medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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[b)]

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

Thomas Lu and Fan-Gang Zeng[c)]

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

Revision date: May 5, 2022

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

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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 May 9, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 (CIs) allow for studies of perceptual performance with minimized effects of large CI individual variability. Different from conventional simulations using continuous sinusoidal or noise carriers, the present study employs Gaussian-enveloped tones (GETs) to simulate pulsatile stimulation in modern CIs. Subject to the time-frequency uncertainty principle, the GET has a well-defined tradeoff between its duration and bandwidth. Two types of GET vocoders were implemented and evaluated in normal-hearing listeners. In the first implementation, constant 100-Hz GETs were used to minimize within-channel temporal overlap while different GET durations were used to simulate electric channel interaction. This GET vocoder could produce vowel and consonant recognition similar to actual CI performance. In the second implementation, 900-Hz/channel pulse trains were directly mapped to 900-Hz GET trains to simulate 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 present results suggest that the pulsatile GETs can be used as alternative vocoders to simulate speech perception with modern CIs.

JASA abstract word limit: 200.

Current word number: 184.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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 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 (Dudley, 1939). 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.

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 (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), due not only to crude F0 and formant extraction methods (i.e., zero-crossing) at that time, but, more importantly, to complicated interactions between sound frequency and electric pitch, 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 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

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

two to four times the highest envelope frequency) pulse train, which is then delivered to a corresponding electrode in an interleaved fashion in which no two electrodes fire simultaneously (Wilson et al., 1991; Skinner et al., 2002). 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 (Zeng et al., 2008).

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 (e.g., cognitive differences, demographic variables, and electrode-neuron interface), which may confound or mask the relative importance of speech processing parameters, e.g., Skinner et al. (2002). Second, acoustic simulations allow the evaluation of relative contributions of different cues to auditory and speech perception, e.g., Xu et al. (2005); Singh et al. (2009). 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 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., Shannon et al. (1998). 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 (Blamey et al., 1984a; Blamey et al., 1984b), filtered harmonic complex tones (Deeks and Carlyon, 2004), and pulse-spread harmonic complexes (Hilkhuysen and Macherey, 2014; Mesnildrey *et al.*, 2016). However, there are limitations to those methods in simulating some important features in

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

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 using continuous carriers to simulate some CI speech processing strategies, e.g., *n*-of-*m*, in which the low-energy bands are abandoned to produce temporally separated envelopes.

Here we identified the Gabor atom (Gabor, 1947), 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 to study a wide range of auditory phenomena in normal hearing or hearing-impaired listeners, e.g., temporal gap detection (Schneider et al., 1994; Trehub et al., 1995), intensity discrimination (Baer et al., 1999; van Schijndel et al., 1999; Baer et al., 2001; Nizami et al., 2001), simultaneous and non-simultaneous masking (Laback et al., 2011; Laback et al., 2013), interaural timing difference (ITD) (Buell and Hafter, 1988), and cortical encoding of pulsatile stimulation (Lu and Wang, 2000; Lu et al., 2001; Johnson et al., 2017). More recently, GET train has been used to simulate some basic tasks on binaural hearing with CIs, e.g., sound localization (Goupell et al., 2010; Jones et al., 2014), lateralization (Ehlers et al., 2016), binaural masking level differences (Lu et al., 2010), temporal weighting of ITD and interaural level difference (ILD) (Brown and Stecker, 2010), effects of electrode place mismatch on binaural cues (Goupell et al., 2013; Kan et al., 2013), and effects of temporal quantization on ITD discrimination (Dieudonne et al., 2020).

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)

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

(Gabor, 1947; Feichtinger and Strohmer, 1998; Gardner and Magnasco, 2006). 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. In much of the existing literature, conventional channel-vocoders with eight channels using continuous noise or sine-wave carriers were used to replicate the sound of 12-24 channel CIs. The main reason is the performance of eight-channel vocoders in normal-hearing listeners usually matches the better performance of actual CI users (Winn and Nelson, 2021).

This study introduces a novel GET vocoder and demonstrates its potential for simulating CI speech perception. In the following sections, the implementation and theory of the proposed GET vocoders are introduced in detail; then two separate experiments of speech perception, each with a different type of GET vocoder, are used to demonstrate the potential of the novel pulsatile vocoders on CI speech perception simulation. Specifically, the first GET (Lu et al., 2007; Goupell et al., 2010) is a naïve type using non-interleaved 100-pps (pulse per second) GET trains as carriers to study the effect of current interaction among channels. The second GET (Meng et al., 2018; Kong et al., 2019) is an advanced type that can directly map 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).

The pulsatile GET vocoder can replicate the temporal (pulsatile), intensity (compressed and quantized), and spectral (maxima-selected) features of an actual CI strategy. Furthermore, current spread at individual electrodes can be simulated by changing the GET bandwidth through

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

the pulse duration parameter. We hypothesized 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.

II. GET THEORY AND VOCODER ALGORITHMS

A. GET Theory

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 (Baer et al., 1999). Its Fourier transform is:

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

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 by a sinusoidal carrier:

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

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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 $s(t)$ 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. 1 illustrates both waveform (a) and spectrum (b) of a unit-amplitude Gaussian-enveloped single pulse (i.e., $a = 1$ in Eq. 4). The carrier frequency f_c is 5 kHz. 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 Eq. (1), i.e.,

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

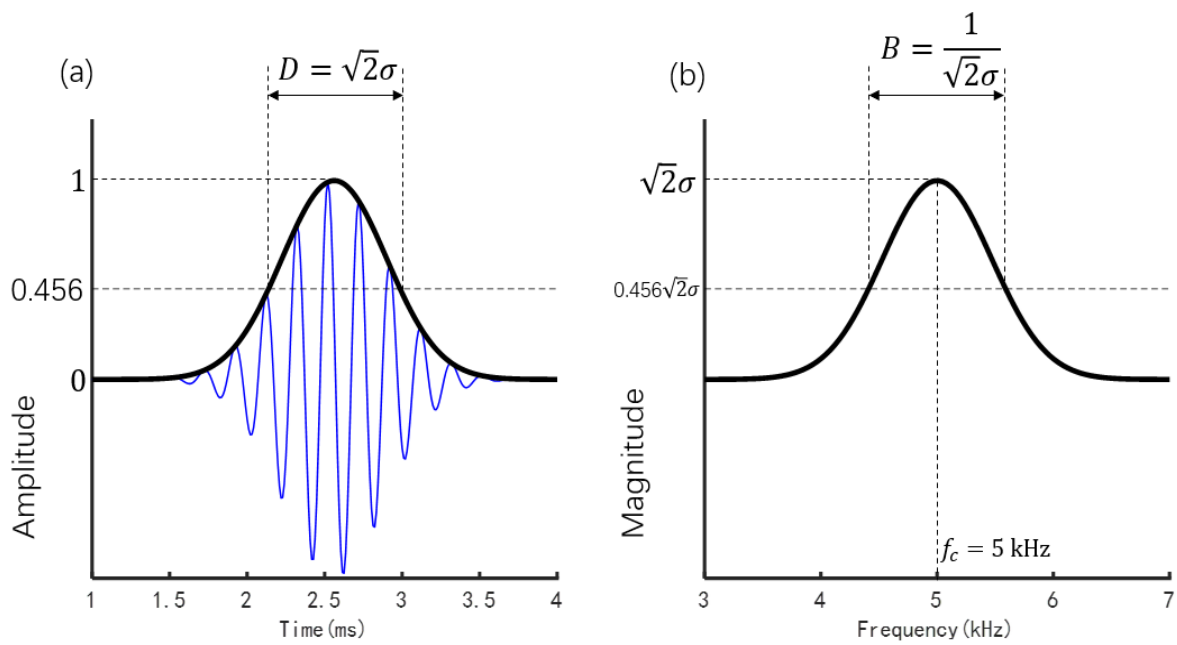

FIG. 1. (Color online) A unit-amplitude single pulse with Gaussian-shaped envelope (black line) in both the time (a) and frequency (b) domains. The carrier frequency is 5 kHz (the blue waveform in the left panel and the frequency with maximum amplitude in the right panel). The σ equals to $3/f_c$ = 0.6 ms in Eq. (1), producing an effective duration of 0.85 ms and an effective bandwidth of 1.2 kHz.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

Using the GET defined by Eq. 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 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 $\sigma = \frac{l}{f_c} = \frac{2}{f_c}, \frac{3}{f_c}$, and $\frac{4}{f_c}$, 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

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

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

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.

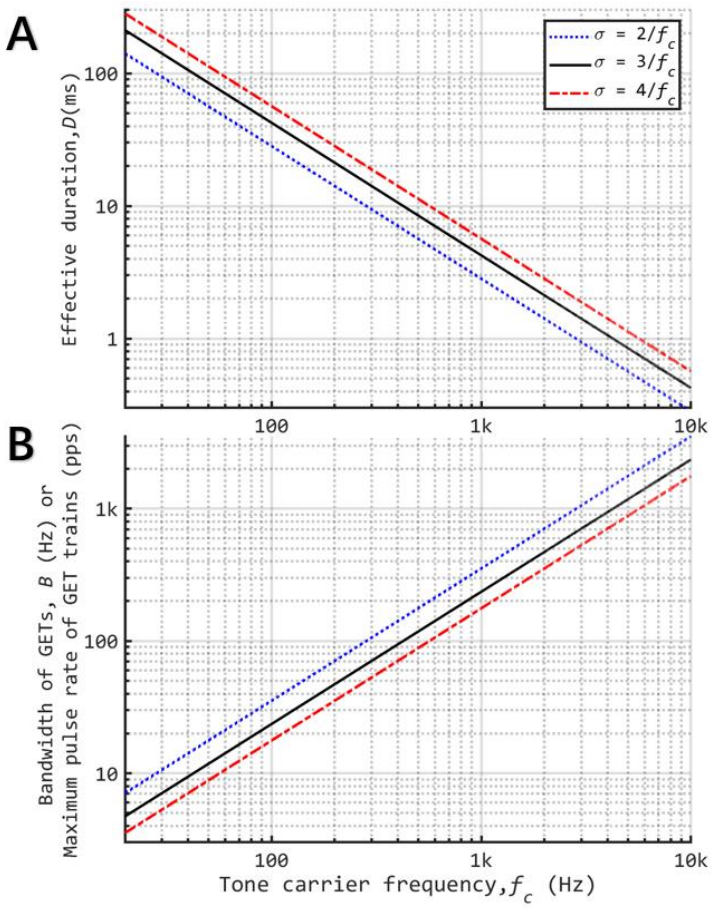

FIG 2. (Color online) The relationship between the tone carrier frequency and the effective duration $D = \sqrt{2}\sigma$ (see Panel **A**) or effective bandwidth $B = 1/D$ (see Panel **B**) of Gaussian-enveloped tones (GETs). All axes are logarithmically scaled. The σ was assumed to be $2/f_c$, $3/f_c$, 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.

The temporal envelopes delivered in electric speech stimuli are often temporally separated across channels in many CI strategies, as nature 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 compression 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

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

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:

$$f_{c_low} = \frac{l}{\sigma} = \frac{\sqrt{2}l}{D_{max}} = \frac{\sqrt{2}l}{d_{gap}} \tag{7}$$

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 (Mehta et al., 2020). For a single GET (defined by Eq. 4), its bandwidth is determined by its duration due to the time-frequency uncertainty principle. Therefore, 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. 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 GETs are discussed above and should be taken into consideration during algorithm design and experiments of CI simulations with GETs.

B. Vocoder Algorithm Frameworks

Fig. 3A shows the conventional acoustic simulation of CI using either noise (Shannon et al., 1995) or sine-wave vocoders (Dorman et al., 1997). The output filters can be used to control the

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

current spread, but no temporal separation feature (e.g., pulsatile timing and temporally separated envelope) can be simulated.

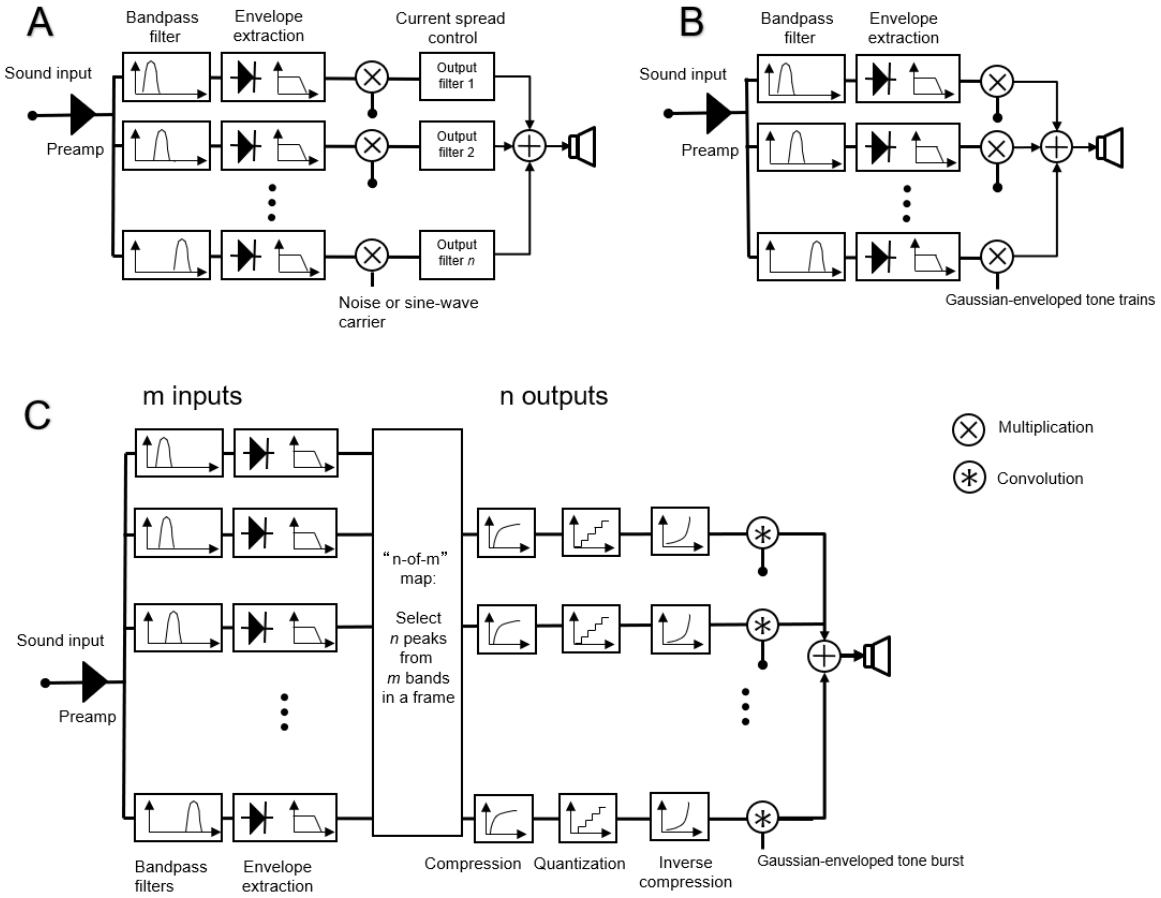

FIG. 3. Block diagrams of conventional channel vocoder (A), the first (B) and second (C) types of GET vocoders. The pulsatile vocoders are using GETs as carriers (the first type; used in Exp. 1) or using a single GET as an impulse response (the second type; used in Exp. 2). The front-end pre-emphasis, bandpass filter, and envelope extraction can be implemented either in the temporal or spectral domain.

The first GET vocoder was proposed by Lu et al. (2007) (see Fig. 3B) and subsequently used in a sound localization study (Goupell et al., 2010). As a naïve implementation, this approach replaces the conventional continuous carriers with pulsatile GET carriers. To demonstrate the effects of current interaction realized by different GET durations, vowel and consonant perception with non-interleaved 100-pps GET carriers was measured in Experiment 1 (Section III).

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

The second GET vocoder was proposed by Meng et al. (2018) (see Fig. 3C). Compared to the naïve implementation of the first type, the second GET vocoder hypothesized that a direct mapping from individual CI electric pulses to individual GET acoustic pulses could transmit similar speech information in both modes 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 Experiment 2 (Section IV), 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.

The front-end processing stages of the three methods in Fig.3 share the same blocks of band-pass filters and envelope extraction, e.g., in a traditional temporal envelope-based continuous interleaved sampling (CIS) (Wilson et al., 1991) or ACE strategy (Vandali et al., 2000). Details about the implementations of the two types of GET vocoders are provided in the following two experiment sections.

III. EXPERIMENT 1: SIMULATION OF CURRENT SPREAD

A. Rationale

Experiment 1 was designed to study vowel and consonant speech perception with the first type of GET vocoder (Lu et al., 2007; Goupell et al., 2010) using non-interleaved GET carriers (where the GET centers for all channels are in alignment with each other in each frame). The

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

interleaved sampling feature of modern CI strategies was not considered. A low pulse rate of 100 pps, which is much lower than the standard clinical rate (e.g., 900 pps or faster), was used in this experiment to minimize the within-channel inter-pulse temporal interaction. The primary purpose of this experiment is to examine the effects of current spread stimulated by manipulating the GET duration based on the uncertainty principle.

There is a significant difference in simulating the spread of excitation between the conventional vocoder (Shannon et al., 1995; Dorman et al., 1997) and the GET implementation (Lu et al., 2007). In the conventional simulation, the spread of excitation is manipulated by changing the filter type and the bandwidth of the synthesis band-pass filters at the vocoder output stage (Croghan and Smith, 2018). 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.

B. Methods

Five vocoders were used: three conventional vocoders - sine-wave, noise-separate, and noise-spread (Fig. 3A) - and two proposed vocoders incorporating the GET simulation -GET-separate and GET-spread (Fig. 3B).

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) (See Tab. I). The filtered signals were half-wave rectified and low-pass filtered (50 Hz 4th order Butterworth) to extract the envelope for each channel. This 50-Hz cutoff requires, in theory, at least a 100-Hz carrier to avoid aliasing.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

TABLE I. Cutoff frequencies of the band-pass filters in Exp. 1 according to a Greenwood map

Band number	Cutoff frequencies (Hz)
2	80, 1250, and 7999
4	80, 424, 1250, 3234, and 7999
8	80, 215, 424, 748, 1250, 2028, 3234, 5103, and 7999
16	80, 140, 215, 308, 424, 568, 748, 972, 1250, 1597, 2028, 2565, 3234, 4067, 5103, 6393, and 7999
32	80, 108, 140, 176, 215, 259, 308, 363, 424, 492, 568, 653, 748, 854, 972, 1103, 1250, 1414, 1597, 1801, 2028, 2282, 2565, 2881, 3234, 3628, 4067, 4556, 5103, 5713, 6393, 7152, and 7999

Synthesis processing for the conventional vocoders: For the sine-wave vocoder, a sine wave with a frequency centered at the corresponding analysis filtering band was used as the carrier. For the noise-separate vocoder, band-pass noise carriers were generated by passing white noise through filters that were the same as the analysis filters. The noise-separate vocoder provides upper-bound performance with a minimum of simulated electrode interaction. For the noise-spread vocoder, low-pass filters (4th order Butterworth) were used to pass white noise for generating low-pass 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 low-pass noises. Low-pass filters were chosen to represent severe interactions between channels (especially on the low-frequency side), 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 all channels.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

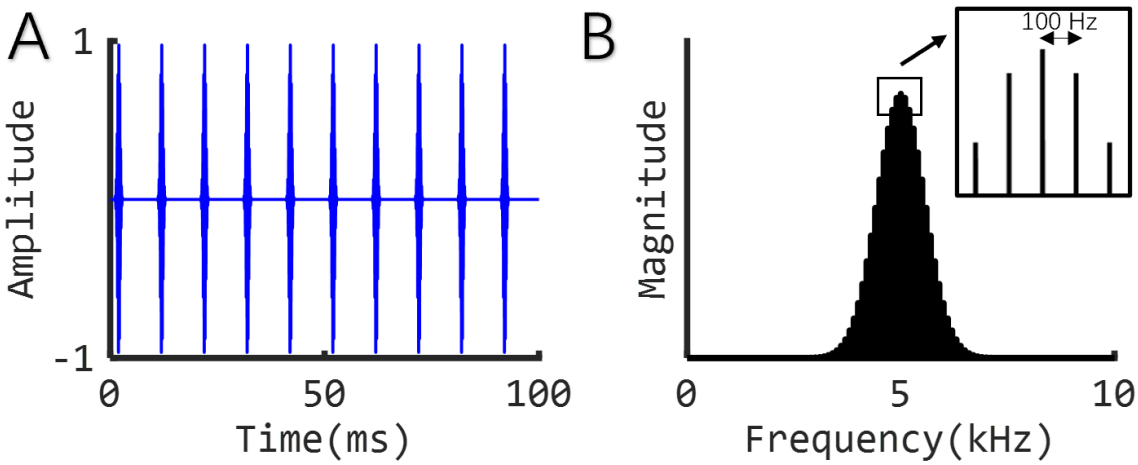

FIG. 4. (Color online) A 100-Hz pulse train, repeating a single pulse every 10 ms, in both the time (left panel) and frequency (right panel) domains. The parameters of the individual pulses are the same as those in Fig. 1.

Synthesis processing for the 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 a GET train. 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-separate vocoder, $D = \sqrt{2}\sigma = 7.0$ ms, while for the GET-spread vocoder, $D = \sqrt{2}\sigma = 1.2$ ms. Because the first experiment focused on the spread of excitation, the pulses among all channels were synchronized, meaning that the "interleaved sampling" feature was not simulated.

CI stimulation was simulated using the above five different vocoders, i.e., sine-wave, noise-separate, noise-spread, GET-separate, and GET-spread. The numbers of channels tested were 2, 4, 8, 16, and 32. There were 12 medial vowels and 14 medial consonants in the vowel and consonant tests, respectively. Fig. 5 provides an example of 16-channel vocoded stimuli for

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

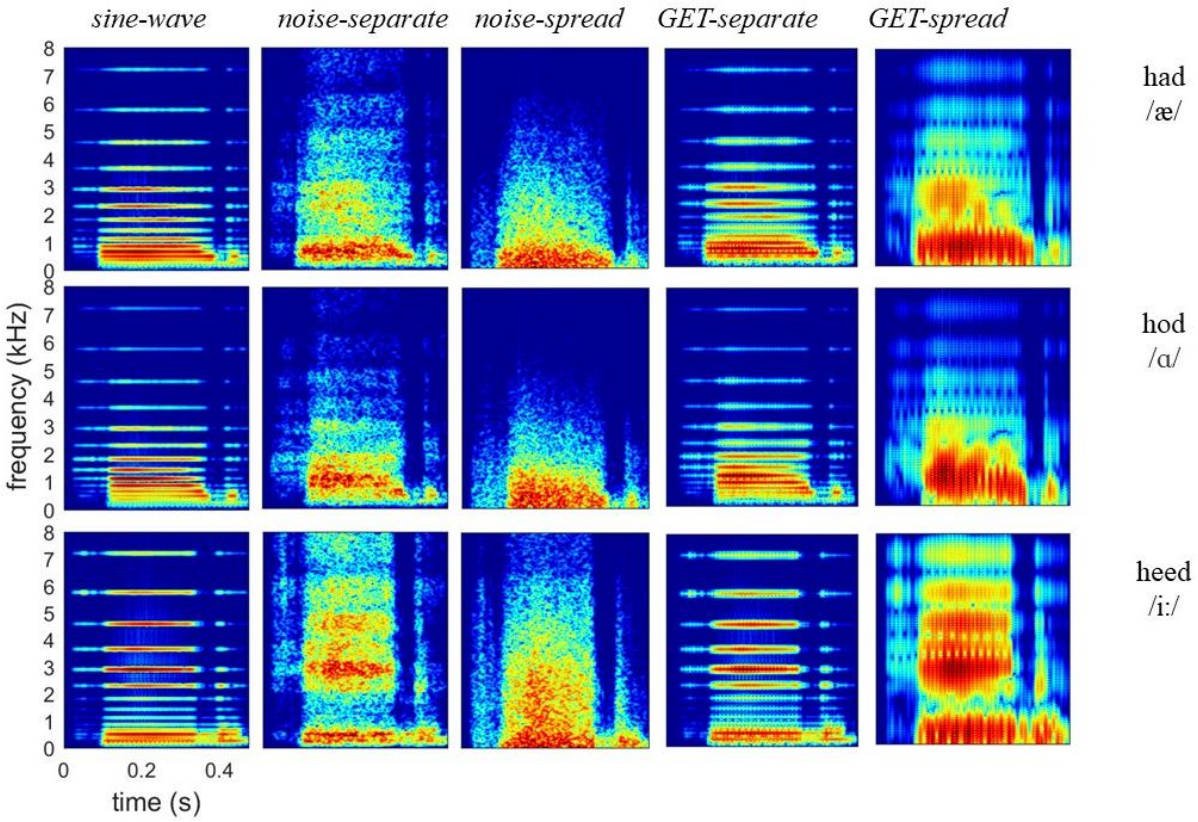

FIG. 5. (Color online) Spectrograms of three vowel stimuli encoded by the sine-wave, noise-separate, noise-spread, GET-separate, and GET-spread vocoders with 16 channels.

vowel tests. 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.

Seven 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).

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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. Results

Results are shown in Fig. 6. For the vowel test, the seven NH participants scored approximately 20% under all simulation conditions with two channels. Increasing the number of channels also improved 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. The noise-separate vocoder and GET-separate vocoder showed similar performance trends. When electrode interaction was simulated with overlapping filters, the subject performance showed a plateau near

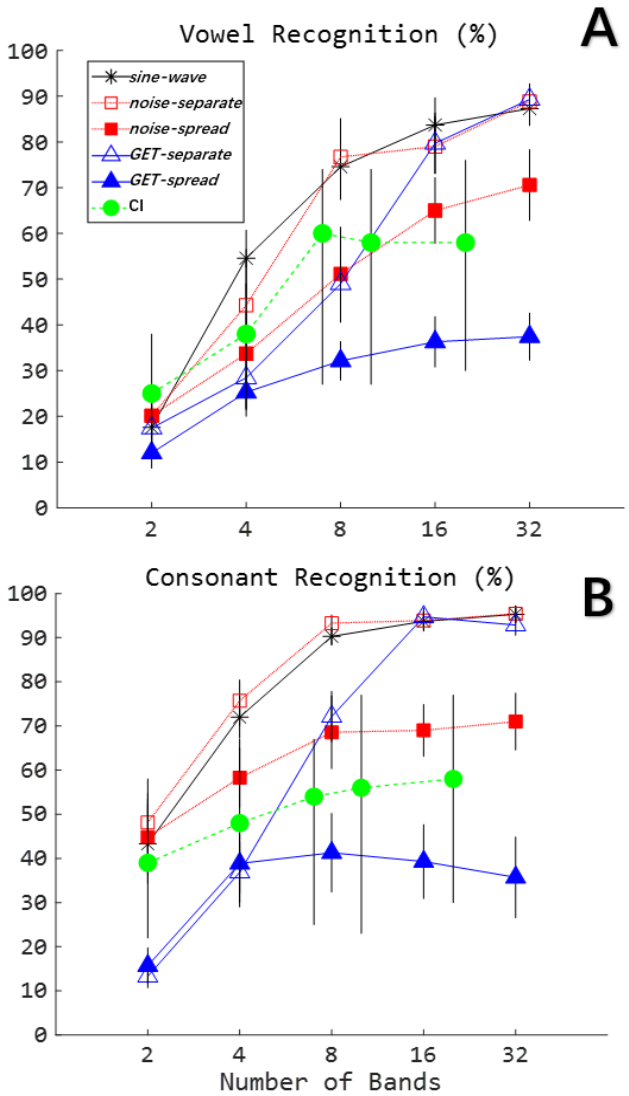

FIG. 6. (Color online) Vowel (**A**) and consonant (**B**) recognition as a function of number of bands (channels). Cochlear implant data is adapted from Friesen et al. (2001). Simulation data are averaged from seven normal hearing subjects listening to vocoded speech. For the simulation data, standard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

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

60% with noise-spread, similar to actual CIs. The GET-spread condition underperformed CI data in this case, saturating near 35% with eight channels.

Further, a two-way repeated-measures ANOVA with Geisser-Greenhouse correction was used to analyze the vowel simulation results with vocoder and number of bands as the main factors. The effect of vocoder ($F_{1.987,\ 11.92} = 49.87$, $p < 0.0001$), number of bands ($F_{2.018,\ 12.11} = 90.66$, $p < 0.0001$), and their interaction ($F_{3.890,\ 23.34} = 9.842$, $p < 0.0001$) were all significant. To further analyze these effects, multiple comparisons with Bonferroni corrections were implemented for each vocoder (to compare the five band numbers) and for each band number (to compare the five vocoders). Table II shows the results of multiple comparisons between different numbers of bands for each vocoder. Generally, there was a trend of better performance with more bands. Still, the mean scores were not significantly different for 8, 16, and 32 bands (the only exception was 8 vs. 32 with GET-separate). Table III shows the results of multiple comparisons between vocoders for each number of bands. Because at 2 and 4 bands most vocoder pairs showed no significant mean difference (the only exception was sine-wave vs. noise-spread at 4 number of bands with $p = 0.009$), the comparison results bands were not listed. GET-spread derived the lowest scores among the five vocoders at 16 and 32 bands, while GET-separate did not show significantly different mean scores from the other three vocoders. The sine-wave, noise-separate, and GET-separate vocoders did not show significantly different mean scores.

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

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

simulation results, consonant recognition scores were analyzed using the same statistical method as the above vowel data analysis. The effects of vocoder ($F_{1.404, 8.427} = 62.55$, $p < 0.0001$), number of bands ($F_{2.234, 13.40} = 379.0$, $p < 0.0001$), and their interaction ($F_{3.080, 18.48} = 10.88$, $p = 0.0002$) were all significant on consonant recognition. Results of multiple comparisons are shown in Table IV and V. The relative scores show similar trends as the results of multiple comparisons for vowel recognition (see Table II and III).

TABLE II. Results of multiple comparisons between vowel recognition scores with five band numbers for each of the five vocoders.

vocoder→	sine-wave		noise-separate		noise-spread		GET-separate		GET-spread	
number of bands pair↓	mean diff	*p*	mean diff	*p*	mean diff	*p*	mean diff	*p*	mean diff	*p*
2 vs. 4	*37.0	0.014	24.1	0.140	13.6	1.000	11.0	0.827	*13.3	0.019
2 vs. 8	**57.0	0.001	**56.6	0.002	31.0	0.085	31.6	0.082	*20.1	0.048
2 vs. 16	***66.1	<0.001	***58.9	<0.001	**44.9	0.001	***62.3	<0.001	*24.3	0.005
2 vs. 32	***69.7	<0.001	***68.7	<0.001	**50.4	0.001	***71.9	<0.001	*25.4	0.012
4 vs. 8	20	0.076	*32.4	0.019	17.4	0.057	20.6	0.209	6.9	0.564
4 vs. 16	*29.1	0.049	*34.7	0.023	***31.3	<0.001	**51.3	0.002	*11.0	0.034
4 vs. 32	*32.7	0.037	*44.6	0.011	***36.9	<0.001	***60.9	<0.001	12.1	0.296
8 vs. 16	9.1	0.199	2.29	1.000	13.9	0.147	30.7	0.058	4.1	1.000
8 vs. 32	12.7	0.288	12.1	0.696	19.4	0.059	*40.3	0.011	5.3	1.000
16 vs. 32	3.6	1.000	9.86	0.131	5.6	0.107	9.6	0.455	1.1	1.000

TABLE III. Results of multiple comparisons between vowel recognition scores with five vocoders for each of the three band numbers (8, 16, and 32).

number of bands→	8		16		32	
vocoder pair↓	mean diff	*p*	mean diff	*p*	mean diff	*p*
sine-wave vs. noise-separate	−2.1	1.000	4.7	1.000	−1.6	1.000
sine-wave vs. noise-spread	23.4	0.054	**18.7	0.009	16.7	0.120
sine-wave vs. GET-separate	25.6	0.065	4.0	1.000	−2.0	1.000
sine-wave vs. GET-spread	**42.4	0.002	***47.4	<0.001	***49.9	<0.001
noise-separate vs. noise-spread	*25.6	0.030	*14.0	0.025	*18.3	0.048
noise-separate vs. GET-separate	27.7	0.080	−0.7	1.000	−0.4	1.000
noise-separate vs. GET-spread	**44.6	0.004	***42.7	<0.001	***51.4	<0.001
noise-spread vs. GET-separate	2.1	1.000	−14.7	0.130	−18.7	0.102
noise-spread vs. GET-spread	19.0	0.459	**28.7	0.002	**33.1	0.004
GET-separate vs. GET-spread	16.9	0.156	**43.4	0.001	***51.9	<0.001

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

TABLE IV. Results of multiple comparisons between consonant recognition scores with five number of bands for each of the five vocoders

vocoder→	sine-wave		noise-separate		noise-spread		GET-separate		GET-spread	
number of bands pair↓	mean diff	*p*	mean diff	*p*	mean diff	*p*	mean diff	*p*	mean diff	*p*
2 vs. 4	*37.0	0.014	24.1	0.140	13.6	1.000	11.0	0.827	*13.3	0.019
2 vs. 8	**57.0	0.001	**56.6	0.002	31.0	0.085	31.6	0.082	*20.1	0.048
2 vs. 16	***66.1	<0.001	***58.9	<0.001	**44.9	0.001	***62.3	<0.001	**24.3	0.005
2 vs. 32	***69.7	<0.001	***68.7	<0.001	**50.4	0.001	***71.9	<0.001	*25.4	0.012
4 vs. 8	20.0	0.076	*32.4	0.019	17.4	0.057	20.6	0.209	6.9	0.564
4 vs. 16	*29.1	0.049	*34.7	0.023	***31.3	<0.001	**51.3	0.002	*11.0	0.034
4 vs. 32	*32.7	0.037	*44.6	0.011	***36.9	<0.001	***60.9	<0.001	12.1	0.296
8 vs. 16	9.1	0.199	2.3	1.000	13.9	0.147	30.7	0.058	4.1	1.000
8 vs. 32	12.7	0.288	12.1	0.696	19.4	0.059	40.3	0.011	5.3	1.000
16 vs. 32	3.6	1.000	9.9	0.131	5.6	0.107	9.6	0.455	1.1	1.000

TABLE V. Results of multiple comparisons between consonant recognition scores with five vocoders for each of the five band numbers

number of bands→	2		4		8		16		32	
vocoder pair↓	mean diff	*p*	mean diff	*p*	mean diff	*p*	mean diff	*p*	mean diff	*p*
sine vs. noi-sep	−4.9	1.000	−3.7	1.000	−3.0	1.000	−0.1	1.000	−0.1	1.000
sine vs. noi-spr	−1.6	1.000	13.7	0.267	21.7	0.292	*24.7	0.016	*24.3	0.022
sine vs. GETsep	30.0	0.065	**35.1	0.002	18.1	0.096	−1.0	1.000	2.4	1.000
sine vs. GETspr	27.6	0.108	**33.1	0.010	**49.0	0.006	**54.4	0.002	**59.6	0.002
noi-sep vs. noi-spr	3.3	1.000	17.4	0.157	24.7	0.111	*24.9	0.021	*24.4	0.038
noi-sep vs. GETsep	**34.9	0.007	**38.9	0.007	21.1	0.115	−0.9	1.000	2.6	1.000
noi-sep vs. GETspr	*32.4	0.018	**36.9	0.002	**52.0	0.007	**54.6	0.003	**59.7	0.004
noi-spr vs. GETsep	*31.6	0.043	**21.4	0.003	−3.6	1.000	**−25.7	0.010	*−21.9	0.040
noi-spr vs. GETspr	29.1	0.078	**19.4	0.003	27.3	0.065	**29.7	0.006	***35.3	<0.001
GETsep vs. GETspr	−2.4	1.000	−2.0	1.000	*30.9	0.034	**55.4	0.002	**57.1	0.003

The current results suggest that the first type GET vocoder is feasible to simulate speech perception with CIs, and the CI current spread also could be simulated by manipulating durations of GETs. 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 good simulation performance in that the actual CI data fell in the range between the separate and spread versions of the GETs.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

IV. EXPERIMENT 2. SIMULATION OF THE N-OF-M STRATEGY ACE

A. Rationale

Some essential features of modern CI processing, including interleaved sampling, maxima selection, amplitude compression and quantization, are omitted in not only conventional continuous-carrier vocoders but also in the first type GET vocoder as used in Experiment 1. All of these features may influence speech perception. According to the analysis in Section II, GETs could be used to simulate them. The second type of GET vocoder (Meng et al., 2018; Kong et al., 2019) is introduced here in detail, and a battery of speech recognition tasks was carried out to demonstrate its performance in Experiment 2. The experiment objective was to demonstrate the potential of CI speech perception simulation with a GET vocoder involving all of the above-mentioned essential features. The ACE strategy with 900-pps pulse rate was simulated by this advanced GET vocoder.

B. 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. 7A 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 (Greenwood, 1990). The cutoff frequencies are 150, 271, 439, 672, 994, 1439, 2057, 2911, 4094,

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

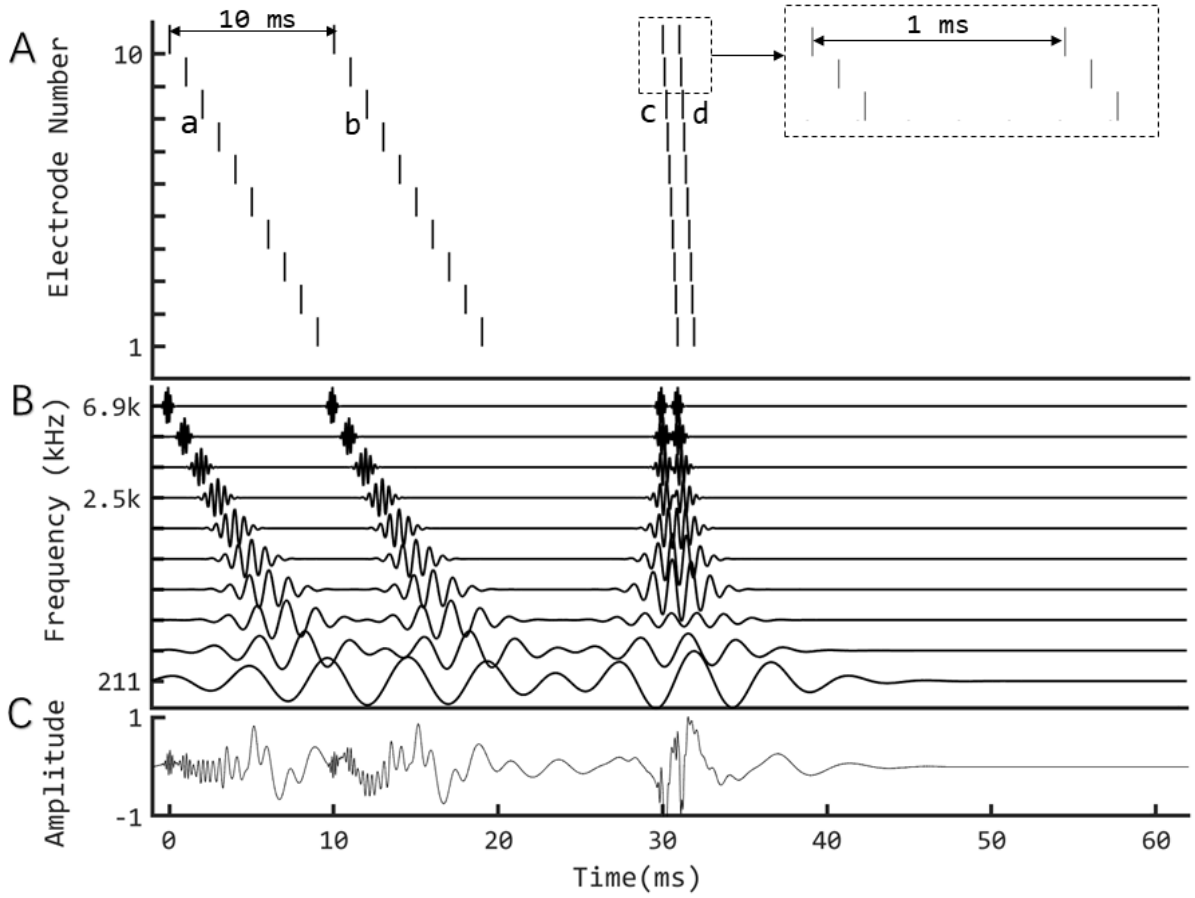

FIG. 7. Mapping a CI electrodogram to a sound using the second type GET vocoder. **A**. An artificial 10-channel CI electrodogram, including two pulse sweeps with a 10-ms difference between **a** and **b**, as well as two additional sweeps with a 1-ms difference between **c** and **d,** corresponding 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.

5732, and 8000 Hz. Then, a band-specific GET was generated in this demonstration by setting the parameters in Eq. 1 as $a = 1$, $t_0 = 0$, and

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

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

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

$$D = \sqrt{2}\sigma = \frac{2\sqrt{2}}{f_c} \tag{9}$$

and a 6.82-dB bandwidth of

$$B = \frac{1}{D} = \frac{\sqrt{2}}{4} f_c \tag{10}$$

Then the acoustic GET train at the kth channel in Fig. 7B is derived by

$$p_{a,k}(t) = (p_{e,k}(t) * e^{-\frac{\pi t^2}{2\sigma^2}}) \cdot \sin(2\pi f_c t + \varphi_0) \tag{11}$$

where $p_{e,k}(t)$ and $p_{a,k}(t)$ denotes the electric and acoustic pulse trains in Fig. 7A and 7B, respectively, "$*$" denotes a convolution calculation, 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. 7B shows the 10-channel GET trains, which have temporally separated waveforms for high-frequency channels, but overlapping waveforms for low-frequency channels. Fig. 7C 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. 7B, 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.

C. 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 ACE strategy which is the current default strategy in Nucleus cochlear implants (Vandali et

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

al., 2000), can be converted to vocoded sounds. 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 (Kong et al., 2019), 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 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 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 cutoffs, 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 (Ghosh et al., 2022) was used to convert input sounds into electrodograms. An inverse-mapping function was used to transfer the electric current value of each electric pulse in the electrodogram to an envelope power value. Single-

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

sample pulse trains from each band were "convolved" with a Gaussian function with $\sigma = 3/f_c$. 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. In the theory and framework analysis in Section II, 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. The output was used to 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 Eqs. 12 and 13, respectively:

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

and

$$L_a = \frac{1}{2.72\alpha}(e^{L_e}(1+\alpha) - 1) \tag{13}$$

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 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 Eqs. 12 and 13, 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. Equation 12 is directly based on the default setting of the acoustic-to-electric compression function in ACE. It was hypothesized that GETlargeDR could

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

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 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). 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 (Souza and Rosen, 2009). Therefore, Sin250 was expected to be better than Sin125.

In Fig. 8, 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.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

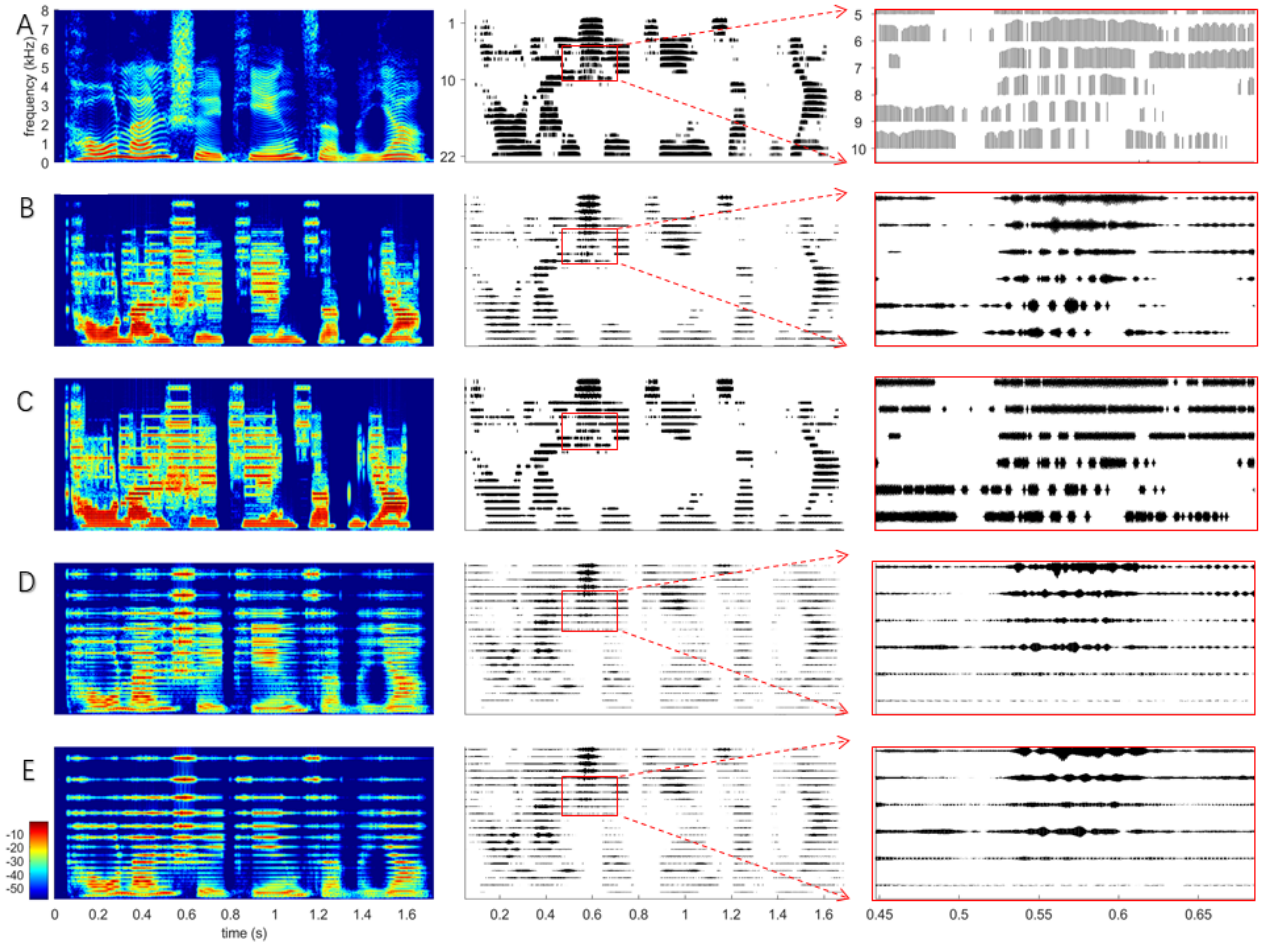

FIG. 8. (Color online) Speech stimulus demonstrations for the ACE-GET simulation experiment. Left: Spectrogram; middle: band-specific signal; right: zoom in of the boxed signals. **A**. Spectrogram and ACE electrodogram of a clear sentence of speech. **B-E**. Spectrogram and band-specific waveforms of vocoded speech using two GET vocoders (GETlargeDR, and GETsmallDR) and two conventional sine-wave vocoders (Sin250 and Sin125), respectively.

D. 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 GETlargeDR, and Group 2 used Sin125 and GETsmallDR. Three open-set Mandarin Chinese recognition tasks were

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

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 our previous experiments (Meng et al., 2019) 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. 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) strictly, 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 sentence list was used. Seven CI participants with various hearing histories were also tested for comparison (See Table VI).

We had three subject groups, two of which were NH listeners each using two different vocoders. A mixed model was used to assess the repeated measures within subjects as well as

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

independent measures between subjects. The 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 criterion of 0.05.

TABLE VI. Detailed information of the 7 CI participants in the speech in reverberation test

Subject	Gender	Age (yr)	CI Experience (yr)	CI Processor	Etiology
C14	F	41	12	CP810	Drug induced
C23	M	31	11	CP810	Sudden deafness
C30	M	13	10	Freedom	LVAS
M5	M	18	15	OPUS-2	Virus infection
C16	F	25	2	Freedom	Unknown
M17	F	18	5	OPUS-2	Genetic
M16	F	17	7	OPUS-2	Unknown

E. Results

The results with the four 22-channel vocoders, i.e., GETlargeDR, GETsmallDR, Sin250, and Sin125 are shown in Fig. 9.

For the TCT test (Fig. 9A), a significant decreasing trend was found from Sin250 (mean = 16.1 syllables/sec), 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.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

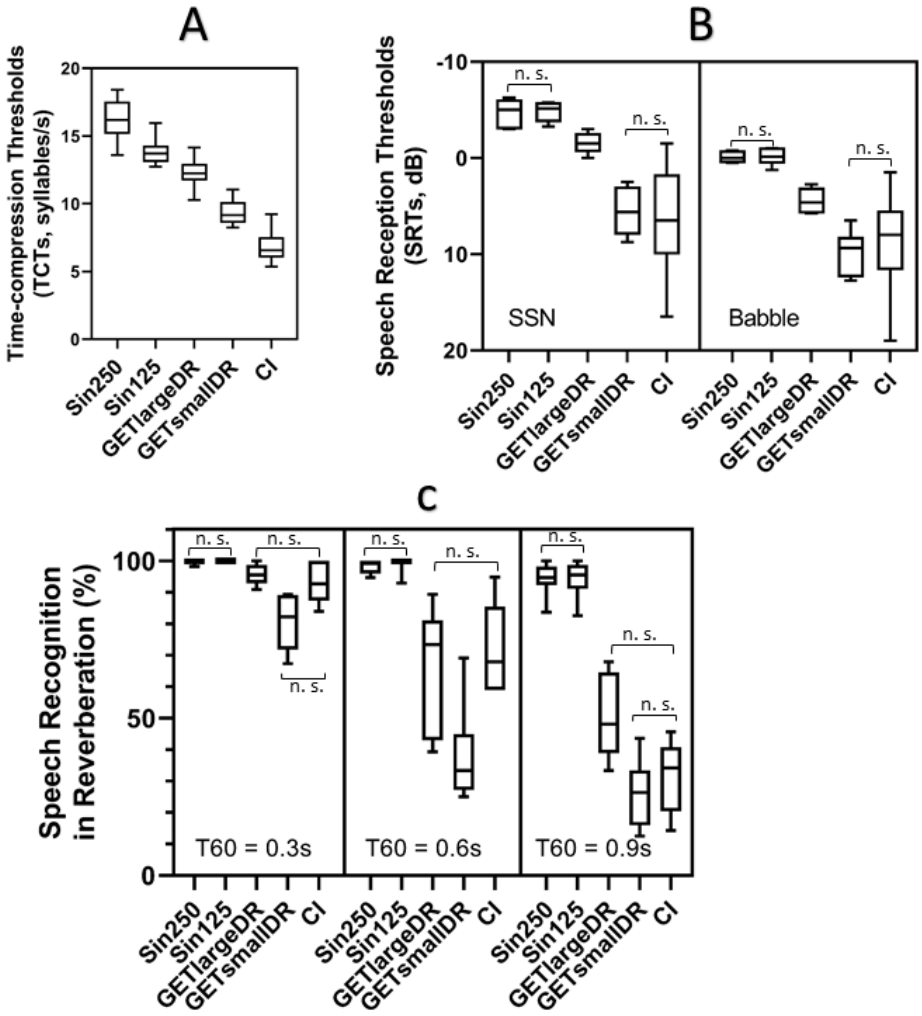

FIG. 9. Results from three speech recognition tasks with two 22-channel sine-wave vocoders (Sin250: 250 Hz cut-off envelope; Sin125: 125 Hz cut-off envelope) and two GET vocoders (GETlargeDR and GETsmallDR; their difference is only in the intensity dynamic range, i.e., 32.7 dB and 5.3 dB for GETlargeDR 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.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

For the SRT test (Fig. 9B), 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. 9C), all vocoders and the actual CI condition showed a significant trend of decreased recognition scores when the reverberation time increased. However, the 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 participants' 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.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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 all three tasks, GET vocoders were able to simulate actual CI performance more closely than sine vocoders. 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. 9A). In the SRT-in-noise test, GETsmallDR and CI produced comparable performance (Fig. 9B). 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. 9C).

V. DISCUSSION

Sounds are transmitted through air as continuous compression waves, but they are encoded by discrete spikes in the neural system and by pulsatile electric stimuli in CIs. Vocoders have been developed to simulate the signal processing and sound perception of CIs. However, the pulsatile feature, which is acknowledged as critical to the success of modern CIs, has not been simulated until now by the most widely used noise and sine-wave excited vocoder (Shannon et al., 1995; Dorman et al., 1997). Some studies have proposed pulsatile vocoders using filtered carriers with strong periodicities including noise burst (Blamey et al., 1984a; Blamey et al., 1984b) and complex tones (Deeks and Carlyon, 2004; Hilkhuysen and Macherey, 2014; Mesnildrey et al., 2016). Instead of using filtered carriers, some CI manufacturers have provided software to directly map electrodograms to vocoded sounds (Ausili et al., 2019; Stam et al., 2019). In this study, a GET-based vocoder was proposed, theoretically analyzed, and evaluated for its performance on CI speech perception simulation.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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 can be used to simulate a "perceivable" atom of sound, which can be traced back to Gabor (1947). 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 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.

An inherent limitation with the GETs is the tradeoff between temporal duration and spectral bandwidth. Shortening the GET duration increases the spectral bandwidth, which introduces temporal or spectral overlaps between different GETs, especially at low frequencies (see Fig. 7 and related text). 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.

B. Speech perception with GET vocoders

In this study, two types of GET vocoders (Fig. 3B&C) were proposed to simulate different aspects of CI processing (Lu et al., 2007; Meng et al., 2018). The first GET vocoder simply replaced the continuous noise or sine-wave carriers in conventional vocoders by a new type of carrier, or GET train. In the first implementation (Fig. 3B), 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. Spread of excitation is an important factor underlying the poor- and large-variance performance for CI participants (Fu and Nogaki, 2005; Bingabr et al., 2008; Strydom and Hanekom, 2011; Grange et al., 2017; O'Neill et al., 2019; Mehta et al., 2020). Different from the noise-or sine-vocoders that produced

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

performance better than actual CI performance even in the case of the severe channel interaction (i.e., using the low-pass filtered noise carriers), the GET vocoder produced a wide range of vowel and consonant recognition performance encompassing actual CI performance (Fig. 6). One limitation in this experiment was that the spectral spread simulated by GET vocoders at low frequency channels might be influenced by the sparsity of the electric pulses. For example (see Fig.7), at the lowest frequency channel, temporal overlap happens between two GETs and the bandwidth of the two overlapped GETs is narrower than an isolated GET. Fortunately, due to the sparse nature of speech signal and narrower GET durations at higher channels, the effects of this limitation should be limited. Another limitation of Experiment 1 was that all vocoders used a 50-Hz envelope cutoff frequency, which was lower than real CIs.

The second vocoder 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 produced a mean and range of speech in noise recognition performance similar to that of actual CI users. In particular, the wider dynamic range simulated better CI performance (Fig. 9). Future studies are 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 (DiNino et al., 2016; Winn, 2020; Bance et al., 2022).

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

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

intensity (or envelope) information. This means that the conventional vocoders can be treated as special cases of GET vocoders.

The MATLAB source code of the GET vocoder for the ACE strategy is provided for academic research purposes[2]. 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.

V. CONCLUSION

This study indicates that pulsatile simulation of speech, which is a key to the success of modern CI and has been omitted in previous vocoders, could be realized by using the proposed GET vocoders. The main conclusions include:

(1) The time-frequency uncertainty principle empowers and imposes constraints on using GETs for CI simulation;

(2) Many features of modern CIs including pulsatile timing, current spread, *n*-of-*m* maxima selection, dynamic compression could be implemented in GET vocoders and then used to derive similar sentence recognition performance to actual CI users;

(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 direct pulse-to-pulse mapping.

Further experiment studies (e.g., in phoneme confusion patterns) are warranted to systematically examine the performance of GET simulation.

[2] Currently as an attachment of the submission and will be open at a permanent website before the final version if the manuscript could be accepted for publication in JASA.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

ACKNOWLEDGMENTS

We thank all the participants in these experiments. J. Carroll and S. Tiaden helped collect the data in Experiment 1. Fanhui Kong and Yulong 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.). Thanks to Drew Cappotto for proof-reading this article.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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

Ausili, S. A., Backus, B., Agterberg, M. J. H., van Opstal, A. J., and van Wanrooij, M. M. (**2019**). "Sound localization in real-time vocoded cochlear-implant simulations with Normal-Hearing Listeners," Trends. Hear. **23**, 2331216519847332.

Baer, T., Moore, B. C. J., and Glasberg, B. R. (**1999**). "Detection and intensity discrimination of Gaussian-shaped tone pulses as a function of duration," J. Acoust. Soc. Am. **106**, 1907-1916.

Baer, T., Moore, B. C. J., and Marriage, J. (**2001**). "Detection and intensity discrimination of brief tones as a function of duration by hearing-impaired listeners," Hear. Res. **159**, 74-84.

Bance, M., Brochier, T., Vickers, D., and Goehring, T. (**2022**). "From microphone to phoneme: an end-to-end computational neural model for predicting speech perception with cochlear implants," IEEE Trans Biomed Eng. Early Access.

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., and Clark, G. M. (**1984a**). "Speech processing studies using an acoustic model of a multiple-channel cochlear implant," J. Acoust. Soc. Am. **76**, 104-110.

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

Brown, A. D., and Stecker, G. C. (**2010**). "Temporal weighting of interaural time and level differences in high-rate click trains," J. Acoust. Soc. Am. **128**, 332-341.

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

Croghan, N. B. H., and Smith, Z. M. (**2018**). "Speech understanding with various maskers in cochlear-implant and simulated cochlear-implant hearing: effects of spectral resolution and implications for masking release," Trend. Hear. **22**. 2331216518787276.

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.

Dieudonne, B., Van Wilderode, M., and Francart, T. (**2020**). "Temporal quantization deteriorates the discrimination of interaural time differences," J. Acoust. Soc. Am. **148**, 815-828.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

DiNino, M., Wright, R. A., Winn, M. B., and Bierer, J. A. (**2016**). "Vowel and consonant confusions from spectrally manipulated stimuli designed to simulate poor cochlear implant electrode-neuron interfaces," J. Acoust. Soc. Am. **140**, 4404-4418.

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.

Ehlers, E., Kan, A., Winn, M. B., Stoelb, C., and Litovsky, R. Y. (**2016**). "Binaural hearing in children using Gaussian enveloped and transposed tones," J. Acoust. Soc. Am. **139**, 1724-1733.

Feichtinger, H. G., and Strohmer, T. (**1998**). *Gabor analysis and algorithms: theory and applications* (Birkhäuser, Boston).

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

Fu, Q. J., Zhu, M., and Wang, X. (**2011**). "Development and validation of the Mandarin speech perception test," J. Acoust. Soc. Am. **129**, EL267-273.

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

Gardner, T. J., and Magnasco, M. O. (**2006**). "Sparse time-frequency representations," Proc. Natl. Acad. Sci. U.S.A. **103**, 6094-6099.

Ghosh, R., Ali, H., and Hansen, J. H. L. (**2022**). "CCi-MOBILE: a portable real time speech processing platform for cochlear implant and hearing research," IEEE Trans. Biomed. Eng. **69**, 1251-1263.

Goupell, M. J., Majdak, P., and Laback, B. (**2010**). "Median-plane sound localization as a function of the number of spectral channels using a channel vocoder," J. Acoust. Soc. Am. **127**, 990-1001.

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 May 9, 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. S. L., and Bergfeld, S. (**2017**). "Cochlear implant simulator with independent representation of the full spiral ganglion," J. Acoust. Soc. Am. **142**, EL484-489.

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

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

Jones, H., Kan, A., and Litovsky, R. Y. (**2014**). "Comparing sound localization deficits in bilateral cochlear-implant users and vocoder simulations with normal-hearing listeners," Trend. Hear. **18**. 2331216514554574.

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 *23rd International Congress on Acoustics* (Aachen, Germany), pp. 3109-3112.

Laback, B., Balazs, P., Necciari, T., Savel, S., Ystad, S., Meunier, S., and Kronland-Martinet, R. (**2011**). "Additivity of nonsimultaneous masking for short Gaussian-shaped sinusoids," J. Acoust. Soc. Am. **129**, 888-897.

Laback, B., Necciari, T., Balazs, P., Savel, S., and Ystad, S. (**2013**). "Simultaneous masking additivity for short Gaussian-shaped tones: Spectral effects," J. Acoust. Soc. Am. **134**, 1160-1171.

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

Lu, T., Liang, L., and Wang, X. (**2001**). "Temporal and rate representations of time-varying signals in the auditory cortex of awake primates," Nat. Neurosci. **4**, 1131-1138.

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 May 9, 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.

Lu, T., and Wang, X. (**2000**). "Temporal discharge patterns evoked by rapid sequences of wide- and narrowband clicks in the primary auditory cortex of cat," J. Neurophysiol. **84**, 236-246.

Mehta, A. H., Lu, H., and 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**, 61-72.

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)* (IEEE), pp. 761-765.

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

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.

Nizami, L., Reimer, J. F., and Jesteadt, W. (**2001**). "The intensity-difference limen for Gaussian-enveloped stimuli as a function of level: Tones and broadband noise," J. Acoust. Soc. Am. **110**, 2505-2515.

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.

Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., and Lamb, M. (**1994**). "Gap detection and the precedence effect in young and old adults," J. Acoust. Soc. Am. **95**, 980-991.

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 Hear. **30**, 160-168.

Skinner, M. W., Holden, L. K., Holden, T. A., Dowell, R. C., Seligman, P. M., Brimacombe, J. A., Beiter, A. L. J. E., and Hearing (**1991**). "Performance of postlinguistically deaf adults with

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

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., and Holden, T. A. (**2002**). "Speech recognition with the nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults," Ear Hear. **23**, 207-223.

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

Stam, L., Goverts, S. T., and Smits, C. (**2019**). "Effect of cochlear implant n-of-m strategy on signal-to-noise ratio below which noise hinders speech recognition," J. Acoust. Soc. Am. **145**, EL417-422.

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.

Svirsky, M. A., Capach, N. H., Neukam, J. D., Azadpour, M., Sagi, E., Hight, A. E., Glassman, E. K., Lavender, A., Seward, K. P., Miller, M. K., Ding, N., Tan, C. T., and Fitzgerald, M. B. (**2021**). "Valid Acoustic Models of Cochlear Implants: One Size Does Not Fit All," Otol. Neurotol. **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**, 1897-1898.

Trehub, S. E., Schneider, B. A., and Henderson, J. L. (**1995**). "Gap detection in infants, children, and adults," J. Acoust. Soc. Am. **98**, 2532-2541.

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 Clarke, 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., and Rabinowitz, W. M. (**1991**). "Better speech recognition with cochlear implants," Nature **352**, 236-238.

Winn, M. B. (**2020**). "Accommodation of gender-related phonetic differences by listeners with cochlear implants and in a variety of vocoder simulations," J. Acoust. Soc. Am. **147**, 174-190.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

Winn, M. B., and Nelson, P. B. (**2021**). "Cochlear Implants," (Oxford University Press).

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. C., 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., Rebscher, S., Harrison, W. V., Sun, X., and Feng, H. (**2008**). "Cochlear Implants: System Design, Integration and Evaluation," IEEE Rev. Biomed. Eng. **1**, 115-142.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

TABLEs

TABLE I. Cutoff frequencies of the band-pass filters in Exp. 1 according to a Greenwood map

Band number	Cutoff frequencies (Hz)
2	80, 1250, and 7999
4	80, 424, 1250, 3234, and 7999
8	80, 215, 424, 748, 1250, 2028, 3234, 5103, and 7999
16	80, 140, 215, 308, 424, 568, 748, 972, 1250, 1597, 2028, 2565, 3234, 4067, 5103, 6393, and 7999
32	80, 108, 140, 176, 215, 259, 308, 363, 424, 492, 568, 653, 748, 854, 972, 1103, 1250, 1414, 1597, 1801, 2028, 2282, 2565, 2881, 3234, 3628, 4067, 4556, 5103, 5713, 6393, 7152, and 7999

TABLE II. Results of multiple comparisons between vowel recognition scores with five band numbers for each of the five vocoders.

vocoder→	sine-wave		noise-separate		noise-spread		GET-separate		GET-spread	
number of bands pair↓	mean diff	*p*	mean diff	*p*	mean diff	*p*	mean diff	*p*	mean diff	*p*
2 vs. 4	*37.0	0.014	24.1	0.140	13.6	1.000	11.0	0.827	*13.3	0.019
2 vs. 8	**57.0	0.001	**56.6	0.002	31.0	0.085	31.6	0.082	*20.1	0.048
2 vs. 16	***66.1	<0.001	***58.9	<0.001	**44.9	0.001	***62.3	<0.001	*24.3	0.005
2 vs. 32	***69.7	<0.001	***68.7	<0.001	**50.4	0.001	***71.9	<0.001	*25.4	0.012
4 vs. 8	20	0.076	*32.4	0.019	17.4	0.057	20.6	0.209	6.9	0.564
4 vs. 16	*29.1	0.049	*34.7	0.023	***31.3	<0.001	**51.3	0.002	*11.0	0.034
4 vs. 32	*32.7	0.037	*44.6	0.011	***36.9	<0.001	***60.9	<0.001	12.1	0.296
8 vs. 16	9.1	0.199	2.29	1.000	13.9	0.147	30.7	0.058	4.1	1.000
8 vs. 32	12.7	0.288	12.1	0.696	19.4	0.059	*40.3	0.011	5.3	1.000
16 vs. 32	3.6	1.000	9.86	0.131	5.6	0.107	9.6	0.455	1.1	1.000

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

TABLE III. Results of multiple comparisons between vowel recognition scores with five vocoders for each of the three band numbers (8, 16, and 32).

number of bands→	8		16		32	
vocoder pair↓	mean diff	*p*	mean diff	*p*	mean diff	*p*
sine-wave vs. noise-separate	−2.1	1.000	4.7	1.000	−1.6	1.000
sine-wave vs. noise-spread	23.4	0.054	**18.7	0.009	16.7	0.120
sine-wave vs. GET-separate	25.6	0.065	4.0	1.000	−2.0	1.000
sine-wave vs. GET-spread	**42.4	0.002	***47.4	<0.001	***49.9	<0.001
noise-separate vs. noise-spread	*25.6	0.030	*14.0	0.025	*18.3	0.048
noise-separate vs. GET-separate	27.7	0.080	−0.7	1.000	−0.4	1.000
noise-separate vs. GET-spread	**44.6	0.004	***42.7	<0.001	***51.4	<0.001
noise-spread vs. GET-separate	2.1	1.000	−14.7	0.130	−18.7	0.102
noise-spread vs. GET-spread	19.0	0.459	**28.7	0.002	**33.1	0.004
GET-separate vs. GET-spread	16.9	0.156	**43.4	0.001	***51.9	<0.001

TABLE IV. Results of multiple comparisons between consonant recognition scores with five number of bands for each of the five vocoders

vocoder→	sine-wave		noise-separate		noise-spread		GET-separate		GET-spread	
number of bands pair↓	mean diff	*p*	mean diff	*p*	mean diff	*p*	mean diff	*p*	mean diff	*p*
2 vs. 4	*37.0	0.014	24.1	0.140	13.6	1.000	11.0	0.827	*13.3	0.019
2 vs. 8	**57.0	0.001	**56.6	0.002	31.0	0.085	31.6	0.082	*20.1	0.048
2 vs. 16	***66.1	<0.001	***58.9	<0.001	**44.9	0.001	***62.3	<0.001	**24.3	0.005
2 vs. 32	***69.7	<0.001	***68.7	<0.001	**50.4	0.001	***71.9	<0.001	*25.4	0.012
4 vs. 8	20.0	0.076	*32.4	0.019	17.4	0.057	20.6	0.209	6.9	0.564
4 vs. 16	*29.1	0.049	*34.7	0.023	***31.3	<0.001	**51.3	0.002	*11.0	0.034
4 vs. 32	*32.7	0.037	*44.6	0.011	***36.9	<0.001	***60.9	<0.001	12.1	0.296
8 vs. 16	9.1	0.199	2.3	1.000	13.9	0.147	30.7	0.058	4.1	1.000
8 vs. 32	12.7	0.288	12.1	0.696	19.4	0.059	40.3	0.011	5.3	1.000
16 vs. 32	3.6	1.000	9.9	0.131	5.6	0.107	9.6	0.455	1.1	1.000

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

TABLE V. Results of multiple comparisons between consonant recognition scores with five vocoders for each of the five band numbers

number of bands→	2		4		8		16		32	
vocoder pair↓	mean diff	*p*	mean diff	*p*	mean diff	*p*	mean diff	*p*	mean diff	*p*
sine vs. noi-sep	−4.9	1.000	−3.7	1.000	−3.0	1.000	−0.1	1.000	−0.1	1.000
sine vs. noi-spr	−1.6	1.000	13.7	0.267	21.7	0.292	*24.7	0.016	*24.3	0.022
sine vs. GETsep	30.0	0.065	**35.1	0.002	18.1	0.096	−1.0	1.000	2.4	1.000
sine vs. GETspr	27.6	0.108	**33.1	0.010	**49.0	0.006	**54.4	0.002	**59.6	0.002
noi-sep vs. noi-spr	3.3	1.000	17.4	0.157	24.7	0.111	*24.9	0.021	*24.4	0.038
noi-sep vs. GETsep	**34.9	0.007	**38.9	0.007	21.1	0.115	−0.9	1.000	2.6	1.000
noi-sep vs. GETspr	*32.4	0.018	**36.9	0.002	**52.0	0.007	**54.6	0.003	**59.7	0.004
noi-spr vs. GETsep	*31.6	0.043	**21.4	0.003	−3.6	1.000	**−25.7	0.010	*−21.9	0.040
noi-spr vs. GETspr	29.1	0.078	**19.4	0.003	27.3	0.065	**29.7	0.006	***35.3	<0.001
GETsep vs. GETspr	−2.4	1.000	−2.0	1.000	*30.9	0.034	**55.4	0.002	**57.1	0.003

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

FIGURE CAPTIONS

FIG. 1. (Color online) A unit-amplitude single pulse with Gaussian-shaped envelope (black line) in both the time (a) and frequency (b) 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 Eq. (1), producing an effective duration of 0.85 ms and an effective bandwidth of 1.2 kHz.

FIG 2. (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 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.

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 (the first type; used in Exp. 1) or using a single GET as an impulse response (the second type; used in Exp. 2). The front-end pre-emphasis, bandpass filter, and envelope extraction can be implemented either in the temporal or spectral domain.

FIG. 4. (Color online) A 100-Hz pulse train, repeating a 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. 1.

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

FIG. 6. (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 are averaged from seven normal hearing subjects listening to vocoded speech. 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 their 19 participants.

FIG. 7. 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

medRxiv preprint doi: https://doi.org/10.1101/2022.02.21.22270929; this version posted May 9, 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.

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.

FIG. 8. (Color online) Speech stimulus demonstrations for the ACE-GET simulation experiment. Left: Spectrogram; middle: band-specific signal; right: zoom in of the boxed signals. **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.

FIG. 9. 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.