoders, instead of modulating a filtered noise signal at the synthesis stage, the envelope in each channel modulates the amplitude of the train of GETs. Because the first experiment focused on the spectral interaction, the pulses among all channels were synchronized, meaning that the "interleaved sampling" feature was not simulated. Fig. 4 shows a 100-Hz pulse train, repeating the single pulse every 10 ms. The pulse train's spectral envelope remains the same as the single pulse but its spectral fine structure becomes discrete with 100-Hz spacing (in this case, the maximum-amplitude frequency is 5 kHz with symmetrically decreasing-amplitude components at 4.9, 4.8, 4.7… and 5.1, 5.2, 5.3… kHz, respectively, see inset in the right panel). For the *GET-adjacent* vocoder, $D = \sqrt{2}\sigma = 7.0$ ms, while for the *GET-overlap* vocoder, $D = \sqrt{2}\sigma = 1.2$ ms.

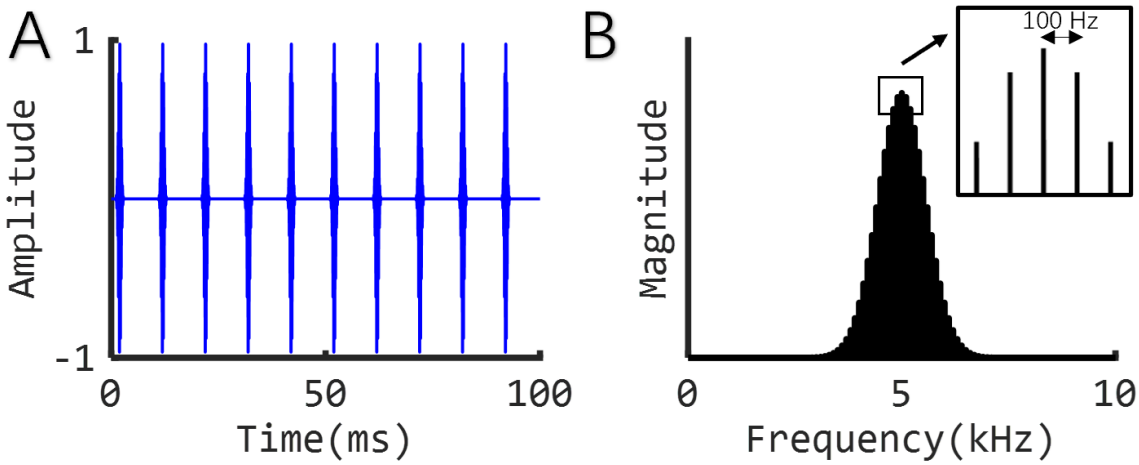

FIG. 4. (Color online) A 100-Hz pulse train, repeating the single pulse every 10 ms, in both the time (left panel) and frequency (right panel) domains. The parameters of the individual pulses are the same as those in Fig. 2.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

CI stimulation were simulated using five different vocoders, i.e., *sine-wave*, *noise-adjacent*, *noise-overlap*, *GET-adjacent*, and *GET-overlap*. The numbers of channels tested were 2, 4, 8, 16, and 32. Six normal hearing (NH) participants, ages 18-21, were tested in an anechoic chamber (IAC) using the English vowel and consonant recognition tests adopted from Friesen *et al.*, (2001). There were 12 medial vowels and 14 medial consonants in the vowel and consonant tests respectively. Fig. 5 demonstrates some 16-channel vocoded stimuli for vowel tests.

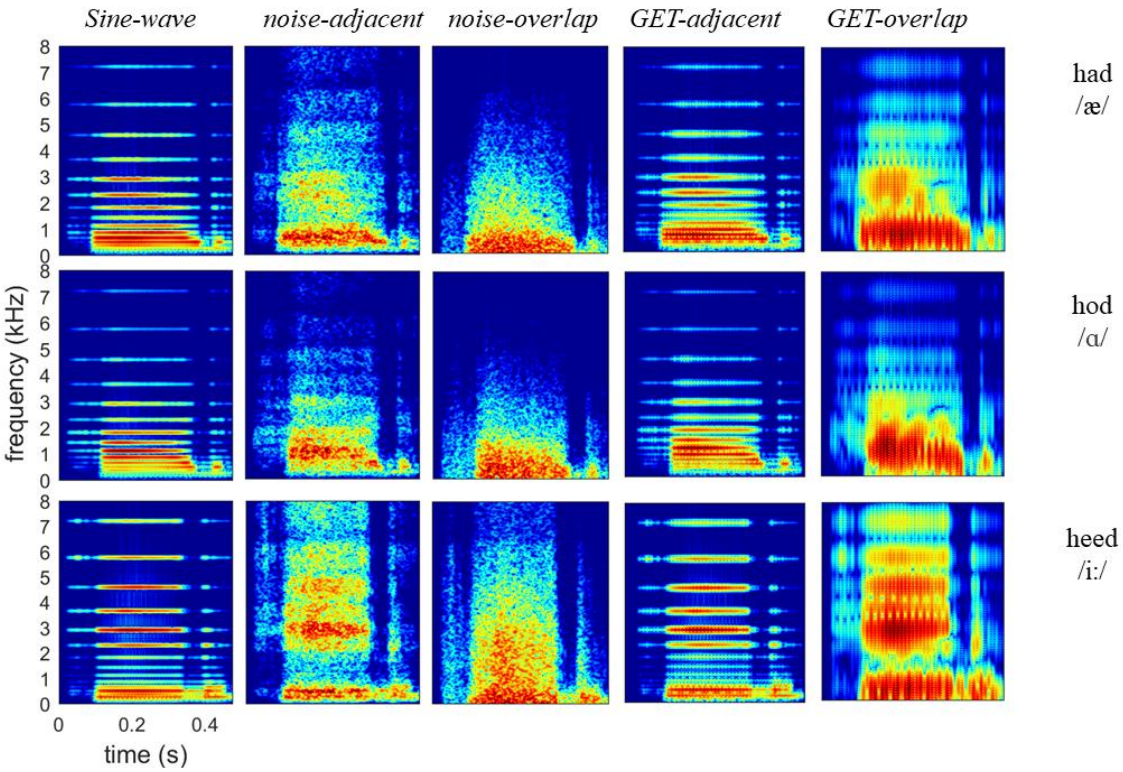

FIG. 5. (Color online) Spectrograms of three vowel stimuli encoded by the *sine-wave*, *noise-adjacent*, *noise-overlap*, *GET-adjacent*, and *GET-overlap* vocoders with 16 channels.

Each stimulus was presented 10 times. Stimuli were presented through headphones (HDA 200, Sennheiser), and the sound level was calibrated to 70 dB SPL. This procedure was conducted following procedures approved by the University of California Irvine Institutional Review Board.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

C. Experiment 2. Simulation of the *n*-of-*m* strategy ACE

1. Method theory: electrodogram to vocoded sound mapping

The GETs are also applicable for directly transferring any pulsatile CI electrodogram to pulsatile vocoded sound. To be more illustrative, Fig. 6A demonstrates a 10-channel electrodogram. (Note: single vertical lines are 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 are not considered in this model.) 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 (Greenwood, 1990). Then, a band-specific GET pulse was generated by setting the parameters in equation (1) as $a = 1$, $t_0 = 0$, and

$$\sigma = \frac{2}{f_c} \tag{5}$$

where f_c denotes the center frequency of the specific band. As a result, the band-specific GET pulse had a 6.82-dB duration of

$$D = \sqrt{2}\sigma = \frac{2\sqrt{2}}{f_c} \tag{6}$$

and a 6.82-dB bandwidth of

$$B = \frac{1}{D} = \frac{\sqrt{2}}{4} f_c \tag{7}$$

Then the acoustic GET train at the kth channel in Fig.6B is derived by

$$p_{a,k}(t) = (p_{e,k}(t) * e^{-\frac{\pi t^2}{2\sigma^2}}) \cdot \sin(2\pi f_c t + \varphi_0) \tag{8}$$

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

where $p_{e,k}(t)$ and $p_{a,k}(t)$ denotes the electric and acoustic pulse trains in Fig.6A and 6B respectively, " $*$ " denotes a convolution calculation, σ and f_c are channel-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.

Fig. 6B shows the 10-channel GET pulse trains, which have temporally separated waveforms for high-frequency channels, but overlapping waveforms for low-frequency channels. Fig. 6C shows the overall waveform summed from the 10 bands.

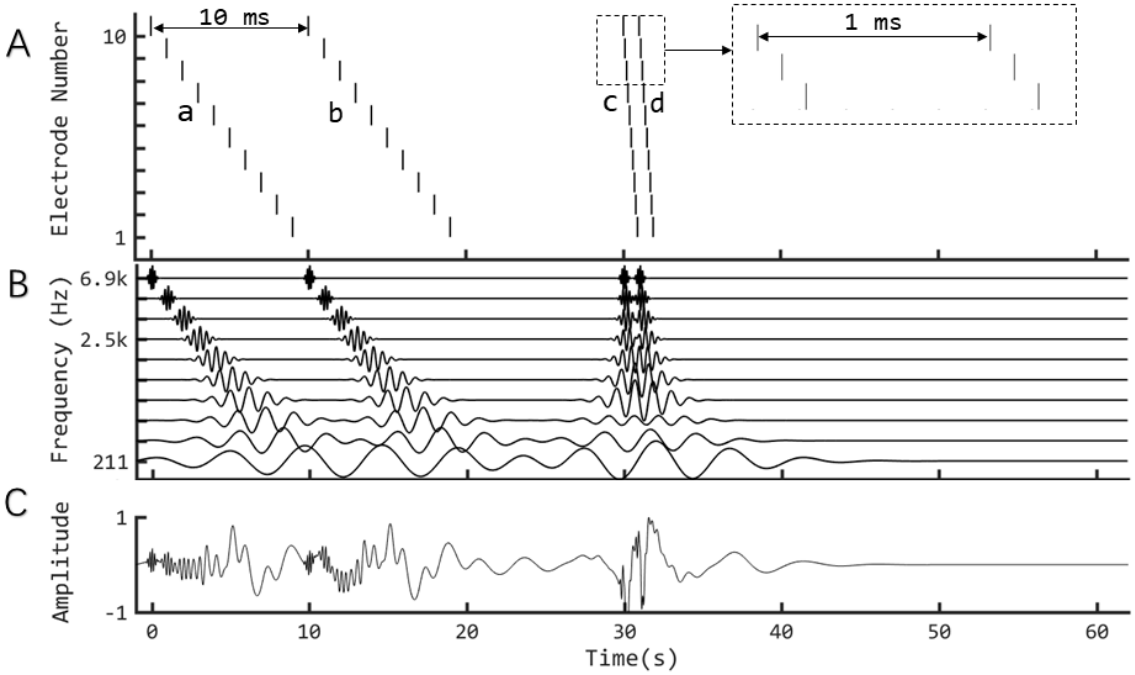

FIG. 6. Mapping a CI electrodogram to a sound using a GET vocoder. **A**. A 10-channel CI electrodogram, including two pulse sweeps with a 10-ms difference between **a** and **b**, as well as two additional sweeps with a 1-ms difference between **c** and **d,** corresponding to stimulation rates of 100 pps and 1000 pps, respectively. **B**. GET pulses mimicking the electric pulse trains. **C**. The final GET waveform resulting from the sum of ten band-specific GET pulses trains in B.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

Using the above method, any electrodograms including the widely used *n*-of-*m* strategy like Advanced Combinational Encoder (ACE) strategy, which is the present default strategy in Nucleus cochlear implants (Vandali *et al.*, 2000), can be converted to vocoded sounds. Following the preliminary results which showed comparable data between the ACE-GET vocoder and actual CI users (Kong *et al.*, 2019), in this paper a battery of speech perception tasks was carried out to further explore the potentials of ACE-GET vocoder on CI simulation.

In the clinical fitting of ACE strategy, the intensity 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 later way 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 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-ch ACE-GET vocoders (denoted by GET1 and GET2) were compared with two 22-ch sine-carrier standard vocoders (125 Hz and 250 Hz envelope cutoffs, denoted by Sin250 and Sin125 respectively) with minimum channel overlapping as shown in Fig. 1A.

The detailed implementation methods of the vocoders: First, an exponential function was used to transfer the electric current values to envelope power values. Then, single-sample pulse trains from each band were convolved with a Gaussian enveloped pulse with $\sigma = 3/f_c$. Then, the output was used to multiply a sinusoidal carrier with a frequency of f_c at the center of the

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

corresponding band and an arbitrary initial phase (a random initial phase in this study). The average power of each band was unchanged. Finally, the modulated signals were summed up to produce the vocoded stimulus.

The difference between GET1 and GET2 is only between their electric-to-acoustic mapping functions, which are Equation (9) and (10) respectively:

$$L_a = \frac{1}{\alpha}((1+\alpha)^{L_e} - 1) \tag{9}$$

and

$$L_a = \frac{1}{2.72\alpha}(e^{L_e}(1+\alpha) - 1) \tag{10}$$

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. In the current 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 Equations (9) and (10), recovered acoustic level ranges were 32.7 dB and 5.3 dB for GET1 and GET2 respectively. Equation (9) is directly based on the default setting of the acoustic-to-electric compression function in ACE. Therefore, it was hypothesized that the GET1 could simulate the best performance of CI listeners with the corresponding ACE strategy and GET2 would significantly degrade the performance because of the much narrower level range. Otherwise, the implementation details of the vocoder were the same as in Meng *et al.* (2018).

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 (Greenwood, 1990). The filtered signals were full-wave rectified and lowpass 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

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

multiplied by the corresponding envelope. The final vocoded stimuli were generated by a summation of the modulated carriers. It was hypothesized that speech intelligibility would be better with a higher cutoff frequency in the envelope extraction (Souza and Rosen, 2009), i.e., Sin250 better than Sin125. Both Sin250 and Sin125 vocoder simulation results were hypothesized to be better than those of most CI listeners.

In Fig. 7, a Mandarin sentence is used to demonstrate the vocoded speech using the four vocoders, i.e., Sin250, Sin125, GET1, and GET2. It shows that the GET vocoders ensembles the ACE-electrodogram more than the sine vocoders. It was hypothesized that the speech intelligibility would be worse but closer to actual CI results with the GET vocoders than with the sine vocoders.

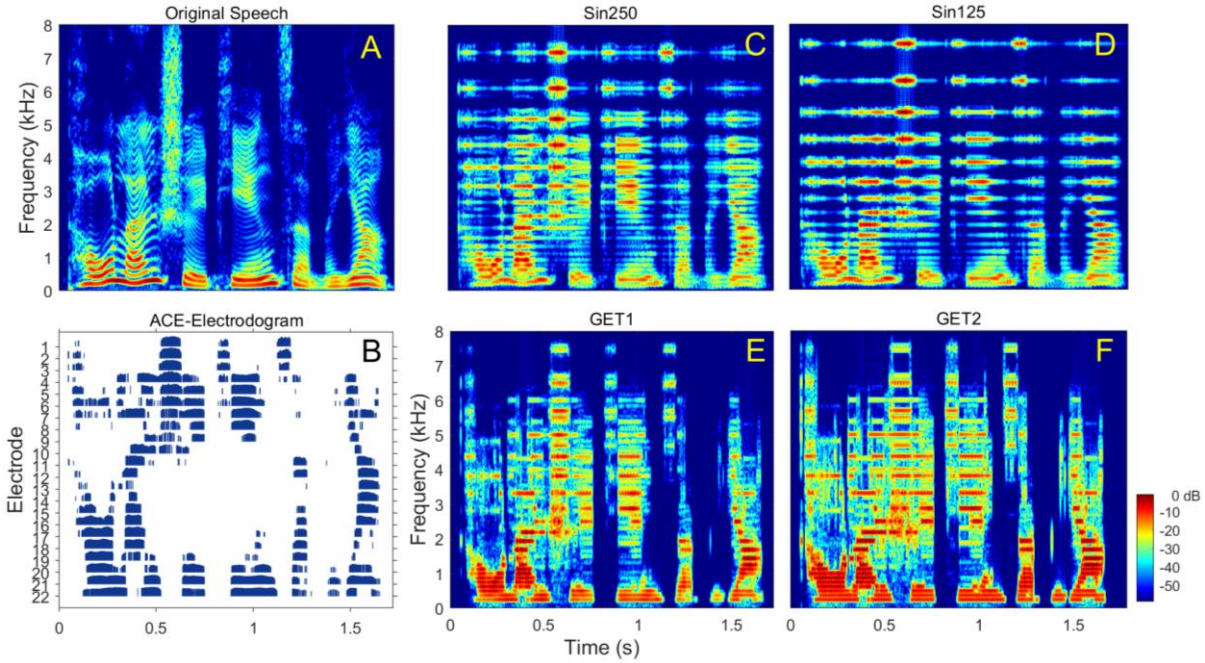

FIG. 7. (Color online) Speech stimulus demonstrations for the ACE simulation experiment. **A**. Spectrogram of a clear sentence speech. **B**. Electrodogram based on an ACE strategy for the original speech. **C-F**. Spectrograms for vocoded speech using Sin250, Sin125, GET1, and GET2 vocoders respectively.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

3. Experiment method: Participants and Tasks

Two groups of NH participants (ten in each group, ages 18-29, and native Mandarin speakers) were tested in a soundproof room. Group 1 used Sin250 and GET1, and Group 2 used Sin125 and GET2. Three Mandarin Chinese recognition tasks were tested, i.e., time-compression threshold, sentence in noise recognition, sentence in reverberation recognition. The results for the four tasks with the two vocoders in these NH participants were compared with actual CI results from the first author's previous experiments (Meng *et al.*, 2019) as well as some newly collected data in this work. These experiments were conducted following procedures approved by the Medical Ethics Committee of Shenzhen University, China. 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 (Fu *et al.*, 2011).

2) Speech reception thresholds (SRTs) in speech-shaped noise (SSN) and 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 (Wong *et al.*, 2007). The TCT and SRT test procedures followed Experiment 2 of Meng *et al.*, (2019) exactly, in which ten CI subjects (9/10 adults) with various hearing histories were tested.

3) Recognition of speech in reverberation was measured using a Mandarin BKB-like sentence corpus (Xi *et al.*, 2012), whose quiet sentences were convolved with simulated room impulse responses (RIRs). The RIRs were generated using a MATLAB function (https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator) with its default setting, except the reverberation times (T60) were set as 0, 0.3, 0.6, and 0.9 s. For each T60, one

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

sentence list was used. Seven CI participants with various hearing histories were also tested for comparison (See Table I).

TABLE I. Detailed information of the seven CI participants in the speech in reverberation test

Subject	Gender	Age range (yr)	CI Experience (yr)	CI Processor	Etiology
C14	F	41-45	12	CP810	Drug induced
C23	M	31-35	11	CP810	Sudden deafness
C30	M	11-15	10	Freedom	LVAS
M5	M	16-20	15	OPUS-2	Virus infection
C16	F	21-25	2	Freedom	Unknown
M17	F	16-20	5	OPUS-2	Genetic
M16	F	16-20	7	OPUS-2	Unknown

The paired-sample *t*-test and two-sample *t*-test functions in MATLAB, i.e., ttest.m and ttest2.m, were used to examine the statistical significance of the means' difference for within-subject comparisons and inter-subject comparisons respectively. A criterion of $p < 0.05$ was used for indicating a significant difference.

III. RESULTS

A. Experiment 1

Results are shown in Figure 8. For the vowel test, the six NH participants scored approximately 20% under all simulation conditions with two channels. Increasing the number

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

of channels also increased performance. With eight channels, performance under the different conditions began to separate. The sine-wave vocoder outperformed actual CI data (adapted from Friesen *et al.*, (2001)), which showed no improvement beyond 8 channels. A *noise-adjacent* vocoder and a *GET-adjacent* vocoder showed similar performance trends. When electrode interaction was simulated with overlapping filters, the subject performance showed a plateau near 60% with noise overlap, similar to actual CIs. The Gaussian overlap condition underperformed CI data in this case, saturating near 35% with eight channels.

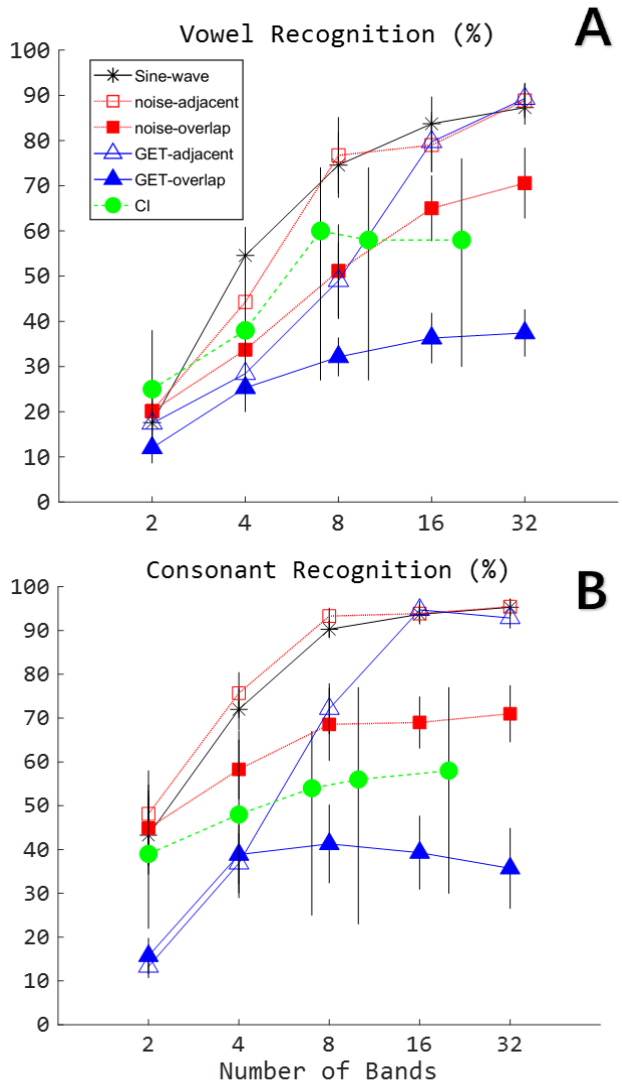

FIG. 8. (Color online) Vowel (**A**) and consonant (**B**) recognition as a function of number of bands (channels). Cochlear implant data is adapted from Friesen et al., (2001). Simulation data is averaged from 6 normal hearing subjects. For the simulation data, standard errors are indicated by the vertical bars. For the CI data, the bars show the entire ranges of performance across all of their 19 participants.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Consonant recognition showed similar performance trends across the simulation types, with *sine-wave*, *noise-adjacent*, and *GET-adjacent* outperforming CIs (adapted from Friesen *et al.* (2001)) when there were eight or more channels simulated. *Noise-overlap* brought the performance closer to actual CI data, while again, *GET-overlap* underperformed CIs. With only two channels, both *GET-adjacent* and *GET-overlap* showed performances that were much lower than actual CIs.

The results suggest that it is feasible to simulate current spread by manipulating durations of GET pulses. In both noise vocoder and GET vocoder, performance was substantially degraded by the increased current spread in both tasks. With eight or more bands, GET vocoders showed better simulation performance than the sine-wave and noise vocoders in that the actual CI data fell in the range between the adjacent and overlap versions of the GETs.

B. Experiment 2

The results with the four 22-ch vocoders, i.e., GET1, GET2, Sin250, and Sin125 are shown in Fig. 9. (**A**) For the TCT test, a significant decreasing trend was found from Sin250 (mean = 16.1 syllables/sec), Sin125 (13.9), GET1 (12.3), GET2 (9.4), to actual CI (6.8) results, while their standard deviations are comparable within the range from 1.0 to 1.2 syllables/s. (**B**) For the SRT test, there was no significant difference 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 GET2 (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 GET1 (means: -1.5 dB in SSN and 4.5 dB in babble noise) were significantly lower than those with Sin250 and Sin125 and significantly higher than those with GET2 and CIs. The mean SRTs in babble noise were always

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

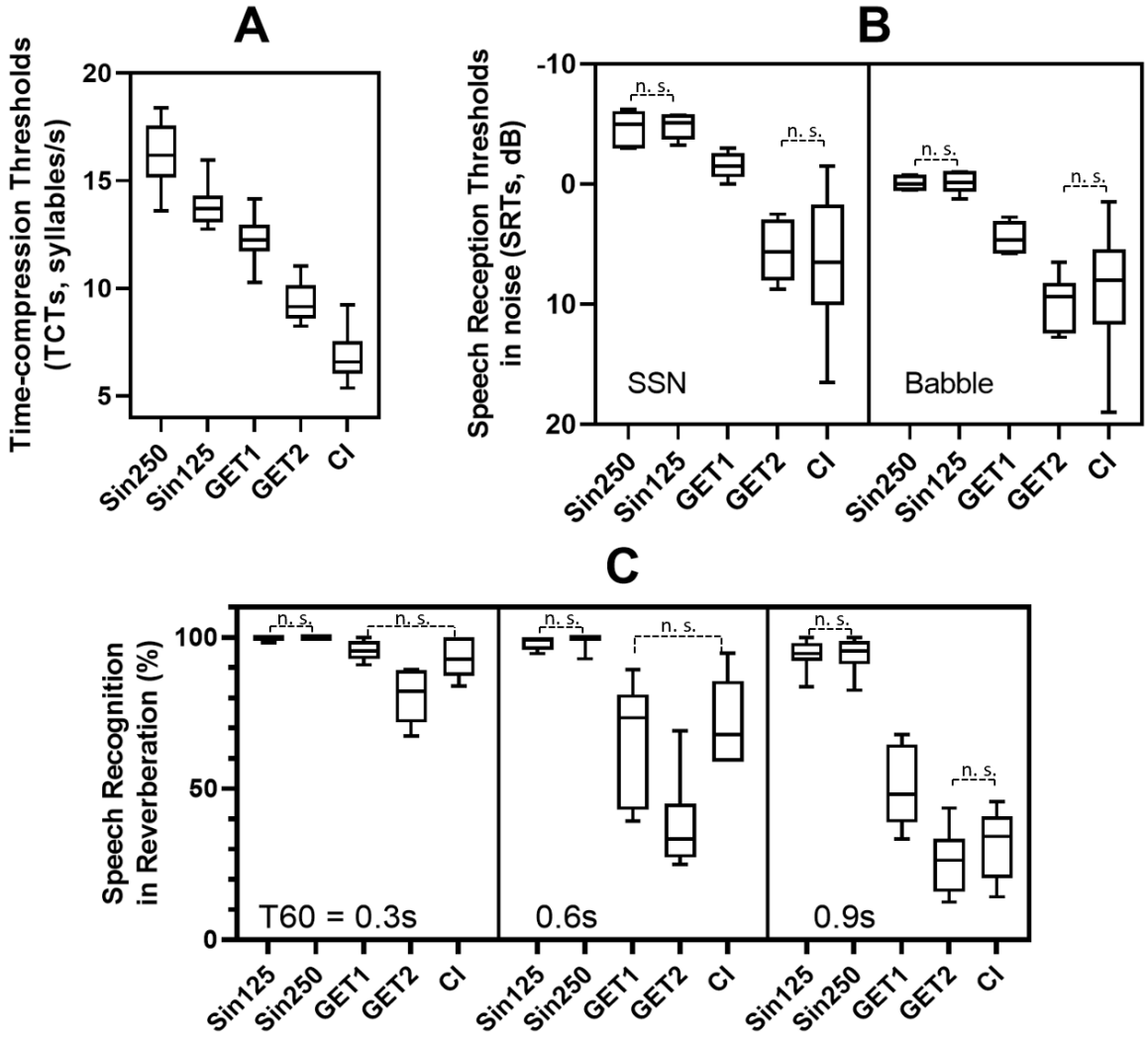

FIG. 9. Results from three speech recognition tasks with two 22-ch sine-wave vocoders (Sin250: 250 Hz cut-off envelope; Sin125: 125 Hz cut-off envelope) and two GET vocoders (GET1 and GET2; their difference is only in the intensity dynamic range, i.e., 32.7 dB and 5.3 dB for GET1 and GET2 respectively) 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 GET1, and the other group used Sin125 and GET2. **A.** Time-compression threshold results. **B.** Speech reception threshold results of a speech in (SSN and babble noise) noise recognition experiment. **C.** Speech recognition scores in reverberation with T60 = 0.3, 0.6, and 0.9s. Pair-wise comparisons were examined. In each sub-figure, "n. s." denotes the non-significant difference ($p \geq 0.05$), otherwise, there was a significant difference.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

significantly lower than those in SSN for all five conditions. (**C**) For the reverberant speech recognition test, all vocoders and the actual CI condition showed a significant trend to face more difficulty when the reverberation time increased. However, the sine vocoder simulations were much less sensitive to reverberation than the CI users. It is showed that even with T60 = 0.9 s, the sine vocoders still derived >94% means, which were much higher than CI participants' 32%. The GET1 and GET2 derived significantly lower scores than the sine vocoders did. Under the T60 = 0.3 s and 0.6 s conditions, there was no significant mean score difference between GET1 and CI, while GET2 derived significantly lower mean scores than GET1 and CI did. Under the T60 = 0.9 s condition, there was no significant mean score difference between GET2 and CI, while GET1 derived significantly higher mean scores than GET2 and CI did.

In all three tasks, GET vocoders were able to simulate the actual CI performance more closely than the sine vocoders. In fact, the sine vocoders overestimated the CI performance in all tasks. In the time-compression task, all vocoders produced better than CI performance with GET2 being the closest (Fig. 9A). In the SRT in noise test, GET2 and CI produced comparable performance (Fig. 9B). In the reverberation task, GET1 produced similar to CI performance in two conditions (T60 = 0.3 and 0.6s in Fig. 9C) and GET2 in the T60 = 0.9s condition (Fig. 9C).

IV. DISCUSSION

This study introduced a novel vocoder that used Gaussian-envelope-tones to simulate not only the pulsatile nature of electric stimulation but also channel interaction and *n*-of-*m* speech processing in modern CIs. The evaluation results showed that GET performance was generally more similar to the actual CI performance than the standard vocoders in all three speech intelligibility tasks. The following discussion focuses on CI simulations in terms of comparing GETs to single electric pulses, channel interactions, and future directions.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

A. GETs and electric pulses

The GET pulse can be used to simulate a "perceivable" atom of sound, which can be traced back to Gabor (1947). In this model, each action potential from an auditory nerve fiber may correspond to a small amplitude GET pulse with the carrier set at the nerve fiber's characteristic frequency, especially when the pulse rate is low (e.g., < 200 Hz). At the other end of the spectrum, the model can be a phenomenological one, where each GET pulse corresponds to a single perceivable unit of an electrical pulse. The amplitude of the GET pulse is scaled proportionally to the electrical stimulus' current levels. It is also possible to closely approximate existing noise and sine-wave vocoders by using wide GET pulses and summing many pulses occurring at high rates (e.g., > 1000 Hz). Several works have proposed to generate pulsatile simulation using harmonic complex (Churchill *et al.*, 2014) or equally related frequency limited pulse trains (Apoux *et al.*, 2018). They used the periodicity cue generated from several harmonics for each channel, but the precise timing information of electric pulse cannot be simulated as the GET vocoders. With the studies that use filtered harmonic complexes (McKay and Carlyon, 1999; Deeks and Carlyon, 2004; Churchill *et al.*, 2014), a concern is that the individual components of the harmonics can be resolved. The GET can be used as a generic tool of CI simulation. Main features in CIs can be independently manipulated: the place of stimulation, pulse time, temporal envelope, spectral envelope and spectral interaction, intensity quantization and maxima-selection, by corresponding features of the acoustic pulses.

B. Channel interactions

Channel interactions is one factor underlying the poor- and large-variance performance for CI participants, which has also been corroborated by several previous studies (Fu and Nogaki, 2005; Bingabr *et al.*, 2008; Strydom and Hanekom, 2011; Grange *et al.*, 2017; O'Neill *et al.*,

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2019; Mehta *et al.*, 2020). In both the standard and GET vocoders, channel interactions can be modeled by varying the overlap and slope of the analysis and synthesis filter banks. Different from the noise-or sine-vocoders that produced performance better than actual CI performance even in the case of the maximal channel interaction, the GET vocoder produced a wide range of performance encompassing the actual CI performance (Figs. 8 and 9).

In addition to channel interaction, distorted frequency mapping may play a role in the discrepancy of simulation to actual CI performance (Shannon *et al.*, 1998). The frequency-shifted simulations were not tested here. However, an upward frequency-shifting is also possible with GETs, which are able to further reduce temporal interaction at higher tone frequencies (Figs. 3 and 6).

C. Limitation and future directions

An inherent limitation with the GETs is the tradeoff between temporal duration and spectral bandwidth. Shortening the pulse duration increases the spectral bandwidth; thus, lower carrier frequencies have stricter limits. The real CIs have no such limitation, in which both pulse duration and pulse rate are the same whether it is a basal or apical electrode.

Similar to the sine- or noise-excited vocoders, the GET also discarded the fine structure (e.g., the bi-phasic waveform and the inter-phase gap) of the pulses in the present implementation. The GET simulation performance on F0 or pitch cues in CIs has not been considered yet. Any improvement is of course limited by the trade-off between temporal and spectral representation. Future studies may explore discrete tones with different envelopes, e.g., Gammatones (de Boer, 1975; Patterson et al., 1987; Ausili et al., 2019). Gammatones are similar in shape to a Gaussian envelope but is asymmetrical with a longer tail.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Regardless of different types of simulation, they all are able to produce good speech performance similar or better than that by modern multi-channel cochlear implant users. This high level of performance re-confirms the classic ideas dating back to Dudley and Fant that while speech quality is affected by different carriers or sound sources, speech intelligibility is determined by the slowly-varying envelope cues controlled by the vocal tract.

V. CONCLUSION

Here we used Gaussian-Enveloped-Tones (GETs) to develop a novel vocoder that simulates pulsatile electric stimulation in cochlear implants. Parametric analysis and systematic evaluation led to the following conclusions:

(1) GETs have produced a more realistic CI performance in speech intelligibility compared to the traditional vocoder simulations.

(2) Similar to the traditional vocoders, GETs are also able to simulate spectral channel interactions to produce asymptotical performance in consonant and vowel recognition beyond 4-7 channels.

(3) GETs can simulate both temporal and spectral aspects of modern CIs from discrete representation of information on a pulse-by-pulse basis in a single electrode to multi-channel speech processing such as the *n*-of-*m* strategy.

ACKNOWLEDGMENTS

We thank all the participants in these experiments. J. Carroll and S. Tiaden helped collect the data in Experiment 1. F. Kong and Y. Xiao helped collect the data in Experiment 2. 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.).

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

REFERENCES

Apoux, F., Carter, B. L., and Healy, E. W. (**2018**). "Effect of dual-carrier processing on the intelligibility of concurrent vocoded sentences," J. Speech. Lang. Hear. Res. **61**, 2804-2813.

Ausili, S. A., Backus, B., Agterberg, M. J., van Opstal, A. J., and van Wanrooij, M. M. (**2019**). "Sound localization in real-time vocoded cochlear-implant simulations with normal-hearing listeners". Trend. Hear. **23**, 1-18.

Bernstein, L. R., and Trahiotis, C. (**2002**). "Enhancing sensitivity to interaural delays at high frequencies by using "transposed stimuli"," J. Acoust. Soc. Am. **112**, 1026-1036.

Bingabr, M., Espinoza-Varas, B., and Loizou, P. C. (**2008**). "Simulating the effect of spread of excitation in cochlear implants," Hear. Res. **241**, 73-79.

Blamey, P. J., Dowell, R. C., Tong, Y. C., Brown, A. M., Luscombe, S. M., & Clark, G. M. (**1984a**). Speech processing studies using an acoustic model of a multiple-channel cochlear implant. J. Acoust. Soc. Am. **76**(1), 104-110.

Blamey, P. J., Dowell, R. C., Tong, Y. C., & Clark, G. M. (**1984b**). An acoustic model of a multiple-channel cochlear implant. J. Acoust. Soc. Am. **76**(1), 97-103.

Buell, T. N., & Hafter, E. R. (**1988**). Discrimination of interaural differences of time in the envelopes of high-frequency signals: Integration times. J. Acoust. Soc. Am., **84**(6), 2063-2066.

Chen, H., Ishihara, Y. C., and Zeng, F. G. (**2005**). "Pitch discrimination of patterned electric stimulation," J. Acoust. Soc. Am. **118**, 338-345.

Churchill, T. H., Kan, A., Goupell, M. J., Ihlefeld, A., and Litovsky, R. Y. (**2014**). "Speech perception in noise with a harmonic complex excited vocoder," J. Assoc. Res. Otolaryngol. **15**, 265-278.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

de Boer, E. (**1975**). "Synthetic whole-nerve action potentials for the cat," J. Acoust. Soc. Am. **58**, 1030-1045.

Deeks, J. M., and Carlyon, R. P. (**2004**). "Simulations of cochlear implant hearing using filtered harmonic complexes: implications for concurrent sound segregation," J. Acoust. Soc. Am. **115**, 1736-1746.

Dorman, M. F., Loizou, P. C., and 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.

Dudley, H. (**1939**). “Remaking speech,” J. Acoust. Soc. Am. **11**, 169-177.

Fant, G. (1970). “Acoustic theory of speech production”. Walter de Gruyter.

Friesen, L. M., Shannon, R. V., Baskent, D., and 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.

Fu, Q. J., Zhu, M., and Wang, X. (**2011**). "Development and validation of the Mandarin speech perception test," J. Acoust. Soc. Am. **129**, L267-L273.

Fu, Q. J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19-27.

Gabor, D. (**1947**). "Acoustical quanta and the theory of hearing," Nature. **159**, 591-594.

Goupell, M. J., Stoelb, C., Kan, A., and Litovsky, R. Y. (**2013**). "Effect of mismatched place-of-stimulation on the salience of binaural cues in conditions that simulate bilateral cochlear-implant listening," J. Acoust. Soc. Am. **133**, 2272-2287.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Grange, J. A., Culling, J. F., Harris, N., and Bergfeld, S. (**2017**). "Cochlear implant simulator with independent representation of the full spiral ganglion," J. Acoust. Soc. Am. **142**, L484.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species--29 years later," J. Acoust. Soc. Am. **87**, 2592-2605.

Hartmann, R., Topp, G., and Klinke, R. (**1984**). "Discharge patterns of cat primary auditory fibers with electrical stimulation of the cochlea," Hear. Res. **13**, 47-62.

Hilkhuysen, G., and Macherey, O. (**2014**). "Optimizing pulse-spreading harmonic complexes to minimize intrinsic modulations after auditory filtering," J. Acoust. Soc. Am. **136**, 1281.

Johnson, L. A., Santina C. C. D., and Wang X. (**2017**) "Representations of time-varying cochlear implant stimulation in auditory cortex of awake marmosets (Callithrix jacchus)." J. Neurosci. **37**, 7008-7022.

Kan, A., Stoelb, C., Litovsky, R. Y., and Goupell, M. J. (**2013**). "Effect of mismatched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users," J. Acoust. Soc. Am. **134**, 2923-2936.

Kong, F., Wang, X., Teng, X., Zheng, N., Yu, G., and Meng, Q., (**2019**). “Reverberant speech recognition with actual cochlear implants: verifying a pulsatile vocoder simulation method,” in *Proceedings of the 23rd International Congress on Acoustics*, pp. 3109-3112.

Lu, T., Carroll, J., and Zeng, F. G. (**2007**). "On acoustic simulations of cochlear implants," in *Conference on Implantable Auditory Prostheses (abstract)* (Lake Tahoe, CA).

Lu, T., Litovsky, R., and Zeng, F. G. (**2010**). "Binaural masking level differences in actual and simulated bilateral cochlear implant listeners," J. Acoust. Soc. Am. **127**, 1479-1490.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

McKay, C. M., and Carlyon, R. P. (**1999**). "Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains," J. Acoust. Soc. Am. **105**, 347-357.

Mehta, A. H., Lu, H., & Oxenham, A. J. (**2020**). "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., **21**(1), 61-72.

Meng, Q., Wang, X., Cai, Y., Kong, F., Buck, A. N., Yu, G., Zheng, N., and Schnupp, J. (**2019**). "Time-compression thresholds for Mandarin sentences in normal-hearing and cochlear implant listeners," Hear. Res. **374**, 58-68.

Meng, Q., Yu, G., Wan, Y., Kong, F., Wang, X., and Zheng, N. (**2018**). "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)*, pp. 761-765.

Mesnildrey, Q., Hilkhuysen, G., and Macherey, O. (**2016**). "Pulse-spreading harmonic complex as an alternative carrier for vocoder simulations of cochlear implants," J. Acoust. Soc. Am. **139**, 986-991.

Nelson, D. A., Donaldson, G. S., and Kreft, H. (**2008**). "Forward-masked spatial tuning curves in cochlear implant users," J. Acoust. Soc. Am. **123**, 1522-1543.

O'Neill, E. R., Kreft, H. A., and Oxenham, A. J. (**2019**). "Speech perception with spectrally non-overlapping maskers as measure of spectral resolution in cochlear implant users," J. Assoc. Res. Otolaryngol. **20**, 151-167.

Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1987**). "An efficient auditory filterbank based on the gammatone function," in *a meeting of the IOC Speech Group on Auditory Modelling at RSRE*.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Ruggero, M. A., and Rich, N. C. (**1983**). "Chinchilla auditory-nerve responses to low-frequency tones," J. Acoust. Soc. Am. **73**, 2096-2108.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science. **270**, 303-304.

Shannon, R. V., Zeng, F. G., and Wygonski, J. (**1998**). "Speech recognition with altered spectral distribution of envelope cues," J. Acoust. Soc. Am. **104**, 2467-2476.

Singh, S., Kong, Y. Y., and Zeng, F. G. (**2009**). "Cochlear implant melody recognition as a function of melody frequency range, harmonicity, and number of electrodes," Ear and Hearing, **30**, 160-168.

Skinner, M. W., Holden, L. K., Holden, T. A., Dowell, R. C., Seligman, P. M., Brimacombe, J. A., and Beiter, A. L. (**1991**). "Performance of postlinguistically deaf adults with the wearable speech processor (WSP III) and mini speech processor (MSP) of the Nucleus multi-electrode cochlear implant," Ear. Hear. **12**, 3-22.

Skinner, M. W., Holden, L. K., Whitford, L. A., Plant, K. L., Psarros, C., & Holden, T. A. (**2002**). "Speech recognition with the nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults," Ear and hearing, **23**, 207-223.

Strydom, T. and Hanekom, J. J. (**2011**). "An analysis of the effects of electrical field interaction with an acoustic model of cochlear implants," J. Acoust. Soc. Am. **129**, 2213-2226.

Souza, P. and Rosen, S. (**2009**). "Effects of envelope bandwidth on the intelligibility of sine-and noise-vocoded speech". J. Acoust. Soc. Am., **126**(2), 792-805.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Svirsky, M. A., Capach, N. H., Neukam, J. D., Azadpour, M., Sagi, E., Hight, A. E., ... & Fitzgerald, M. B. (**2021**). Valid acoustic models of cochlear implants: One size does not fit all. Otology & Neurotology, **42**, S2-S10.

Tong, Y. C., Clark, G. M., Seligman, P. M., and Patrick, J. F. (**1980**). Speech processing for a multiple-electrode cochlear implant hearing prosthesis. J. Acoust. Soc. Am, **68**(6), 1897-1899.

van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**1999**). "Intensity discrimination of Gaussian-windowed tones: indications for the shape of the auditory frequency-time window," J. Acoust. Soc. Am. **105**, 3425-3435.

Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M. (**2000**). "Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system," Ear. Hear. **21**, 608-624.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., Rabinowitz, W. M. (**1991**). "Better speech recognition with cochlear implants," Nature. **352**, 236-238.

Wong, L. L., Soli, S. D., Liu, S., Han, N., and Huang, M. W. (**2007**). "Development of the Mandarin Hearing in Noise Test (MHINT)," Ear. Hear. **28**, 70S-74S.

Xi, X., Ching, T. Y., Ji, F., Zhao, Y., Li, J. N., Seymour, J., Hong, M. D., Chen, A. T., and Dillon, H. (**2012**). "Development of a corpus of Mandarin sentences in babble with homogeneity optimized via psychometric evaluation," Int. J. Audiol. **51**, 399-404.

Xu, L., Thompson, C. S., and Pfingst, B. E. (**2005**). “Relative contributions of spectral and temporal cues for phoneme recognition,” J. Acoust. Soc. Am., **117**, 3255-3267.

Zeng, F. G. (**2017**). “Challenges in improving cochlear implant performance and accessibility”. IEEE Transactions on Biomedical Engineering, **64**, 1662-1664.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted March 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Zeng, F. G., Rebscher, S., Harrison, W., Sun, X., & Feng, H. (**2008**). "Cochlear implants: system design, integration, and evaluation". IEEE Reviews in Biomedical Engineering, **1**, 115-142.