International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 4 Issue 1 ǁ Jan. 2016 ǁ PP.89-102

Simulations of cochlear implant signal processing

Subhadip Goswami, Subrat Kumar Panda, Salila Malla

Department of Electronics and Communication Engineering, NM Institute of Engineering and Technology,Bhubaneswar , Odisha
Department of Electronics and Communication Engineering,Aryan Institute of Engineering and Technology Bhubnaeswar , Odisha
Department of Electronics and Communication Engineering,Capital Engineering College,Bhubaneswar,Odisha

ABSTRACT: *Conventional signal processing implemented on clinical cochlear implant (CI) sound processors is based on en-velope signals extracted from overlapping frequency regions. Conventional strategies do not encode temporal envelope or temporal fine-structure cues with high fidelity. In contrast, several research strategies have been developed recently to enhance the encoding of temporal envelope and fine-structure cues. The present study ex-amines the salience of temporal envelope cues when encoded into vocoder representations of CI signal processing.*

Normal-hearing listeners were evaluated on measures of speech reception, speech quality ratings, and spatial hearing when listening to vocoder representations of CI signal processing. Conventional vocoder techniques using envelope signals with noise- or tone-excited reconstruction were evaluated in comparison to a novel approach based on impulse-response reconstruction. A variation of this impulse-response approach was based on a research strategy, the Fundamentally Asynchronous Stimulus Timing (FAST) algorithm, designed to improve temporal precision of envelope cues.

The results indicate that the introduced impulse-response approach, combined with the FAST algorithm, produces similar results on speech reception measures as the conventional vocoder approaches, while providing significantly better sound quality and spatial hearing outcomes. This novel approach for stimulating how temporal envelope cues are encoded into CI stimulation has potential for examining diverse aspects of hearing, particularly in aspects of musical pitch perception and spatial hearing.

Keywords:
Cochlear implants
Signal processing
Speech reception
Binaural hearing

I. INTRODUCTION

Cochlear implants (CIs) are medical devices that restore a degree of hearing to people with severe to profound hearing loss. Conventional signal processing used on clinical devices is based on envelope signal extracted from overlapping frequency regions, which span the relevant spectrum for speech. These envelope-based strategies have been success-ful in restoring a remarkable degree of hearing for CI users. In recent years, several new strategies have been proposed to encode temporal envelope and temporal fine-structure cues with enhanced fidelity (van Hoesel and Tyler 2003; Arnoldner et al., 2007). The motivation for these strategies is that precise encoding of temporal cues may improve pitch perception and spatial hearing for CI users. The present study exam-ines a novel vocoder approach, combined with an algorithm to enhance the encoding of temporal envelope cues (FAST: Smith et al., 2014), to evaluate the effect of precise encoding of temporal envelope cues for speech reception in background noise, speech quality rating, and spatial hearing.

Vocoders are signal processing methods that provide a degree of in-dependent control over the envelope and fine-structure characteristics of a signal. The term vocoder is a combination of the words “voice” and “coder”, as the original class of vocoders was built on the princi-pal of coding the voice and then reconstructing the acoustic signal in accordance with this code (Dudley 1939; Schroeder 1966). More re-cently, a class of vocoders referred to as channel vocoders was devel-oped to model the perceptual effects of CI signal processing (Shannon et al., 1995; Dorman et al., 1997). Channel vocoders provide insight into which acoustic features are relevant for speech reception in differ-ent acoustic environments (Fu et al., 1998; Nelson et al., 2003; Chen and Loizou 2011).

The various channel vocoders are similar in that they implement analysis schemes to separate sound into frequency channels, extract temporal envelopes, then reconstruct band-limited signals before com-bining across channels for the acoustic output. The two methods that have been most studied as models of CI signal processing are the noise (Shannon et al., 1995) and tone (Dorman et al., 1997) vocoders. Other methods for reconstructing channel envelopes into band-limited signals have been considered including methods that use Gaussian-envelope tones (Lu et al., 2010), harmonic complexes (Deeks and Carlyon 2004; Hervais-Adelman et al., 2011), and neural modeling (Boghdady et al., 2016). These differing methods produce similar results on speech comprehension measures, while producing qualitatively different representations. Presently, the large variability in CI speech reception outcomes makes it difficult to conclude whether any of these methods are in practice better models of the perceptual consequences of CI signal processing.

The vocoder methods introduced in this article are based on recon-structing stimulation sequences, which directly correspond to CI stim-ulation sequences. In typical CI signal processing, the acoustic signal is processed through a bank of filters, the envelope is extracted, and then a pulsatile stimulation sequence is generated using pulse-generating logic such as Continuous Interleaved Sampling (CIS: Wilson et al., 1991). Typical channel vocoders reconstruct an acoustic signal from envelopes extracted from each frequency band. In contrast, the impulse-response vocoders introduced here use stimulation sequences directly produced by CI signal processing. Specifically, sequences are filtered through a reconstruction filter bank to produce band-limited signals that are summed across channels.

The advantage of using pulsatile stimulation rather than envelopes is that it allows temporal differences in CI stimulation strategies to be examined. For example, several stimulation strategies have been intro-duced that use temporal synchronization of stimulation to the underly-ing temporal envelope or fine structure of the acoustic signal (van Hoe-sel and Tyler 2003; Arnoldner et al., 2007; van Hoesel 2007; Vandali and van Hoesel 2012). The motivation for developing algorithms that control stimulation timing with greater precision is to improve aspects of hearing that may depend on temporal fine structure such as spatial hearing and pitch perception. Two impulse-response vocoders are intro-duced in this article: one is based on the CIS stimulation strategy that is in clinical use, the other is based on the FAST algorithm (Smith et al., 2014), which triggers pulsatile stimulation based on temporal maxima of channel envelopes.

The present study examined hearing for normal-hearing listeners attending to vocoder representations of CI signal processing. Novel impulse-response vocoders were examined in comparison to the well-established noise and tone vocoders on measures of speech reception in noise, speech quality in quiet, and spatial lateralization based on inter-aural timing cues. For speech reception in background noise, speech re-ception was measured in stationary speech-spectrum noise and in time-reversed speech. These noise conditions were selected to examine speech reception differences between unmodulated and modulated background noise. It has been established that normal-hearing listeners have better performance in modulated than unmodulated background noise, but this masking release in modulated noise is reduced in hearing-impaired lis-teners (Bacon et al., 1998; Peters et al., 1998; Moore et al., 1999). CI users typically receive little to no modulation masking release (Nelson et al., 2003; Jin and Nelson 2006), and in some cases have been shown to have poorer speech reception in modulated background noise (Kwon and Turner 2001; Kwon et al., 2012). The vocoder conditions examined in the present study include reconstruction bandwidth as a parameter to allow consideration of how spectral resolution impacts masking release in modulated background noise. Consequently, in addition to providing validation of the introduced impulse-response vocoders, the study also provides insight into the issue of masking release in modulated back-ground noise.

Speech quality in quiet was evaluated with each vocoder method using two different reconstruction bandwidths to control spectral res-olution. The purpose of evaluating speech quality was to demonstrate that these different methods could provide comparable speech reception while providing significantly different quality ratings. As vocoder mod-els become increasingly successful in predicting CI perceptual outcomes, there is a risk of assuming these models represent the quality of hear-ing through a CI. Such conclusions regarding sound quality should be made cautiously since, as others have pointed out, neither the noise or tone vocoders are expected to produce an auditory nerve response sim-ilar to that produced by actual CI stimulation (Boghdady et al., 2016). Consequently, a simple proof that various vocoder methods can produce comparable speech reception, while producing different quality ratings, is an important demonstration that encourages caution when drawing conclusions regarding the quality of hearing with CIs.

Spatial lateralization was measured using an interaural timing differ-ence (ITD) discrimination task for the different vocoder methods. The rationale for measuring ITD discrimination thresholds with the various vocoder methods was to provide an initial examination of a perceptual measure that strongly depends on precision of stimulation timing. It was expected that the noise and tone vocoder methods, which do not explicitly encode stimulation timing, would provide relatively weak access to interaural cues; whereas, the impulse-response vocoders introduced here, particularly the version based on the FAST algorithm, would pro-vide access to temporal cues needed for ITD discrimination. The combi-nation of speech reception, quality ratings, and spatial hearing measures were selected to validate the novel vocoder methods on relevant speech reception measures, while extending the vocoder technique more gen-erally for encoding temporal envelope cues for binaural hearing.

II. MATERIAL AND METHODS

2.1. Subjects

Subjects consisted of 8 normal-hearing listeners. The University of Southern California's Institute Review Board approved the study pro-tocol. All subjects provided informed consent and were paid for their participation. All subjects were native English speakers who had pure tone audiometric thresholds of 20 dB HL or better at octave frequencies between 125 and 8000 Hz.

2.2. Speech and noise materials

The Coordinate Response Measure (CRM) sentence database (Bolia et al., 2000) was used to measure speech reception in noise and speech quality in quiet. The CRM materials consist of sentences of the form "Ready call sign go to color number now," with all 256 combina-tions of 8 call signs ("Arrow", "Baron", "Charlie", "Eagle", "Hopper", "Laker", "Ringo" "Tiger"), 4 colors ("blue", "green", "red", "white"), and 8 numbers (1 through 8). These sentence materials were recorded using 4 female and 4 male talkers. For speech reception and quality measures used in the present study, only one of the speakers (a male) was used as target speech. The rationale for using the same male talker on each trial is that in typical conversations one is aware of whom one is speak-ing to, while there are circumstances (e.g., answering the phone) that this assumption does not hold, it is typically true; therefore, we chose not to include talker variability as one of the perceptual dimensions to examine.

Background noise included stationary speech-spectrum noise and time-reversed speech. Speech-spectrum noise was generated by filter-ing Gaussian noise through a spectral-shaping filter estimated from the average power spectral density of 1 min of CRM sentences using the target male talker. Time-reversed speech was generated by randomly concatenating 4 of the CRM sentences using a female talker, selecting a segment of equal length to the target sentence from within that con-catenation, then time reversing the segment. A spectral-shaping filter was applied to the time-reversed speech to compensate for any aver-age power spectral density differences between the target male talker and the female masker. Thus, the speech-spectrum noise and the time-reversed speech had the same average power spectral density. All ma-terials were down-sampled to 16,000 Hz and all subsequent signal pro-cessing was performed at this sample rate.

2.3. Signal processing

Speech reception and speech quality were measured for four dif-ferent vocoder algorithms, which will be referred to as noise, tone, CIS, and FAST vocoders. The noise and tone vocoders were based on algorithms that have been established as useful models of speech

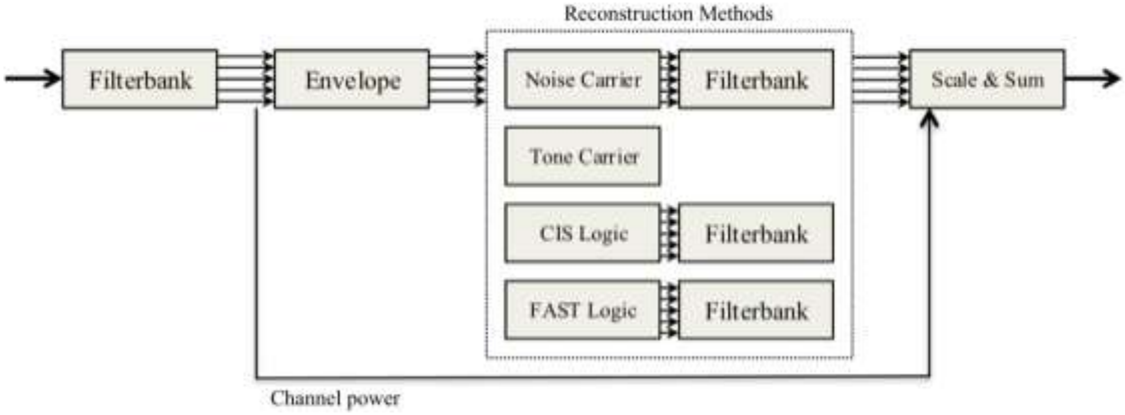

Fig. 1. Diagram of the noise, tone, CIS, and FAST vocoders. These vocoders use the same filter bank and envelope extraction procedures but differ in how the extracted envelopes are used to produce acoustic representations of CI signal processing.

reception for CI users. The CIS and FAST vocoders are introduced here as impulse-response methods that directly reconstruct pulsatile stimula-tion sequences.

The general structure of the vocoder algorithms is depicted in Fig. 1. Each algorithm used identical analysis filter bank and envelope extrac-tion routines. The algorithms differ in the methods used to reconstruct the channel envelopes into acoustic signals. For the noise vocoder, inde-pendent Gaussian noise was generated for each channel, and then mul-tiplied by the corresponding channel envelope and processed through a reconstruction filter bank. For the tone vocoder, sinusoidal signals with frequencies corresponding to the center frequencies of the analysis filter bank were generated, and then multiplied by the corresponding channel envelope. For the CIS and FAST algorithms, logic (described in a subsequent paragraph) was implemented to convert the channel envelopes into a pulsatile stimulation sequence. These pulsatile stimu-lation sequences were then filtered through a reconstruction filter bank. For each sentence that was processed, the average power across time was calculated for each channel output from the analysis filter bank, which was then used to scale the outputs of the reconstruction stage just prior to adding the signals together for the vocoder output.

The vocoders used in this study used 16-channel filter banks with center frequencies logarithmically spaced between 250 and 4000 Hz. Each filter was implemented as a 256th-order finite impulse-response filter constructed using the Hann window method. For analysis filter banks, the bandwidth of the filters was defined such that the 6-dB crossover point occurred midway between center frequencies with log-arithmic spacing. Given that the filter bank used 16 filters spanning 4 octaves, the corresponding bandwidth of these filters is 1/4th octave; specifically, the 6-dB crossover points occurred at $2^{\pm 1/8}$ times the center frequency of the filter. The channel envelopes were extracted from the filter bank outputs using the Hilbert transform method. These envelopes were further processed using an 8th-order infinite impulse-response low-pass filter of Butterworth design having a cutoff frequency of 300 Hz.

The CIS and FAST vocoders included routines to convert the chan-nel envelopes to pulsatile stimulation sequences. For the CIS vocoder, the channel envelopes were continually sampled at an overall rate of 4000 Hz such that the pulsatile rate was 250 Hz per channel. Specifi-cally, every 0.25 ms (equivalently, every 4th sample point) the channel envelopes were sampled starting with the highest frequency channel and in turn sampling from progressively lower frequency channels. For the FAST vocoder, all temporal local maxima of the channel envelopes were selected as pulsatile values, with all other values set to zero. The resulting pulsatile stimulation patterns were then filtered through a re-construction filter bank, scaled and added together to produce the cor-responding vocoder output.

To clarify the reconstruction mechanism of the FAST impulse-response vocoder, Fig. 2 illustrates the decision logic and acoustic reconstruction for a single processing channel for a brief portion of a vowel. The FAST processing logic triggers impulses based on the lo-cal temporal maxima of each band-limited envelope. These impulses are then used to reconstruct a band-limited acoustic stimulus by filter-ing the impulse through a reconstruction filter. Hence the reconstructed "pulses" are impulse responses of the reconstruction filter.

For reconstruction, two different bandwidths were implemented, re-ferred to as narrow and broad reconstruction. The narrow reconstruc-tion filter bank was identical to the analysis filter bank; the broadly tuned filter bank was specified to contain filters with one-octave band-

√ √

width, specifically the 6 dB attenuation points occurred at 1⁄2 and 2 times the filter center frequencies; aside from this tuning difference, the narrow and broad filter banks were identical. For implementing a com-parable broad reconstruction for the tone vocoder, the filter envelopes were spectrally smeared directly prior to modulating the sinusoidal car-riers. This spectral smearing was implemented by convolving the filter envelopes across the 16 channels with the smearing filter [$\frac{1}{2}$ 1 1 1 $\frac{1}{2}$]. Since the filters are spaced 1/4th octave apart, this smearing achieves a comparable level of spectral smearing as the broadened reconstruction filters used in the other methods.

Measured speech reception thresholds also depend on the resolution of the vocoder filtering. In the present study, vocoder filtering and re-construction was implemented using 16 filters logarithmically spaced between 250 and 4000 Hz resulting in $\frac{1}{4}$ -octave spacing between fil-ters. Such spacing is comparable but slightly downshifted to the default frequency allocation using by Advanced Bionics clinically programming software, which allocates filters between 333 and 6665 Hz with $\frac{1}{4}$ -octave filtering.

The CIS and FAST impulse-response vocoders have special consid-erations as to how the acoustic impulse responses interact to affect the sound quality of the synthesized signal. Specifically, close spectrotem-poral proximity of impulse responses can interact to produce distortions. It was determined that an overall CIS pulse rate of 4000 Hz (250 Hz per channel) produced a qualitatively acceptable reconstruction and intelli-gible speech. However, such a low per channel pulse rate might be in-sufficient to characterize fundamental frequency of voicing. While pitch perception associated with pulse-rate presentation of fundamental fre-quency was not explicitly considered in the present study, it is an impor-tant issue of future study since many older implants still employ pulse rates in the 250 to 350 Hz range.

The FAST impulse-response vocoder minimizes the interaction between spectrotemporally proximal impulse responses by using a relatively sparse pulse rate that is synchronized to the fundamen-tal frequency. In this way, the per channel pulse rate provides a clear representation of the fundamental frequency but the spectral fine structure thus leading to the smeared spectral response neces-sary for a cochlear implant simulation. The acoustic results of the

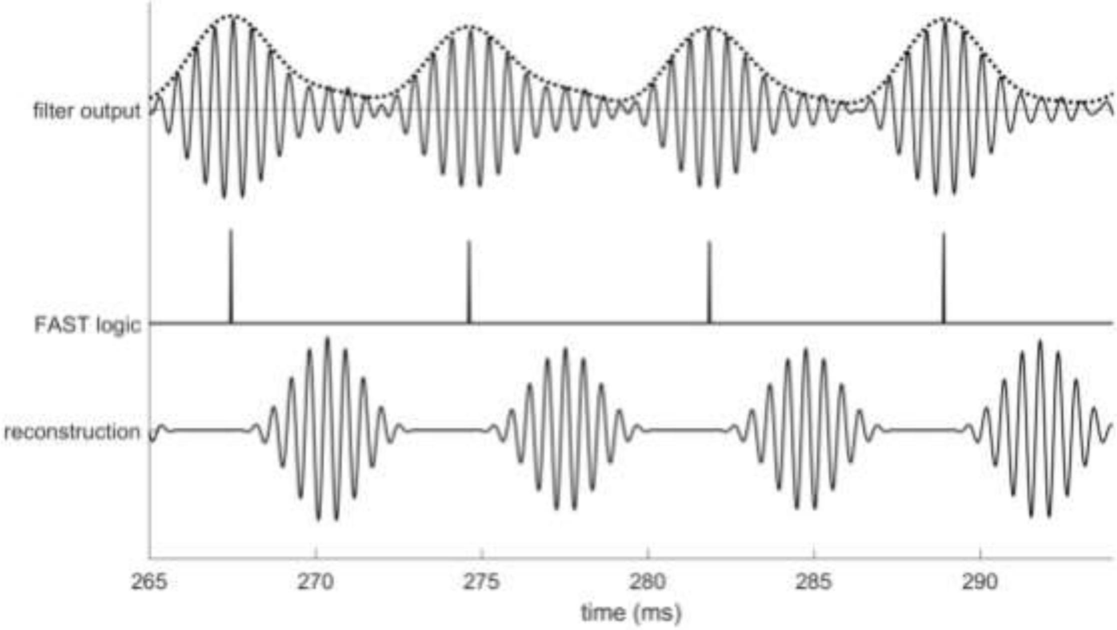

Fig. 2. Illustration of the FAST analysis logic and reconstruction procedure for four glottal periods of a spoken vowel as observed for an individual processing channel. The filter output shows the output from the filtering and envelope extraction stages. The FAST logic generates an impulse at local temporal maxima of the envelope signals. Reconstruction is performed by filtering the impulses through a reconstruction bandpass filter.

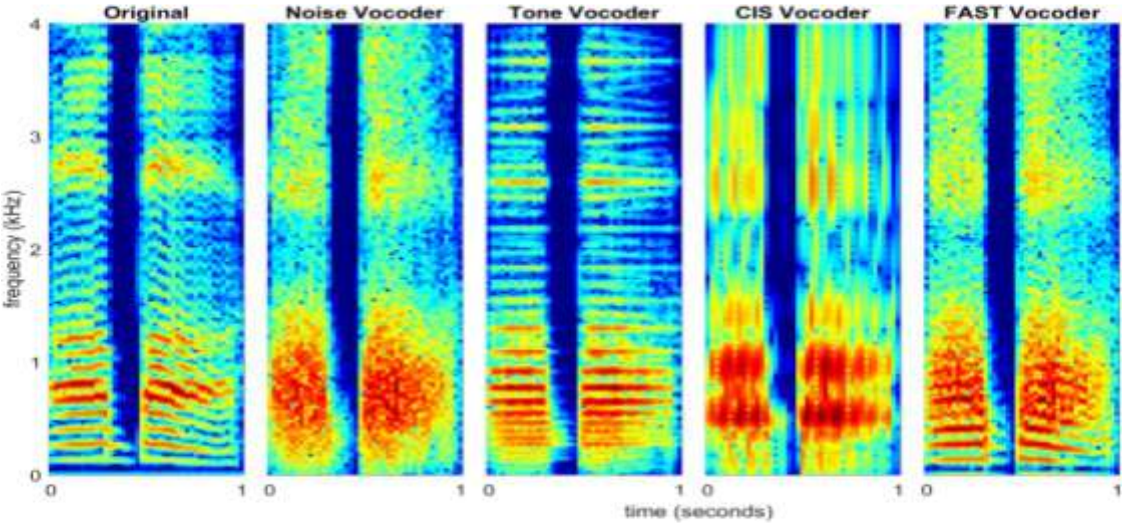

Fig. 3. Spectrogram representations of the vowel-consonant-vowel utterance /aba/ spoken by a man. Each of the vocoder processing schemes provide comparable spectral representations of the formant frequencies with degradation of the place-of-excitation cues. Each processing scheme introduces characteristic distortions with the noise vocoder introducing broadband stochastic distortion across frequency regions and with the tone

vocoder distorting the spectrum as it is sampled by a set number of tonal frequencies. The impulse vocoders introduce distortion somewhat between the noise and tone vocoder methods, but with the FAST vocoder providing a cleaner representation of fundamental frequency.

analysis/synthesis approaches are illustrated in Fig. 3, which provides spectrograms of the reconstructed signals for the phoneme /aba/ spo-ken by a man with an average fundamental frequency of 120 Hz for this utterance. The various vocoder methods represent the formant frequencies in different ways but with formants spectrally smeared.

For the FAST vocoder, the fundamental frequency of voicing is repre-sented in a manner more comparable to the original signal. This rep-resentation is hypothesized to lead to a better quality of reconstruction while producing spectral smearing essential to cochlear implant simulation.

2.4. Speech reception in noise

Speech reception thresholds (SRTs) were measured for 16 conditions consisting of every combination of 2 background noise types (speech-spectrum noise and time-reversed speech), 4 vocoder algorithms (noise, tone, CIS, and FAST), and 2 reconstruction methods (narrow and broad). These 16 conditions were tested in random order with 3 repetitions of each condition. For each trial in the procedure, a target sentence was randomly selected from the CRM database always using the same male talker. The target sentence was combined with background noise and processed through the vocoder algorithm for the condition. The pro-cessed speech was presented to the left ear at 65 dB SPL. Sentences were scored correct when the subject identified both the color and number of the sentence. The initial signal-to-noise ratio (SNR) of the procedure was set to 12 dB SNR, which was decreased/increased by 2 dB after each trial that was scored correct/incorrect. The procedure continued for 8 reversals and the average of SNR values from the last 4 reversals was taken as the SRT for the run.

2.5. Speech quality in quiet

Speech quality was measured using a quality rating procedure in which the subject listened to speech in quiet for 8 conditions consisting of every combination of the 4 vocoder algorithms (noise, tone, CIS, and FAST) and 2 reconstruction bandwidths (narrow and broad). For this quality rating procedure, the same sentence from the CRM database was used for every trial (i.e., “Ready Charlie go to blue one now”). The same sentence was used for every trial since the purpose of this procedure was to measure subjective rating of speech quality, so acoustic and linguistic differences between sentences were purposely minimized. The sentence was processed with the corresponding vocoder and was presented to the left ear at 65 dB SPL. The processed sentence was presented by itself (the unprocessed sentence was not presented as a reference).

The subject responded with a computer interface to rate the quality of the perceived speech from 1 to 10, with 10 indicated as the highest quality. A 32-trial familiarization procedure was implemented in which the 8 conditions were measured with 4 trials per condition in random order. This familiarization procedure allowed the subject to orient how he/she would distribute the varying processing conditions across the quality rating scale. This familiarization procedure was immediately fol-lowed by a 64-trial procedure consisting of the 8 processing conditions measured with 8 trials per condition in random order.

2.6. Spatial lateralization

Spatial lateralization was measured using a lateralization procedure based on ITD discrimination thresholds, which were measured for a modulated sinusoid for unprocessed (i.e., not vocoded) sounds and for the 8 vocoder conditions consisting of the noise, tone, CIS, and FAST vocoders using narrow and broad reconstruction. Discrimination thresh-olds were measured using an adaptive two-alternative, two-interval, forced-choice procedure. The stimulus was a 500 Hz sinusoid modu-lated by a 100 Hz raised-cosine modulator. Listeners were presented with two sequential sounds with either the first interval having a left-leading interaural delay followed by the second interval having a right-leading interaural delay or vice versa. Listeners were instructed to in-dicate whether the sound moved from left to right or from right to left across the two intervals. Correct answer feedback was given in the form of flashing the user interface response button as green for correct and red for incorrect responses.

ITDs were set to 1000 s for the first trial and were decreased by a factor of $2^{1/3}$ following correct responses and increased by a factor of 2 following incorrect responses. The maximum ITD value allowed by the procedure was 1000 s. The adaptive procedure was implemented for 8 reversals and the logarithmic average of the last 4 reversals was calculated as the ITD threshold for the condition. This adaptive rule converges to 75% discrimination accuracy (Kaernbach 1991).

III. RESULTS

3.1. Speech reception in noise

Speech reception thresholds (SRTs) were measured for 16 condi-tions consisting of two background noise types (speech-spectrum noise and time-reversed speech), 4 vocoder algorithms (noise, tone, CIS, and FAST), and 2 reconstruction methods (narrow and broad). Fig. 4 plots SRTs averaged across repetitions and subjects, with average SRTs mea-sured in time-reversed speech plotted against the corresponding SRTs measured in speech-spectrum noise. For example, SRTs measured with the noise vocoder using narrow reconstruction is plotted as a solid square with an average SRT in speech-spectrum noise of −5.4 dB SNR and an average SRT in time-reversed speech of −16.4 dB SNR for an av-erage SRT difference of 10 dB with better performance achieved in time-reversed speech. Each of the vocoder algorithms using narrow recon-struction produced comparable results with average SRTs measured in speech-spectrum noise ranging from −6.9 dB SNR for the FAST vocoder to −4.8 dB SNR for the CIS vocoder, and with average SRTs measured in time-reversed speech ranging from −16.8 dB SNR for the FAST vocoder to −12.2 dB SNR for the CIS vocoder. Similarly, each of the vocoder al-gorithms using broad reconstruction produced comparable results with average SRTs measured in speech-spectrum noise ranging from −2.8 dB SNR for the FAST vocoder to −0.6 dB SNR for the CIS vocoder, and with average SRTs measured in time-reversed speech ranging from −3.7 dB SNR for the tone vocoder to −0.8 dB SNR for the CIS vocoder. These results indicate that speech reception in noise is similar with each of these four vocoder algorithms with a common trend of better speech reception, particularly in modulated background noise, occurring when narrow reconstruction methods are used.

Measured SRTs were analyzed using a 3-way repeated-measures analysis of variance (ANOVA) with background noise type, vocoder al-gorithm, and reconstruction bandwidth as factors. As expected, noise type ($F_{1,256}$ = 37.2, $p < 0.001$) and reconstruction bandwidth ($F_{1,256}$

= 767.2, $p < 0.001$) were both significant indicating that speech recep-tion is differentially affected by modulated and unmodulated back-ground noise as well as by the spectral resolution of the vocoder algorithm. More relevant to the present study, vocoder algorithm was significant ($F_{3,256}$ = 32.4, $p < 0.001$) indicating SRT differences across vocoder algorithms. The second order interactions between background noise type and reconstruction bandwidth was significant ($F_{1,256}$ = 222, $p < 0.001$) reflecting the trend illustrated in Fig. 4 that SRTs were rel-atively lower when measured in time-reversed speech compared to speech-spectrum noise when using narrow rather than broad recon-struction bandwidths. The interaction between background noise type and vocoder algorithm was also significant ($F_{3,256}$ = 8.4, $p < 0.001$) indi-cating relative differences for each vocoder algorithm when measuring SRTs in different background noise types.

Post-hoc analyses were implemented to examine the differences between SRTs associated with each vocoder algorithm. A multiple-comparisons analysis based on Tukey's honest significant difference cri-terion was implemented using the ANOVA statistics described in the pre-vious paragraph. This multiple-comparisons analysis indicated that only a few comparisons were significantly different ($p < 0.05$) when compar-ing SRTs across vocoders for a given background noise type and re-construction bandwidth. Specifically, SRTs measured in time-reversed speech and using narrow reconstruction bandwidths were significantly higher for the CIS vocoder than for the SRTs measured with the noise and FAST vocoders for the same conditions. Similarly, SRTs measured in time-reversed speech using broad reconstruction methods were sig-nificantly higher for the CIS vocoder than the SRTs measured with the noise and tone vocoders for the same conditions. No other comparisons of SRTs for the same background noise type and reconstruction method

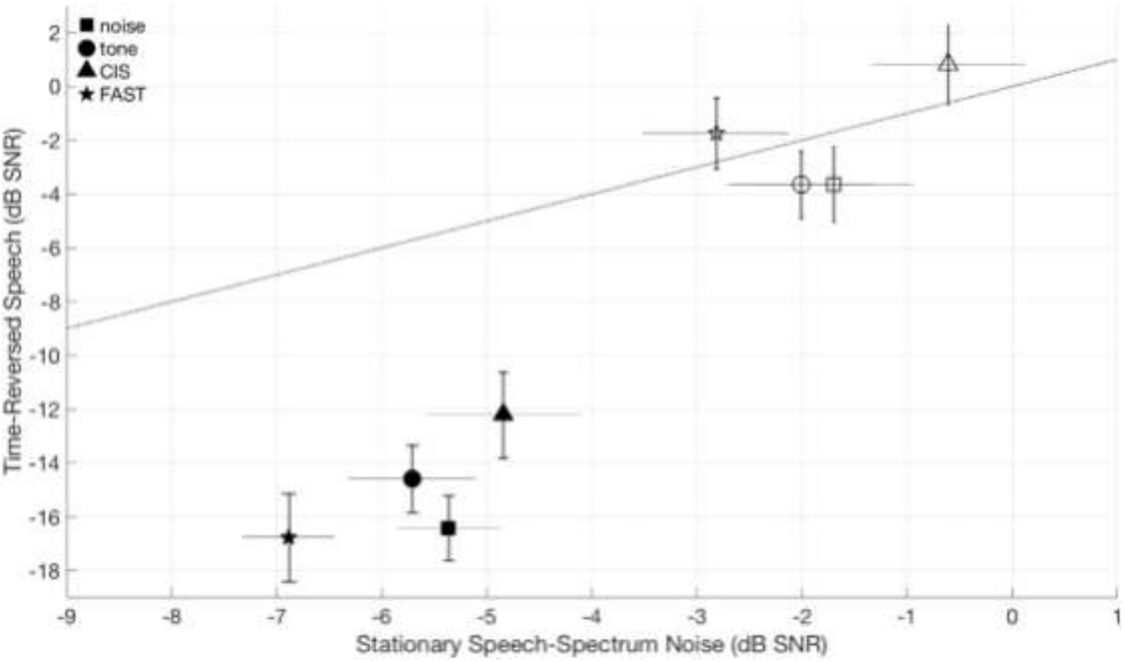

Fig. 4. Speech reception thresholds measured in time-reversed speech plotted against speech reception thresholds measured in stationary speech-spectrum noise for the four examined vocoders. Each vocoder is implemented using narrow (filled symbols) and broad (open symbols) reconstruction bandwidths. Error bars indicate standard error of the mean.

were significantly different ($p > 0.05$). Consequently, the results indi-cate that measured SRTs were generally similar with the caveat that the CIS vocoder tended to produce significantly higher SRTs than the other vocoders when measured in time-reversed speech.

3.2. Speech quality in quiet

Speech quality was measured using the quality rating procedure for 8 conditions consisting of the 4 vocoder algorithms (noise, tone, CIS, and FAST), and 2 reconstruction bandwidths (narrow and broad). Speech quality was also measured for an unprocessed condition to serve as a reference. Fig. 5 plots quality ratings averaged across repetitions and subjects. Measured quality ratings were lowest for the noise vocoder. Measured quality ratings were highest for the FAST vocoder when using narrow reconstruction and highest for the CIS vocoder when using broad reconstruction. For comparison, quality ratings for the unprocessed con-dition was 9.56 with a standard deviation of 0.32 across subjects.

Quality ratings were analyzed using a 2-way repeated-measures ANOVA with vocoder algorithm and reconstruction bandwidth as fac-tors. The effect of vocoder algorithm ($F_{3,21} = 26.8$, $p < 0.001$), recon-struction bandwidth ($F_{1,21} = 71.5$, $p < 0.001$), and the interaction be-tween vocoder algorithm and reconstruction bandwidth ($F_{3,21} = 8.3$, $p < 0.001$) were all significant effects on speech quality.

To further analyze these effects, a multiple-comparisons analysis based on Tukey's honest significant difference criterion was imple-mented using the ANOVA statistics described in the previous paragraph. For vocoders implemented with narrow reconstruction (i.e., Fig. 5, Panel A), the p value associated with the quality comparison between the highest rated vocoder (FAST) and the next highest rated vocoder (CIS) was 0.14. The p value associated with the comparison between the CIS and tone vocoders was 0.98. All other comparisons between quality ratings for vocoders with narrow reconstruction were significant ($p < 0.05$). For vocoders implemented with broad reconstruction (i.e., Fig. 5, Panel B), the p value associated with the quality comparison between the highest rated vocoder (CIS) and the next highest rated vocoder (tone) was 0.12. The p value associated with the comparison between the tone and FAST vocoders was 0.99. All other comparisons between quality ratings for vocoders with broad reconstruction were significant ($p < 0.05$). Consequently, these results generally indicate that there are significant differences between the perceived quality of speech processed through these different vocoder algorithms.

3.3. Spatial lateralization

Spatial lateralization was measured as ITD discrimination thresholds for 8 normal-hearing listeners using 500 Hz sinusoids modulated by 100 Hz raised-cosine envelope. Thresholds were measured for 9 vocoder conditions consisting of the 4 vocoder methods (noise, tone, CIS, and FAST) for each reconstruction method (narrow or broad), and for an un-processed condition in which no vocoder was implemented. The results are summarized in Fig. 6, which plots the ITD discrimination thresholds averaged across subjects.

The dotted line of Fig. 6 represents chance performance estimated us-ing a Monte Carlo procedure. Specifically, the adaptive rule described in the Methods section was implemented for 10,000 iterations using ran-dom assignments of correct/incorrect responses. The probability distri-bution of these 10,000 iterations were analyzed to determine the 95% probability threshold, which is the lower bound on thresholds that have a 95% probability of being greater than chance.

The ITD discrimination results plotted in Fig. 6 indicate that the av-erage ITD discrimination thresholds of subjects listening through noise, tone, or CIS (using narrow reconstruction bandwidths) vocoders did not differ from chance performance by more than one standard error of the mean. This result is not surprising as those vocoder methods discard tem-poral fine structure differences across ears, which is replaced by diotic carriers, whether those carriers are tones, noise, or impulse responses. Within this context, it is important to realize that the impulses gener-ated for CIS are diotic in the sense of being binaurally identical. What is surprising is that some subjects could perform better than chance using the CIS vocoder with broad reconstruction. This is somewhat surpris-ing as the CIS reconstruction process is diotic with regards to tempo-ral positioning of impulses. However, for any of these vocoder meth-ods, interaural level and envelope-timing differences are preserved in the channel envelope signals. That some subjects could perform the ITD

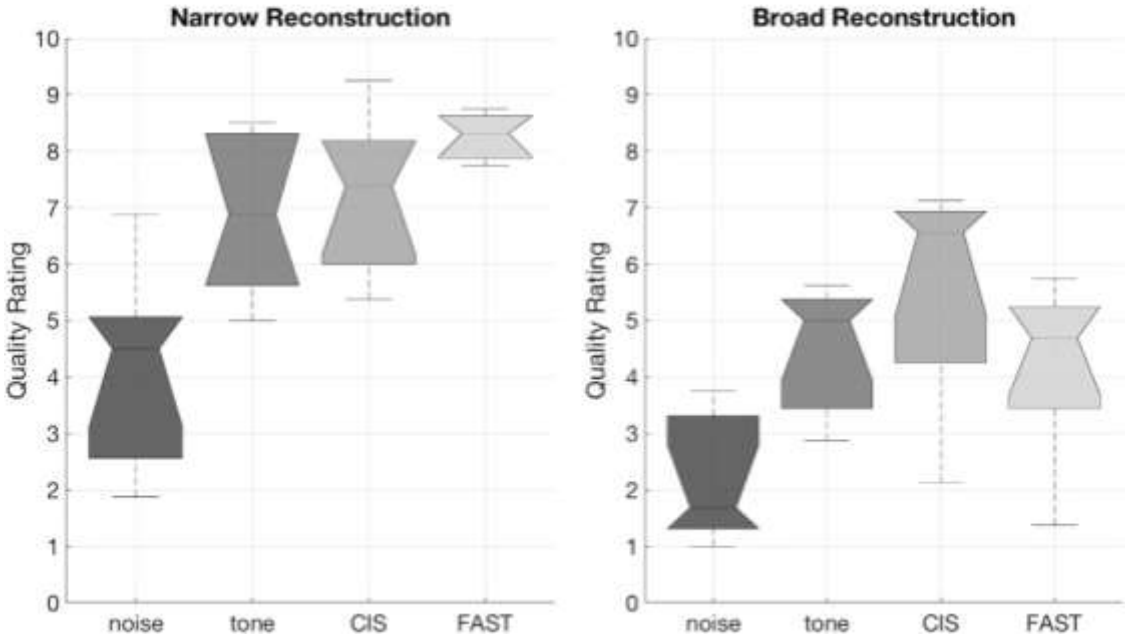

Fig. 5. Quality ratings averaged across repetitions and subjects for each vocoder condition. Quality ratings were measured in quiet for each of the four examined vocoders using two different reconstruction bandwidths. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points.

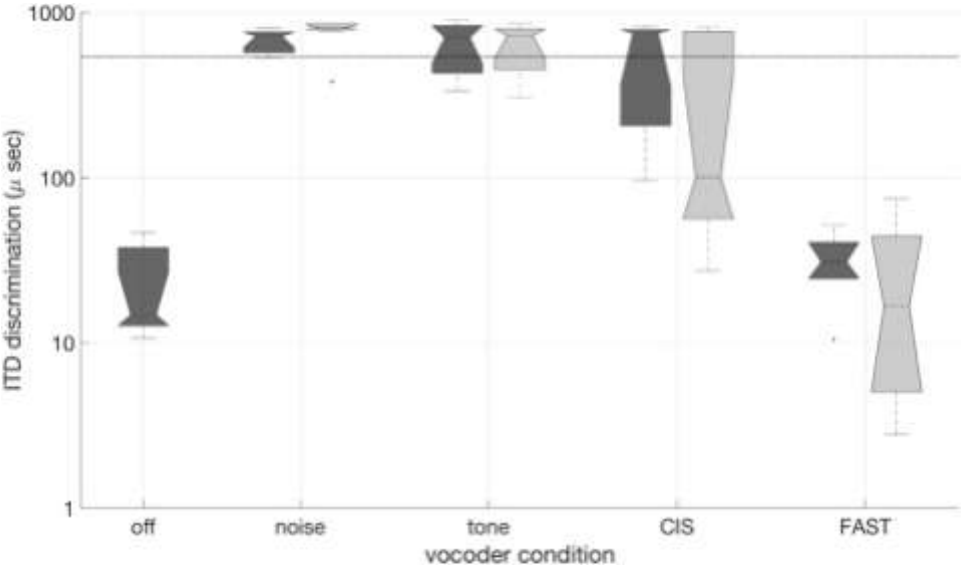

Fig. 6. Interaural timing discrimination (ITD) thresholds averaged across subjects. Thresh-olds were measured using 500 Hz tones mod-ulated by a 100 Hz raised-cosine envelope for each vocoder condition, as well as for an un-processed condition. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not treated as outliers, and outliers are plotted separately.

discrimination task better than chance when listening through the CIS vocoder with broad reconstruction is likely explained by better encoding of such envelope diff erences when using pulsatile carriers as compared to tones or noise bands.

As expected, ITD discrimination thresholds were well above chance for the FAST vocoder with either narrow or broad reconstruction meth-ods. This result was expected since the FAST algorithm places impulses at local maxima in the channel envelopes. While the FAST logic discards temporal fine structure of the underlying carrier signal, it encodes enve-lope timing cues with precision. Specifically, while the other approaches use diotic carriers to convey envelope modulations, the FAST approach uses impulse responses that are triggered by acoustic envelope events.

An analysis of variance was calculated on the measured ITD thresh-olds with vocoder type as the main factor and subject as a random factor. Vocoder type was significant ($F_{8,56}$ = 40.0, $p < 0.001$) and accounted for 82.2% of the variance. A multiple-comparisons analysis of ITD thresh-olds for the diff erent vocoder implementations were calculated based on the ANOVA model using a Tukey's honest significance diff erence criterion at a 0.05 significance level. At that level, ITD discrimination thresholds with the unprocessed sound and the two FAST vocoders did not significantly diff er. At that level, measured thresholds with the CIS vocoder using broad reconstruction diff ered from all other results, and measured thresholds with the noise, tone, and CIS (using narrow recon-struction) vocoders did not significantly diff er.

IV. DISCUSSION

Encoding temporal envelope and temporal fine-structure cues into CI stimulation is a challenge. It has been suggested that enhanced encoding of temporal cues could improve aspects of hearing for CI users including speech reception in noise and reverberation, musical pitch perception, and spatial hearing. The precise encoding of tem-poral cues, however, is complicated by stimulation issues such as electrical field interactions as well as by plasticity issues concerning

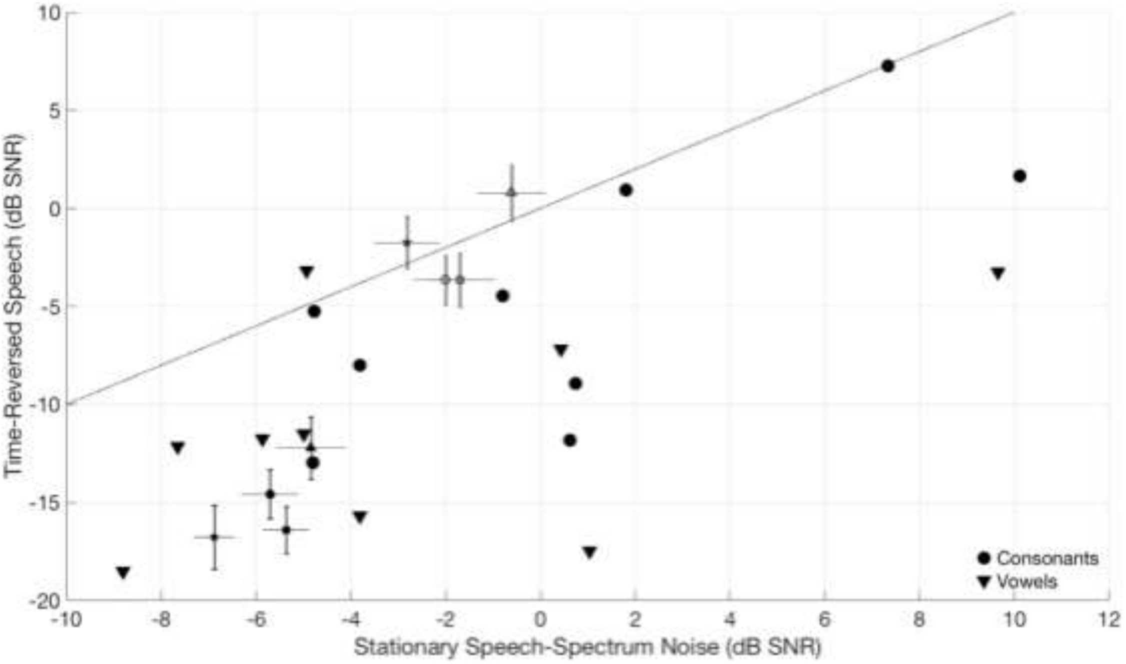

Fig. 7. Speech reception thresholds from Fig. 3 overlaid with the speech reception thresholds reported in Goldsworthy (2015). This figure illustrates how speech reception thresholds for normal-hearing listeners for the evaluated vocoders compare to actual CI users on similar measures. See body of article for detailed differences between acoustic conditions.

acquisition of encoded sensory cues. For example, it is highly likely that precise encoding of temporal cues would require substantial time for au-ditory training for the newly encoded cues. Because of the complexities associated with encoding temporal envelope cues in CI signal process-ing, the present study first examined the potential of encoding temporal envelope cues into vocoder representations of CI signal processing.

The present study compared vocoder methods in terms of speech re-ception in noise, speech quality in quiet, and sound lateralization based on interaural timing cues. The noise and tone vocoders have been established in the literature as useful models of CI signal processing as it pertains to speech reception. The impulse-response method introduced in this article is based on reconstructing pulsatile stimulation sequences produced by CI signal processing. A key advantage of such impulse-response vocoders compared to noise and tone vocoders is that the impulse-response reconstruction method allows the detailed temporal cues associated with CI stimulation to be examined. In this study, speech reception and quality ratings were examined to provide a framework for understanding the new vocoder methods with established ones, while ITD discrimination was examined to extend this framework to important dimensions of binaural psychophysics that depend on precise temporal stimulation.

The results from this study indicate that the FAST impulse-response vocoder produces comparable results to the noise and tone vocoders for speech reception. Speech reception thresholds measured in station-ary speech-spectrum noise and in time-reversed speech were not sig-nificantly different between the noise, tone, and FAST vocoders. The CIS vocoder produced speech reception results with similar trends, but was found to produce higher speech reception thresholds than the other vocoders when measured using time-reversed speech as background noise. This difference should be considered in future studies when con-sidering the CIS impulse-response vocoder. The details of how vocoder reconstruction transmits spectrotemporal information of band-limited signals may produce speech reception differences affected by across fre-quency smearing of information, demonstrated by others as important when considering modulated noise (Oxenham and Kreft 2014).

While such detailed consideration of reconstruction methods is im-portant for understanding differences in predicting speech reception in modulated noise, actual speech reception outcomes for CI users is presently not sufficiently characterized to conclude whether any of these methods is an overall better predictor. Specifically, the results of this study indicate that speech reception thresholds with the CIS vocoder were higher compared to the other vocoders, but it is important to keep in mind that the scale of predicted difference between the evaluated methods is smaller than the range of actual speech reception outcomes observed in CI users.

To illustrate this comparison of vocoder results against actual CI speech reception data, Fig. 7 overlays the SRTs measured for the present study with CI users' SRTs measured under similar acoustic conditions (Goldsworthy 2015). There are two important differences between the SRTs measured for the present study and the SRTs col-lected for Goldsworthy (2015). First, the speech materials used in Goldsworthy (2015) were consonant and vowel materials rather than CRM sentences. Second, the modulated background noise used in Goldsworthy (2015) was speech-spectrum noise gated on and off at a 10 Hz gating frequency rather than the time-reversed speech. Neverthe-less, this comparison of SRTs provides an indication of the wide range of speech reception differences observed with CI users. This comparison also indicates that the modeled differences between narrow and broad reconstruction methods captures the trend of the better performing CI users having lower SRTs in modulated compared to unmodulated back-ground noise. In other words, the better performing CI users do benefit from masking release in modulated background noise and this benefit is well modeled by the spectral resolution of vocoder methods.

Measured speech reception thresholds depend on the choice of speech materials. An advantage of the CRM materials is that they do not contain syntactic cues and thus can be used repetitively. However, since the CRM task is closed set, measured SRTs are typically much lower than when measured using relatively open set materials such as HINT and IEEE databases. For instance, Qin and Oxenham (2003) mea-sured SRTs in vocoder-processed speech with IEEE materials and gen-erally found SRTs to be around 0 dB SNR for a 24-channel vocoder and around +5 dB SNR for an 8-channel vocoder. Consequently, care must be given when comparing vocoder results across studies using different speech materials. These speech reception results contribute to the grow- ing body of knowledge regarding the importance of spectral resolution for masking release in fluctuating background noise. As an average indi-cator of this effect, the difference between speech reception thresholds measured in time-reversed speech compared to speech-spectrum noise averaged across vocoder algorithms was only −1.9 dB SNR for the broad reconstruction methods compared to −10.3 dB SNR for the narrow re-construction methods. Such evidence supports the hypothesis that spec-tral resolution is essential for providing CI users with improved mask-ing release in fluctuating noise. That a 10.3 dB masking release can be achieved with as little as 16 processing channels supports the notion that as few as 16 electrodes might provide substantial masking release in fluctuating noise if it were not for spectral smearing associated with current spread and auditory nerve pathology. Consequently, speech re-ception in fluctuating noise, or more specifically the masking release between fluctuating and stationary noise, is a sensitive measure for eval-uating the subtle perceptual effects resulting from novel CI stimulation strategies designed to spectrally sharpen electrode stimulation.

Before discussing the implications of the quality assessment and ITD discrimination results, it is important to consider the spectrotemporal details of how impulse vocoders differ from other vocoder methods. This impulse-response reconstruction method introduced in this article is comparable to Gaussian envelope tone vocoders (Lu et al., 2010). In fact, Gaussian envelope tone vocoders can be thought of as specific cases of the impulse-response method based in which the impulse response is defined as the Gaussian envelope tone, and the pulse logic is the use of a fixed, relative slow, rate pulse train. Typically, variations of Gaussian envelope tone vocoders have used modulation frequencies with a 100 Hz fundamental to drive the reconstruction process. Using a relatively low 100 Hz fundamental, thus constraining the temporal support, allows the temporal interaction of subsequent pulses to be avoided. What distin-guishes the impulse-response method described in this article, is that it was used to reconstruct arbitrary temporal pulse patterns from vocoder analysis.

For the CIS impulse-response vocoder described here, the underlying pulsatile rate was set to 250 Hz per channel. Typically, CIS implemented on clinical CI sound processors use stimulation rates of 800 Hz or higher. During the piloting stages for this study, it was determined that such high rates when used in an impulse-response vocoder produced sim-ulations with substantial audible distortions. Such audible distortions can occur for impulse-response vocoders from either spectrotemporal interference of subsequent impulse responses. As the bandwidth of the reconstruction filter decreases, the temporal duration of the impulse re-sponses increases and will consequently acoustically interfere with sub-sequent pulses. In other words, the acoustic pulses temporally interfere in a way that CI stimulation does not. For broader reconstruction filters, the temporal resonance of the filter is shorter and temporal interfer-ence is reduced across subsequent pulses, but the spectral interference increases. This is presumably why existing considerations of Gaussian envelope tone vocoders generally use relatively low modulations rates. Specifically, impulse responses do not have a sufficient time to decay before the subsequent pulse is generated. This consideration perhaps indicates that these vocoder methods are approaching a limit in how well any acoustic reconstruction method can "simulate" an electrical biphasic pulse delivered by a CI.

In terms of speech quality, when the vocoders were implemented with narrow reconstruction methods, the FAST vocoder was rated as having the highest quality. This is most likely because the FAST algo-rithm uses pulsatile patterns that are synchronous to the fundamental frequency of voicing. While the noise and tone vocoders convey a sense of pitch by envelope modulation of the respective carriers, it is likely that using filter banks to reconstruct pulsatile patterns that are syn-chronous to the fundamental frequency will produce a more salient rep-resentation of pitch. However, when the vocoders were implemented with broad reconstruction methods, the CIS vocoder was rated as hav-ing the highest quality, followed by equivalent performance between the tone and FAST vocoders, and with the noise vocoder having the lowest perceived quality. That the quality of the FAST vocoder decreased rela-tive to the other vocoders when using broad reconstruction can be un-derstood in terms of across channel pulsatile interference, which may be relevant to CI-user perception. Specifically, for the broad reconstruction methods, even though pulse timing is synchronous to the fundamental frequency within each channel, the broad reconstruction filters produce spectrally broad impulse responses that will interfere across cochlear locations in a normal-hearing listener. Such channel interaction is a rel-evant issue in the design of CI stimulation strategies, including stimu-lation strategies that attempt to use stimulation that is synchronous to the temporal fine structure within individual processing channels. How-ever, such conclusions should be drawn carefully since there are funda-mental differences between the band-limited impulse responses used to acoustically stimulate the cochlea compared to actual electrode bipha-sic pulses. A key difference is that band-limited impulse responses must temporally resonate over a moment of time to produce the band-limited characterization.

With regards to ITD discrimination, the impulse-response vocoder methods introduced in this article provide a way to explore sound pro-cessing strategies for CIs at the level of individual electrical pulses. The manner that pulse timing is encoded by these vocoder methods is mecha-nistically accurate in the sense that individual acoustic impulses are used to model individual electrical pulses. This mechanistic accuracy con-trasts with vocoder implementations that are based on such techniques as using correlated noise that instill a degree of interaural spectral and temporal fine structure to the signal (Swaminathan et al., 2016). How-ever, while the impulse-response vocoders are mechanistically accurate in how individual acoustic impulses are used to model individual elec-trical pulses, it is important to realize that there are differences between acoustic impulse responses and electrical biphasic pulses. Specifically, acoustic impulse responses ring in time dependent on the characteristics of the reconstruction filters, whereas electrical biphasic pulses do not. Consequentially, while the introduction of impulse-response vocoders provides a way to model individual biphasic pulses, there remains im-portant differences between these acoustic simulations and electrical stimulation.

The results from the present study indicate that acoustically degrad-ing modulated tones using an impulse-response vocoder method that incorporates FAST pulse generating logic is effective for preserving the acoustic cues needed for ITD discrimination. This is an important find-ing as it demonstrates a mechanistically accurate method for exploring CI signal processing that depend on precise stimulation timing. Such a model is timely in that it is widely anticipated that bilateral CIs may be improved in terms of binaural coordination of devices to encode in-teraural differences more precisely. Presently, clinical devices do not provide sufficient synchronization of bilateral devices to enable recipi-ents to take full advantage of latent binaural abilities. Until technology advances to the point that binaural CIs are programmed in a coordinated manner, in terms of loudness and pitch balancing as well as temporal synchronization, acoustic modelling approaches such as the impulse-response vocoders introduced here will be a useful method for explor-ing CI signal processing strategies that are being designed for the next generation of coordinated bilateral devices.

V. CONCLUSIONS

An impulse-response method for channel vocoders was introduced that provides flexibility in considering the temporal stimulation charac-teristics of CI signal processing. These impulse vocoders are mechanis-tically accurate in that individual impulse responses are used to model each electrical pulse generated by CI signal processing logic. The results indicate that two specific impulse-response vocoders based on CIS and FAST produce similar speech reception in noise as the well-established noise and tone vocoders, while producing higher ratings of speech qual-ity in quiet. Further, it was demonstrated that the FAST vocoder ap-proach successfully encodes the temporal envelope cues required for lateralizing sounds based on ITD cues. These results provide initial val-idation of impulse-response vocoders as models of CI signal processing that provide more control over the temporal properties of pulsatile stim-ulation.

REFERENCES

[1]. Arnoldner, C., Riss, D., Brunner, M., Durisin, M., Baumgartner, W., Hamzavi, J., 2007. Speech and music perception with the new fine structure speech coding strategy: preliminary results. Acta Oto Laryngol. 127 (12), 1298–1303.

[2]. Bacon, S.P., Opie, J.M., Montoya, D.Y., 1998. The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. J. Speech Lang. Hear. Res. 41 (3), 549–563.

[3]. Boghdady, N.E., Kegel, A., Lai, W.K., Dillier, N., 2016. A neural-based vocoder implemen-tation for evaluating cochlear implant coding strategies. Hear. Res. 333, 136–149.

[4]. Bolia, R.S., Nelson, W.T., Ericson, M.A., Simpson, B.D., 2000. A speech corpus for mul-titalker communications research. J. Acoust. Soc. Am. 107 (2), 1065–1066.

[5]. Chen, F., Loizou, P.C., 2011. Predicting the intelligibility of vocoded speech. Ear Hear. 32 (3), 331–338.

[6]. Deeks, J.M., Carlyon, R.P., 2004. Simulations of cochlear implant hearing using filtered harmonic complexes: implications for concurrent sound segregation. J. Acoust. Soc. Am. 115 (4), 1736.

[7]. Dorman, M.F., Loizou, P.C., 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 (4), 2403–2411.

[8]. Dudley, H., 1939. Remaking speech. J. Acoust. Soc. Am. 11 (1), 169.

[9]. Fu, Q., Shannon, R.V., Wang, X., 1998. Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing. J. Acoust. Soc. Am. 104 (6), 3586–3596.

[10]. Goldsworthy, R.L., 2015. Correlations between pitch and phoneme perception in cochlear implant users and their normal hearing peers. JARO 16 (6), 797–809.

[11]. Hervais-Adelman, A.G., Davis, M.H., Johnsrude, I.S., Taylor, K.J., Carlyon, R.P., 2011. Generalization of perceptual learning of vocoded speech. J. Exp. Psychol. Hum. Per-cept. Perform. 37 (1), 283–295.

[12]. van Hoesel, R.J.M., 2007. Sensitivity to binaural timing in bilateral cochlear implant users.

[13]. J. Acoust. Soc. Am. 121 (4), 2192.

[14]. van Hoesel, R.J.M., Tyler, R.S., 2003. Speech perception, localization, and lateralization with bilateral cochlear implants. J. Acoust. Soc. Am. 113 (3), 1617.

[15]. Jin, S., Nelson, P.B., 2006. Speech perception in gated noise: the effects of temporal reso-lution. J. Acoust. Soc. Am. 119 (5), 3097.

[16]. Kaernbach, C., 1991. Simple adaptive testing with the weighted up-down method. Percept.Psychophys. 75 (3), 227–230.

[17]. Kwon, B.J., Perry, T.T., Wilhelm, C.L., 2012. Sentence recognition in noise promoting or suppressing masking release by normal-hearing and cochlear-implant listeners. J. Acoust. Soc. Am. 131 (4), 3111–3119.

[18]. Kwon, B.J., Turner, C.W., 2001. Consonant identification under maskers with sinusoidal modulation: masking release or modulation interference? J. Acoust. Soc. Am. 110 (2), 1130.

[19]. Lu, T., Litovsky, R., Zeng, F.-G., 2010. Binaural masking level differences in actual and simulated bilateral cochlear implant listeners. J. Acoust. Soc. Am. 127 (3), 1479–1490.

[20]. Moore, B.C.J., Peters, R.W., Stone, M.A., 1999. Benefits of linear amplification and mul-tichannel compression for speech comprehension in backgrounds with spectral and temporal dips. J. Acoust. Soc. Am. 105 (1), 400–411.

[21]. Nelson, P.B., Jin, S., Carney, A.E., Nelson, D.A., 2003. Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. J. Acoust. Soc. Am. 113 (2), 961.

[22]. Oxenham, A.J., Kreft, H.A., 2014. Speech perception in tones and noise via cochlear im-plants reveals influence of spectral resolution on temporal processing. Trends Hear. 18, 1–21.

[23]. Peters, R.W., Moore, B.C., Baer, T., 1998. Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. J. Acoust. Soc. Am. 103 (5), 577–587.

[24]. Qin, M.K., Oxenham, A.J., 2003. Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. J. Acoust. Soc. Am. 114 (1), 446–454.

[25]. Schroeder, M.R., 1966. Vocoders: analysis and Synthesis. Proc. IEEE 54 (5), 720–735. Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270 (5234), 303–304.

[26]. Smith, Z.M., Kan, A., Jones, H.G., Buhr-Lawler, M., Godar, S.P., Litovsky, R.Y., 2014.

[27]. Hearing better with interaural time differences and bilateral cochlear implants. J. Acoust. Soc. Am. 135 (4), 2190–2191.

[28]. Swaminathan, J., Mason, C.R., Streeter, T.M., Best, V., Roverud, E., Kidd, G., 2016. Role of binaural temporal fine structure and envelope cues in cocktail-party listening. J. Neurosci. 36 (31), 8250–8257.

[29]. Vandali, A.E., van Hoesel, R.J.M., 2012. Enhancement of temporal cues to pitch in cochlear implants: effects on pitch ranking. J. Acoust. Soc. Am. 132 (1), 392–402.

[30]. Wilson, B.S., Finley, C.C., Lawson, D.T., Wolford, R.D., Eddington, D.K., Rabinowitz, W.M., 1991. Better speech recognition with cochlear implants. Nature 352 (6332), 236–238.