Formant priority channel selection for an "*n*-of-*m*" sound processing strategy for cochlear implants

Juliana N. Saba, Hussnain Ali, and John H. L. Hansen

Citation: The Journal of the Acoustical Society of America **144**, 3371 (2018); doi: 10.1121/1.5080257
View online: https://doi.org/10.1121/1.5080257
View Table of Contents: https://asa.scitation.org/toc/jas/144/6
Published by the Acoustical Society of America

Formant priority channel selection for an "*n*-of-*m*" sound processing strategy for cochlear implants

Juliana N. Saba, Hussnain Ali, and John H. L. Hansen[a)]
Cochlear Implant Processing Laboratory—Center for Robust Speech Systems, University of Texas at Dallas, Richardson, 800 West Campbell Road, Richardson, Texas 75080, USA

(Received 16 November 2017; revised 19 October 2018; accepted 6 November 2018; published online 17 December 2018)

The Advanced Combination Encoder (ACE) signal processing strategy is used in the majority of cochlear implant (CI) sound processors manufactured by Cochlear Corporation. This "*n*-of-*m*" strategy selects "*n*" out of "*m*" available frequency channels with the highest spectral energy in each stimulation cycle. It is hypothesized that at low signal-to-noise ratio (SNR) conditions, noise–dominant frequency channels are susceptible for selection, neglecting channels containing target speech cues. In order to improve speech segregation in noise, explicit encoding of formant frequency locations within the standard channel selection framework of ACE is suggested. Two strategies using the direct formant estimation algorithms are developed within this study, FACE (formant-ACE) and VFACE (voiced-activated-formant-ACE). Speech intelligibility from eight CI users is compared across 11 acoustic conditions, including mixtures of noise and reverberation at multiple SNRs. Significant intelligibility gains were observed with VFACE over ACE in 5 dB babble noise; however, results with FACE/VFACE in all other conditions were comparable to standard ACE. An increased selection of channels associated with the second formant frequency is observed for FACE and VFACE. Both proposed methods may serve as potential supplementary channel selection techniques for the ACE sound processing strategy for cochlear implants. https://doi.org/10.1121/1.5080257

[MIM] Pages: 3371–3380

I. INTRODUCTION

Cochlear implants (CI) are recognized as the most effective clinical device that restores auditory function of individuals with moderate to severe hearing loss. In general, CI sound coding strategies analyze the frequency content of incoming acoustic signals in an attempt to extract and represent speech features within the constraints of a CI system (e.g., limited temporal/spectral resolution). Some of these speech features are comprised of formants, vocal tract shape, glottal source excitation, and signal periodicity (Choi and Lee, 2012; Loizou, 1999; Rubinstein, 2004; Shannon *et al.*, 1995; Wouters *et al.*, 2015). In general, "*n*–of–*m*" sound processing strategies, popular in some clinical processors, determine the frequency bands associated with the highest spectral energy in order to activate "*n*" out of "*m*" electrodes in each stimulation cycle. For acoustic scenarios where noise is absent, this approach is both reasonable and effective; however, it may become much less effective in noisy or reverberant environments (i.e., noise-dominated channels are selected over more critical speech-content-dominated channels). Important for conveying the phonetic content of the speech signal, it is well known that formant frequencies (*F*1, *F*2, and *F*3) correspond to resonances in the vocal tract. The location and time variations of formants convey important speech perception cues to both normal hearing as well as hearing impaired individuals (Kewley-Port *et al.*, 2007). These cues, however, can be easily masked when the listener is exposed to diverse acoustic environments consisting of degrading noise types, reverberation, and additional talkers. CI users, in particular, experience additional speech understanding challenges in noise due to poor spectral resolution and the absence of fine temporal and spectral cues, which makes speech segregation in noise a very challenging task (Assmann and Summerfield, 2004; Fu *et al.*, 1998; Hazrati and Loizou, 2012; Kokkinakis and Loizou, 2011; Parikh and Loizou, 2005; Wouters and Vanden Berghe, 2001; Wouters *et al.*, 2015). Speech understanding with CI devices, to a large extent, depend on the efficacy of the sound processing strategies to effectively encode the speech signal (i.e., its temporal and spectral characteristics).

Phonetic information is important as it shapes the envelope spectrum and conveys both harmonic and periodic structure of speech. Linear predictive coding (LPC), for example, uses a linear speech production and synthesis model based on vocal-tract modeling and source excitation (Markel and Gray, 1976) to model the overall frequency response of the speech segments. LPC, which is a very popular prediction method for formant estimation, is a computationally efficient, all-pole representation of speech spectra and determines the frequencies associated with the resonances of the linear system (i.e., formants) (Deller *et al.*, 2000; Deng and O'Shaughnessy, 2003). Representing formants in an all-pole manner has been shown to have an envelope structure similar to the natural harmonic peaks (El-Jaroudi and Makhoul, 1991). This type of signal processing strategy can be referred to as "peak-picking." Combined with the short-term average energy, the mean-squared prediction error along with autocorrelation, the filter coefficients of the linear speech predictor define the digital filter governing the speech spectrum (Schafer and Rabiner,

a)Electronic mail: john.hansen@utdallas.edu

1970). Due to the physical nature of speech production, where time versus frequency structure reflects the phonetic content over more than a single frame, LPC techniques have been modified for computation efficiency and have been used successfully in automatic speech recognition applications (Alku *et al.*, 2013; El-Jaroudi and Makhoul, 1991; Schafer and Rabiner, 1970; Tao *et al.*, 2008).

Accurate formant estimation can be a challenging task depending on the speech signal characteristics [e.g., signal-to-noise ratio (SNR), back vowels vs other phonemes]. Historically, average vowel formant frequencies have been estimated by monitoring the relationship between $F1$ and $F2$ with the use of a Sonograph (Flanagan, 1955; Peterson and Barney, 1952). Vowels tend to follow a spectrally smooth pattern with rising and falling trends, whereas other phonemic classes such as nasals, stops, affricates, and fricatives can present missing formants, zeros, or high frequency content similar to that of noise which may introduce unintended distortions for traditional peak-picking methods (Assmann and Summerfield, 2004; Parikh and Loizou, 2005). For example, in the presence of noise, spectral tilt can be hidden, minor peaks can be added, and additional harmonics past the third formant may be lost (Assmann and Summerfield, 2004; Deng and O'Shaughnessy, 2003). Therefore, the accuracy of formant estimation may depend on the intelligible differences in frequency, known as difference limens (DL). Several studies have shown a 3%–5% difference in formant frequencies from monitoring the perception of small variations of $F1$ and $F2$ and their transitions (Flanagan, 1955, 1956b; Hawks, 1994). Depending on the spectral content of the noise, larger differences may be observed for $F2$ formant location than for $F1$ (Loizou, 2007). In addition, formant estimation may become challenging when the proximity of two formants are close together, as this may cause one large peak instead of two smaller peaks to represent each formant (i.e., a well-known challenge for back vowels) (Deller *et al.*, 2000; Hawks, 1994). Identification of individual formants that appear to be merged due to close approximation in the frequency spectrum can be achieved with the use of the chirp z-transform (CZT) (Rabiner *et al.*, 1969; Schafer and Rabiner, 1970). All techniques and challenges aside, effective encoding/delivery of formant cues has been shown to increase speech intelligibility (SI) for CI users (Blamey *et al.*, 1987; Dorman *et al.*, 1997; Fu *et al.*, 1998; Geurts and Wouters, 1999; Shannon *et al.*, 1995; Vandali *et al.*, 2000).

In the past few decades, advancements in CI technology has been a multi-disciplinary effort that includes improvements in sound processors, novel stimulation techniques, improved electrode designs, and surgical techniques, all of which have had a positive impact on the field (Loizou, 1998, 1999; Wilson *et al.*, 1993; Wilson and Dorman, 2008; Wouters *et al.*, 2015). In this study, a new channel selection technique is proposed which expands on existing "n-of-m" strategies and the prior success of formant estimation, mainly inspired by feature extraction based signal processing. The very first generation of Nucleus multi-electrode CI (by Cochlear Ltd.), used the $F0/F2$ sound coding strategy to convey voicing input from the fundamental frequency ($F0$). $F0$ has since been documented as an important cue for SI, as it provides information on the temporal structure of speech and can define the stimulation rate of CI (Assmann and Summerfield, 2004; Dowell *et al.*, 1987; Skinner *et al.*, 1991; Tao *et al.*, 2008). Peak amplitudes of the second resonant ($F2$) achieved by zero-crossing detection and bandpass filtering techniques in average ranges of the second formant peak slowly began providing additional frequency-based cues to listeners (Blamey *et al.*, 1984; Blamey *et al.*, 1987; Seligman *et al.*, 1984; Tong *et al.*, 1980). This strategy assigned bandpass filters to specific electrodes along a logarithmic scale, much like the physiological place coding of the cochlea. To further increase SI of those with Nucleus CIs, an additional zero-crossing measure was later added to provide information concerning the first two formants in the $F0/F1/F2$ strategy (Blamey *et al.*, 1987). The Multipeak (MPEAK) strategy was developed soon after as a solution that provides high frequency information as well as a means to convey more acoustic cues (Dowell *et al.*, 1987; Patrick and Clark, 1991; Skinner *et al.*, 1991).

The trend of increasing frequency-based features, and formants in particular, shifted towards a spectral analysis approach with the development of the spectral peak strategy (SPEAK) and the introduction of continuous interleaved sampling (CIS) sampling approach (Skinner *et al.*, 1994; Wilson *et al.*, 1993). Instead of feature extraction, bandpass filters were used to analyze frequency characteristics of short speech segments and deliver biphasic pulses to electrodes corresponding to frequency channels with the highest spectral energy (McDermott *et al.*, 1992; Wilson *et al.*, 1993). In each stimulation cycle, six electrodes out of the possible 16 stimulation sites were activated, which not only enabled the selection of the three formant peaks (or at least aimed to), but also excitation of adjacent filters to provide a larger spectral representation of speech (Loizou, 1999; McDermott *et al.*, 1992; Seligman and McDermott, 1995; Vandali *et al.*, 2000; Wilson *et al.*, 1993). Cochlear Ltd. incorporated this approach with the spectral peak strategy (SPEAK), which increased the number of bandpass filters and allowed for a variable range of maximum channels for stimulation (Seligman and McDermott, 1995; Skinner *et al.*, 1994). Advanced Combination Encoder (ACE) strategy, commonly used in a vast number of CI devices today, is classified as an "n-of-m" strategy. This approach commonly selects 8–12 ("n") out of 22 ("m") channels that encompass the highest spectral energy in each stimulation cycle and activates the corresponding intracochlear electrodes (Vandali *et al.*, 2000). This is one of many adapted strategies using the CIS approach which has proven to be effective in reducing channel interaction and can control the unnecessary activation of all electrodes simultaneously (Wilson *et al.*, 1991).

In a previous study by the authors, a channel selection technique was developed to designate three of the selected channels to the locations of $F1$, $F2$, and $F3$ for each stimulation cycle (Ali *et al.*, 2014). Improved sentence recognition performance was observed at 10 dB and 5 dB SNR conditions as well as in reverberation (Ali *et al.*, 2014). In this current study, VFACE is developed as a rationale for improved formant estimation accuracy by explicit encoding of formant frequencies only during the voiced portions of speech. This approach allows for both an accuracy check based on traditional signal processing and speech science approaches

TABLE I. Biographical data for CI subject participation.

Subject ID	Age (years)	Years implanted	Implant Type	Active Electrodes	Stim. Rate (Hz)	"*n*-maxima"
S1	60	7	CI24RE	20	900	8
S2	64	5	CI24RE	20	500	8
S3	70	9	CI24RE	21	900	8
S4	37	4	CI422	18	900	10
S5	64	7	CI24RE	21	900	8
S6	55	5	CI422	22	500	8
S7	69	10	CI512	22	900	8
S8	55	14	CI24R	18	900	12

(referred to as the enhanced formant estimation, or EFE algorithm) and a soft decision of formant channels for prioritization. Channels with high accuracy formants are prioritized for selection and the remaining selection are identified using the original "*n-maxima*" criteria. Formant estimation and channel selection was investigated in a prior version of FACE implemented for the entire speech signal (both voiced and unvoiced segments) using a high order LPC model without any error catching structure (Ali *et al.*, 2014).

In this study, the proposed FACE and VFACE strategies are also considered to be "*n*-of-*m*" strategies. The purpose of the improved formant estimation and proposed channel selection is to help increase SI performance of CI users in a wide variety of challenging acoustic conditions such as babble noise, speech shaped noise (SSN), and reverberation, where formant frequencies may be masked by interference. For noisy environments, sub–optimum channel selection may prevent the speech-dominant channels from being selected and stimulated. This study aims at increasing the accuracy of the previously proposed formant estimation algorithm with the use of a combination of improved formant detection techniques. An improvement or comparable performance is hypothesized as a result of the newly proposed channel selection criteria. The goal of this study is to test the efficacy of proposed algorithms in wide variety of noise types.

II. MATERIALS AND METHODS

A. Subjects

Eight adult post-lingually deafened CI users (two unilateral, six bilateral) were recruited for this study and paid for their participation. All participants were native speakers of American English, implanted with Nucleus 24 (Cochlear Ltd.) implant systems, and had at least 5 years of experience with their device. Inclusion in this study required the routine use of the ACE sound coding strategy in their clinical processor. Biographical data of subjects is summarized in Table I. For bilateral users, the information demonstrated in Table I reflects the ear tested in this study. The average age of the participants was 59 years old with a range of 37–70 years old.

B. Speech material

Sentences from IEEE (IEEE, 1969), Arizona Bioindustry Association (AzBio) (Spahr *et al.*, 2012), and the consonant-nucleus-consonant (CNC) databases (Peterson and Lehiste, 1962) were chosen to evaluate SI in 11 different acoustic conditions. All eleven conditions are outlined and referenced in Table II. Continuous SSN was added to IEEE sentences to simulate 10 dB and 5 dB SNR conditions. Two reverberation conditions ($T_{60} = 300,\ 600$ ms) as well as a combination of reverberation and noise ($T_{60} = 600$ ms and 10 dB SSN) were also included. Reverberant conditions were generated by convolving room impulse responses (RIR) as described in Neuman *et al.* (2010). In addition, four-talker babble noise was added to AzBio sentences to simulate 10 and 5 dB SNR conditions. The CNC word database consisted of three-syllable words with a consonant-nucleus (vowel)-consonant structure, where the whole word as well as individual phonemes were scored separately for the same word (Hillenbrand *et al.*, 1995). All speech material were directly streamed using the UT-Dallas CCi-MOBILE research platform (Ali *et al.*, 2014; Ali *et al.*, 2016; Ali *et al.*, 2017) to the subject's implant unilaterally.

C. Signal processing

Two algorithms, EFE and channel prioritization (CP), were implemented within the pipeline of the ACE strategy (Cochlear Ltd.) (Vandali *et al.*, 2000). VFACE and FACE, the strategies using these algorithms are referred to in this study as individual sound processing strategies. Note, the difference between FACE/VFACE and ACE strategies lies in different channel selection processes as demonstrated in Fig. 1.

For clinical standard ACE, the acoustic signal is passed through a 22-channel filterbank implemented using the Fast-

TABLE II. Simulated acoustic conditions presented to CI users in this study include two different noise types with two different SNR ratios, two reverberation conditions where T_{60} values represent the reverberation time, a reverberation, and noise combination, as well as individual words and phonemes presented without any noise or reverberation. Acoustic conditions are specified in the remainder of the study using the shorthand notation in the first column.

Condition	Full Description	Database	Noise Type	SNR	T_{60}
C-1	Quiet	IEEE	—	—	—
S-1	10 dB SSN	IEEE	SSN	10 dB	—
S-2	5 dB SSN	IEEE	SSN	5 dB	—
R-1	300 ms reverb	IEEE	Reverb-only	—	300 ms
R-2	600 ms reverb	IEEE	Reverb-only	—	600 ms
R-3	Reverb and noise	IEEE	SSN	10 dB	600 ms
C-2	Quiet	AzBio	—	—	—
B-1	10 dB babble	AzBio	Babble	10 dB	—
B-2	5 dB babble	AzBio	Babble	5 dB	—
W-1	CNC words	CNC	—	—	—
P-1	CNC phonemes	CNC	—	—	—

Fourier-transform (FFT). The spectral energy ("*n-maxima*") of 22 channels is calculated according to the bandpass filters with frequency allocations determined from Cochlear Corporation (Cochlear Ltd.). The signal is then compressed logarithmically, and current levels are generated within the dynamic range(s) of the selected electrodes governed by the subject's frequency-to-electrode configuration (MAP) for each stimulation cycle. VFACE is processed in the same way as FACE, except that custom channel selection is only employed for the voiced segments of speech signal. An open-source, voiced activity detector (VAD) was adapted and used to estimate voice and unvoiced speech segments (Kadir, 2008).

1. EFE algorithm

To estimate formant frequencies, the short term log energy (STLE) of speech frames is calculated to determine the presence of a primary speaker before LPC. A 28-order LPC model generates coefficients for every 8 ms stimulation cycle (i.e., a temporal frame). Unlike the previous version of FACE (Ali *et al.*, 2014), formant bandwidths and frequencies are validated for accuracy against the range of average formant frequencies developed for eleven English vowels and then subjected to a five-step error catching structure (Deller *et al.*, 2000; Flanagan, 1956a,b; Hillenbrand *et al.*, 1995; Rabiner *et al.*, 1969). Three formants from the previous 8 ms window are stored and used to compare against formants of the current window as a continuity check to ensure fluid vowel movement characterized with a 150–200 Hz bandwidth (O'Shaughnessy, 2008). Before formants are prioritized in channel selection, each formant undergoes accuracy/error verification. If error is detected for $F1$, the CZT is calculated between 200–900 Hz to determine an additional peak in the spectrum. Based on a similar strategy from (Flanagan, 1955, 1956a), the corrected formant calculation is compared to the average vowel formant range within 3% of the previously selected formant frequency to stay within the range of difference limens (DLs) (Flanagan, 1955; Hillenbrand *et al.*, 1995). For $F2$ and $F3$, additional formant candidates are calculated from the LPC coefficients that fit within the average vowel formant range to serve as an alternative peak (Hillenbrand *et al.*, 1995).

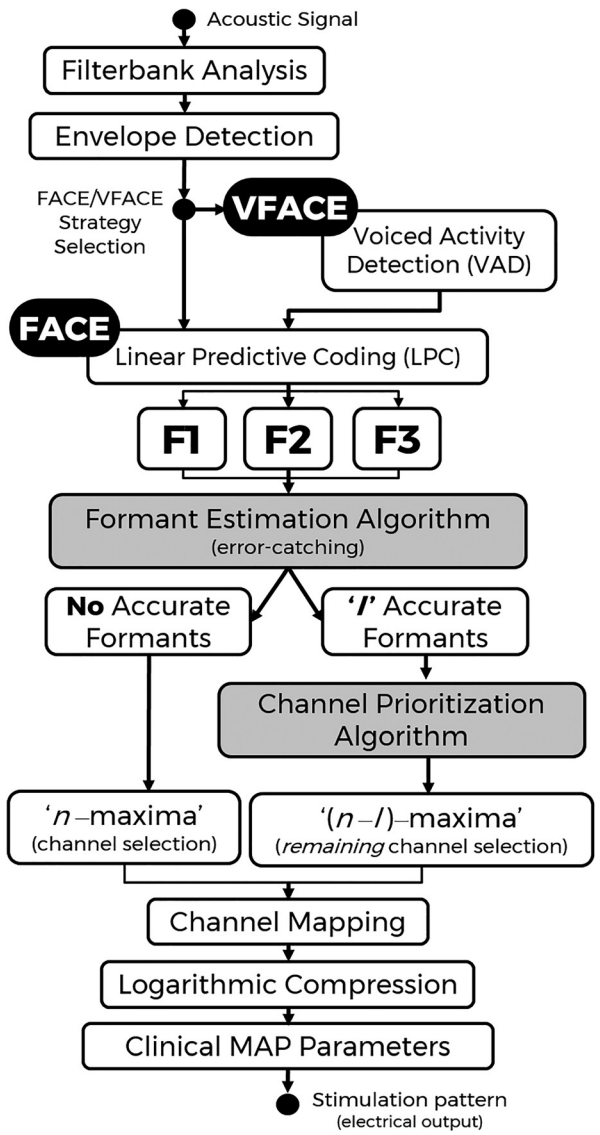

FIG. 1. Block diagram of the formant estimation strategies (FACE and VFACE) embedded within the ACE processing strategy. EFE and CP algorithms used for channel selection in FACE/VFACE are shaded in gray. '*l*' represents the soft decision of the EFE algorithm to prioritize channels outside of the 'n-maxima' decision.

2. CP algorithm

The CP algorithm is employed if formants individually meet the accuracy criteria determined by the EFE algorithm. Any number of channels (0–3), representing any combination of $F1$, $F2$, or $F3$ are selected for stimulation where the remaining channels are selected using the "*n-maxima*" criteria, i.e., highest spectral energy. Generally, the channels with the lowest spectral energy are deselected to ensure selection of prioritized channels representative of formant frequencies. If accurate formant estimation cannot be achieved for a particular frame, the channels with the highest spectral energy ("*n-maxima*") are selected, as in standard ACE processing.

D. Procedure

Speech intelligibility (SI) tests were conducted for the two proposed CI sound coding strategies against standard clinical ACE processing for eleven simulated acoustic conditions. Prior to testing, MAP files were obtained for each subject and programmed into the CCi-MOBILE sound processor (Ali *et al.*, 2014; Ali *et al.*, 2016; Ali *et al.*, 2017). Informed consent and approval from the institutional review board were obtained before subject testing. For bilateral implant users, the subjects were asked to subjectively select their better ear for testing. Subjects with residual hearing in the non-implanted ear (unilateral subjects) were provided with an ear plug to place in the contralateral ear to prevent listener distraction if residual hearing existed (both unilateral subjects had profound hearing-loss in the hearing-aid ear). Speech battery was tested within a lab-environment (in a direct–connect scenario) using offline processing techniques.

Both conditions and algorithms (ACE, FACE, VFACE) were randomized across all subjects in this study. The

subjects were asked to verbally repeat back each sentence (or word) they perceived, and their response was recorded for intelligibility measures. No repetitions of an individual token were provided to the subjects. Average intelligibility for each sentence was scored as a percentage of the total number of words presented. Twenty sentences were used to score each condition from the IEEE and AzBio databases. Fifty words were used from the CNC database to assess the effect of each strategy. Subjects were able to adjust the volume of the stimuli only at the start of the test (during the training phase) and were given the option to stop or to take a break throughout the entire testing duration to avoid fatigue. Each subject was presented 150 words and 540 sentences; the average testing time was 3.25 h/subject.

Post-processing of SI performance included statistical analysis of SI per condition, analysis of channel selection per strategy, and individual subject performance in each strategy for each condition. Histograms were used to determine if channel selection differed in each strategy and if there existed trends in channel selection according to the types of noise investigated. Each histogram was normalized to the total number of channels selected in each individual sentence from either database (1800 total sentences). Averages channel selection profiles were calculated for babble and SSN noise types, reverberation, as well as all nine conditions combined (excluding CNC words and phonemes). Statistical significance was determined from within—subjects repeated measures analysis of variance (ANOVA) to determine the effect of the three signal processing strategies (ACE, FACE, VFACE), 11 simulated acoustic conditions (B-1, B-2, C-1, C-2, R-1, R-2, R-3, S-1, S-2, W-1, P-1), and their effect on speech intelligibility. Statistical analysis was performed using SPSS (IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0, IBM Corp., Armonk, NY) with an α set to 0.05. Pairwise comparisons were adjusted using Bonferroni corrections.

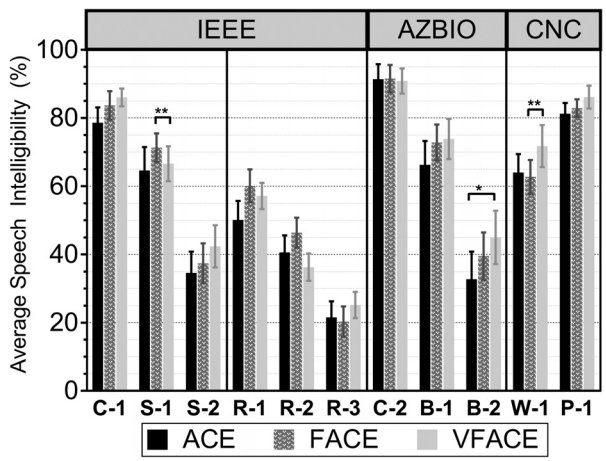

FIG. 2. Mean intelligibility scores of each of the eleven conditions used in this study. Standard error of the mean is represented by the error bars. Statistical significance is represented from strategy interaction ($p < 0.05$) by repeated measures ANOVA using Bonferroni corrections. * denotes significance ($p < 0.05$) between ACE and VFACE; ** denotes significance ($p < 0.05$) between FACE and VFACE.

III. RESULTS

A two-way, repeated-measures ANOVA was performed to determine the effect of strategy and condition on SI. A statistically significant effect of sound processing strategy [$F(2,14) = 7.909$, $p = 0.005$] and acoustic condition [$F(10,70) = 43.775$, $p < 0.001$] was observed. The interaction of strategy and condition was also significant [$F(20\,140) = 2.356$, $p = 0.02$]. For all conditions within the speech battery, comparable results (within five percentage points) were obtained for both FACE and VFACE strategies versus the standard ACE strategy. As expected, quiet conditions (C-1, C-2) yielded the highest SI results out of all the tested conditions. As the difficulty within each noise/reverb type increased (decreasing SNR level, increasing T_{60} value), the average intelligibility decreased. Pairwise comparisons of strategy using Bonferroni corrections resulted in a significant difference between ACE and FACE ($p = 0.047$), but not with VFACE ($p = 0.075$) or between FACE and VFACE ($p = 0.595$).

Figure 2 demonstrates mean intelligibility scores of each processing strategy. A significant improvement ($p < 0.05$) of 12.3 percentage points was observed with VFACE for B-2 (AzBio, 5 dB SNR). In B-1 (AzBio, 10 dB SNR), an increase of SI was observed from 66.29% with ACE to 73.84% with VFACE ($p > 0.05$). Observation of the two proposed strategies for S-1 resulted in a 6.7 percentage point difference ($p < 0.05$) between 71.30% (F) and 66.59% (V). Average SI of 62.75% (F) and 71.50% (V) in W-1 was also observed to be significant ($p < 0.05$). No significant improvements in SI were found for the reverberation conditions (R-1, R-2, or R-3), regardless of a 10.1 (F) and 7.0 (V) percentage point difference.

Figure 3 shows channel selection patterns for nine acoustic conditions (excluding W-1, P-1), where percent differences were calculated against standard ACE. Overall, ACE demonstrated more frequent selection of channels 1–7 (corresponding

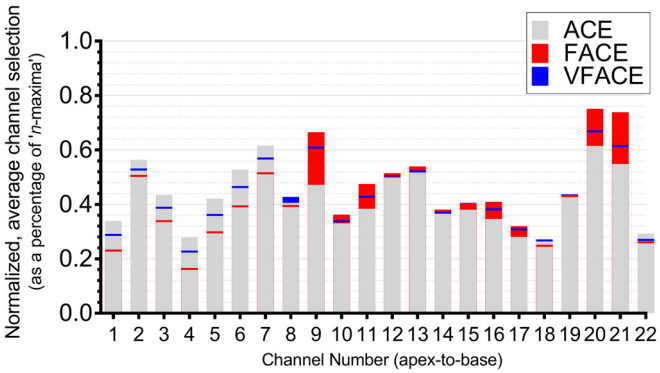

FIG. 3. (Color online) Normalized, average channel selection as a percentage of the "*n-maxima*" selected in each stimulation cycle, across all sentence tokens for nine conditions used in the study (with the exception of CNC words). Light gray bars represent the normalized percentage of each channel selection using standard ACE strategy with "*n-maxima*" channel selection; red bars represent the normalized percentage of individual channel selection using the formant-prioritization strategy FACE; the blue lines represent the individual channel selection using the formant-prioritization strategy VFACE. For channels 1–8, FACE and VFACE were selected on average less frequently than those associated channels with ACE. Conversely, for channels 9–22, red bars and blue lines can be seen to demonstrate the increased selection of channels compared to standard ACE.

to frequencies: 188–1063 Hz), whereas FACE and VFACE selected channels 9–12 (corresponding to frequencies: 1188–2063 Hz) more frequently. FACE and VFACE selected channels 1–7 34.88% and 22.83% less than ACE, respectively; FACE and VFACE selected channels 8–12 18.61% more than ACE. Figure 4 demonstrates the average channel selection for channels 8–12 for sentence tokens indicative of the test battery for all subjects in this study. The most noticeable difference in channel selection occurs for channel 9 (corresponding to frequency range of 1188–1313 Hz). Channels 20–21 (corresponding to frequencies: 5313–6983 Hz) for FACE and VFACE were also selected on average more than ACE.

IV. DISCUSSION

The overall objective of this study was to evaluate SI performance of CI users using a new channel selection criteria embedded in clinical/standard ACE strategy. It is well known that as the SNR of the speech signal decreases (e.g., in environments with noise and reverberation), CI user's ability to decode speech decreases. It was hypothesized that the standard channel selection process in "*n*-of-*m*" strategies may not be an ideal solution for the representation of complex sounds masked by broad-band noise because channel selection is entirely based on highest spectral energy in each frequency band. The two proposed strategies, FACE and VFACE, were developed as a solution to improve SI by implementing a different criteria for channel selection based on accurate formant estimation. The original hypothesis of this work was to determine the effects of selection/stimulation of channels representing the formant frequencies (*F*1–*F*3) of a speech segment on SI for CI users.

To evaluate the behavior of channel selection each strategy, a one-to-one comparison of "*n-maxima*" between ACE, FACE, and VFACE was performed on a frame-by-frame basis. Discrepancy in channel selection indicate the lack of selecting channels with formant frequencies (either *F*1, *F*2, or *F*3). When a difference is identified, FACE and VFACE strategies will ensure the stimulation of prioritized channels via the corresponding electrode. Individual channel (1–22) differences were quantified for the speech battery experienced by CI users for 9 of the 11 acoustic conditions (C-1, C-2, S-1, S-2, B-1, B-2, R-1, R-2, and R-3). CNC words and phonemes were excluded due to token length and decreased formant transitions for voiced and unvoiced portions of speech. Using a standard 22-electrode MAP (threshold-clinical-level, THR = 100, maximum-comfort-level, MCL = 200), FACE and VFACE on average resulted in a range of 4%–22% difference for individual channel selection. The use of the CP algorithm used in FACE and VFACE, increased alternate channel selection from quiet, to reverberation, to noisy conditions. The data from this study indicates that CI users may or may not be perceptually sensitive to small spectral changes resulting from channel selection, especially for acoustic conditions where FACE or VFACE did not differ from ACE. FACE and VFACE were presented to each CI user in an acute manner; therefore, if channel selection was 80%–96% similar to ACE, CI users may not be able to perpetually discriminate the effects of such small changes. The lack of significance in the intermediate conditions (S-1, B-1, R-1) may be further explained by a future investigation of individual perceptual sensitivity levels of each CI user in each condition.

The baseline performance (i.e., with standard ACE) of some of the subjects appeared to be at a high-performing level (80%–100% SI on average) in quiet, noise-free conditions. Splitting the subjects by their performance with ACE in quiet conditions provided the ability to analyze possible ceiling effects. Subjects S2, S5, S6, and S7 were considered high performing while subjects S1, S3, S4, and S8 were considered low performing. A one-way ANOVA was performed

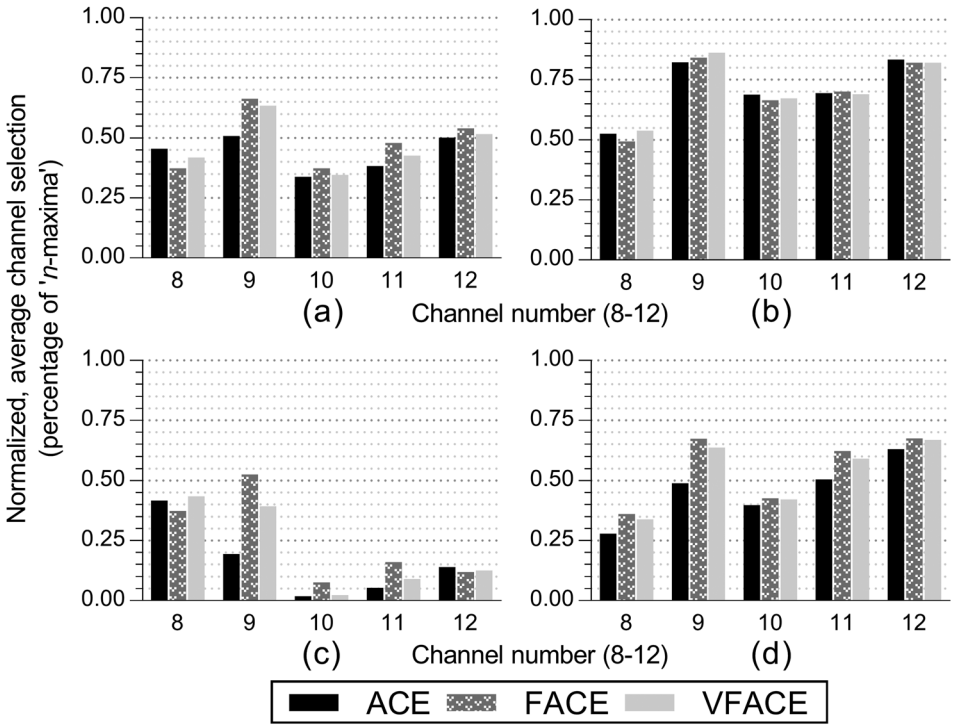

FIG. 4. Average channel selection histograms for: (a) all nine conditions: C-1, C-2, S-1, S-2, R-1, R-2, R-3, B-1, and B-2, (b) R-1 and R-2, (c) S-1 and S-2, and (d) B-1 and B-2 for channels 8–12 (or electrodes 11–15) associated with frequency range 1125–1938 Hz (according to the default frequency allocations for analysis bands used in ACE sound processing strategy). Channel selection is represented as a percentage of channel selection according to "*n-maxima*" for each sentence token, averaged across the total number of sentences used.

to analyze the effects of sound processing strategy of each group and revealed a non-significant ($p > 0.05$) trend of improved performance with the FACE/VFACE strategies for subjects who exhibited lower performance with ACE (60%–80% SI). However, for high-performers, the lower degree of change in channel selection observed for FACE/VFACE strategies may explain the performance variability in subjects with little improvement for ACE. At most, the number of channels selected according to formant frequencies is three, and the remaining channels are selected using the "*n-maxima*" criteria. Changes in channel selection were also found to be subject–dependent (i.e., due to individual MAP characteristics, proximity of formants may not always result in three distinct channel selections, one per formant). If the proximity of formants are close together in frequency (for example, a back vowel such as /a/), one channel may be selected or prioritized instead of two distinct channels to represent those individual formants. Individual subject channel selection was not analyzed as part of this study, but may be further investigated to determine how MAP parameters play an important role in channel selection for representing each formant.

FACE/VFACE strategies selected on average more channels in the low–mid frequency range as opposed to the low frequency channels. Figure 4 illustrates the histogram for channels 8–12 with frequencies associated with the second formant range (in Hz). A slight increase in selection of high frequency channels 19–21 (5313–6938 Hz) is also noted. For R-1 and R-2, FACE and VFACE demonstrated comparable channel selection behavior. It should be noted that the channel selection is one of the many parameters that affect speech perception; a large number of dependent factors, such as encoding of temporal cues, "*n-maxima*," stimulation rate, etc., all contribute towards overall SI performance (Brown and Bacon, 2011). Both strategies demonstrated comparable SI to ACE in quiet and can be implemented within the processing framework of ACE without impacting performance. Nevertheless, the results of this study demonstrate the presence of a relationship between perceptual outcomes and channel selection.

Despite numerical improvements in SI scores in the formant-prioritization strategies, no significant difference was observed in the reverberation conditions nor the SSN conditions with either of the proposed strategies. The lack of a clear, reproducible relationship between FACE/VFACE and SI prove that further experimentation is needed to determine the effects of the signal processing approach in reverberant conditions (R-1, R-2, and R-3). Reverberation alone, and in combination with noise, presents a higher degree of difficulty for CI users as shown by a wide range of studies (Assmann and Summerfield, 2004; Hazrati and Loizou, 2012; Kokkinakis and Loizou, 2011; Neuman *et al.*, 2012; Parikh and Loizou, 2005). S4, S6, and S8 admitted difficulty understanding and hearing in natural environments similar to the simulated testing conditions of reverberation. S4, S5, and S6 stated the duration of some sentences were too short to become accustomed to the noise, which hindered their ability to segregate speech from noise. The spectral characteristics of reverberation can be described as a time-frequency smearing phenomenon, where the same peak-to-peak structure will repeat in subsequent frames causing the peak-to-valley ratio to decrease. Results from the prior evaluation of FACE (Ali *et al.*, 2014), indicated approximately a 20% gain in speech intelligibility in the SSN and reverberation combination. However, the original hypothesis regarding possible improvements in SI performance was not upheld as data from the present study suggests. Traditional approaches to address reverberation, such as speech enhancement, noise-suppression, including front-end pre-processing algorithms, are likely to be better options for speech signal enhancement and boost performance levels (Furuya and Kataoka, 2007; Hu and Loizou, 2008; Kokkinakis and Loizou, 2011; Wouters and Vanden Berghe, 2001).

Channel selection in FACE and VFACE used the extraction of formant frequencies to ensure the stimulation of a channel/electrode (spectral analysis), but these schemes do not improve the temporal fine structure associated with formant(s). Feature extractions strategies such as *F*0/*F*2, *F*1/*F*2/*F*3, MPEAK (Dowell *et al.*, 1987; Seligman *et al.*, 1984) were originally used in earlier generation of sound processors. Former approaches, including *F*0/*F*1/*F*2 and the MPEAK strategies, have been investigated using different stimuli and thus compared to determine the effects of increasing speech understanding with increasing frequency representation (Firszt *et al.*, 2009; McDermott *et al.*, 1992; Nogueira *et al.*, 2005; Skinner *et al.*, 1991; Vandali *et al.*, 2000; Wouters *et al.*, 2015). In a comparative study with 63 post-lingually deafened CI users, 80% of individuals subjectively reported that the SPEAK strategy enabled them to listen as well as, or better, in a wide variety of listening conditions versus the MPEAK strategy (Skinner *et al.*, 1994). Although the average scores on the CUNY/SIT sentences at 15 and 10 dB SNR babble noise reflected statistically insignificant improvement, 34 out of 40 subjects and 43 out of 58 subjects, respectively, demonstrated significant ($p < 0.05$) improvement in words correct using SPEAK strategy (Skinner *et al.*, 1994). This improvement in SI reflected the comparable and positive improvement of increasing the number of electrodes stimulated in each stimulation cycle as well as increasing the amount of temporal and spectral information.

In general, improvements in this study from standard ACE were observed with VFACE in the most difficult noise conditions (S-2, B-2, R-3). This improvement is an indication that in low-level SNR conditions, explicit channel selection of formant frequencies may facilitate in obtaining improved intelligibility. By combining front-end speech enhancement algorithms with the proposed channel selection, performance levels may potentially be improved further. Prioritizing the stimulation of channels containing formant locations can be implemented independent of the processing strategy. FACE and VFACE have the ability to alter the number of bands selected based on the highest spectral energy. A number of researchers have investigated SI as a function of number of the electrodes. Shannon *et al.* (1995) varied the number of frequency bands to determine the effect of temporal cues for speech recognition. The authors of that particular study found that intelligibility can be achieved with only a small number of bands for spectrally-rich or spectrally-degraded speech because of the cues given by the temporal structure (Shannon *et al.*, 1995). Similarly,

Dorman *et al.* (1997) showed that signal processors with a small number of channels, between 5 and 8, resulted in high levels of speech understanding. These studies provide rationale for the importance of effective identification and selection of frequency channels that contribute most to speech intelligibility.

Similar to the motivation of this study, Nogueira *et al.* (2005) developed an "*n-of-m*" strategy using psychoacoustic-masking models (PACE) to identify speech-dominant bands while stimulating a smaller number of effective electrodes. In that work, the psychoacoustic-based model was applied within the ACE processing framework and their data indicated that performance with eight electrodes using ACE could be achieved using PACE with four channels. Noise and reverberation within the speech signal may reduce intelligibility, regardless of the number of channels stimulated. The two proposed strategies in this study can serve as a possible solution to improve performance within these adverse acoustic conditions by prioritizing the formant bands in addition to the "*n-maxima*" selected by standard ACE. Other strategies such as HiRes, HiRes120, fine structure processing (FSP), and MP3000 use different techniques (not mentioned here) to deliver temporal information by means of higher stimulation rates and current steering (virtual channels) to increase the spatial resolution of speech (Clay and Brown, 2007; Firszt *et al.*, 2009; Koch *et al.*, 2004; Wouters *et al.*, 2015). While concepts of recently developed strategies demonstrate promise, the current proposed strategy in this investigation is one of the envelope-based strategies, specifically for "*n*-of-*m*" type methods.

The proposed channel selection is dependent on both LPC and VAD. The VAD involved in the processing scheme of VFACE attempts to increase formant accuracy by only calculating formant estimation during voiced portions of speech. Analysis using CNC phonemes indicated the strength of VFACE without interference of noise, but not for the FACE strategy. Although a significant difference ($p < 0.05$) was observed between FACE and VFACE for W-1, a full phoneme analysis is needed to explain the effects of the signal processing approaches on both vowel and phoneme recognition. Parikh and Loizou (2005) studied the effects of manually selecting the formant frequencies ($F1$ and $F2$ estimated using a 22-pole LPC model) to compare the same vowel spoken in quiet and in noise. Their results demonstrated approximately 10% difference in accurate formant detection for 10 and 5 dB SNR SSN. The proposed strategy estimates formant locations on a frame-by-frame basis using the pipeline shown in Fig. 1. Thus, better formant estimation techniques could be used to iteratively determine the accuracy needed to significantly improve intelligibility, independent of the acoustic condition.

Some limitations to the approach used in the development of the two proposed strategies include: LPC and VAD parameters, computational efficiency, and formant estimation methods. Formant estimation from LPC in the presence of any simulated noise condition were shown to be susceptible to the synergistic effects of noise and reverberation, much like any algorithm without noise suppression, preprocessing, or speech enhancement. Computational power is an important factor in the real-time development of FACE and VFACE into commercial sound coding strategies. LPC in noisy conditions may result in a more flattened envelope spectrum of speech and reduce the visible valleys between each of the formant peaks (Chen and Loizou, 2004). The components of the error catching structure within FACE and VFACE were implemented for a high order LPC model at a low computational cost (Snell and Milinazzo, 1993; Zapata *et al.*, 2004); however, offline computation of R-2, R-3, and B-2 conditions appeared to require more processing time than other conditions. As model order of LPC increases, the increased number of pole pairs can lead to misrepresentation of formant resonances. Formant estimation can be quantified independent of the VAD by determining ground truth of the formant frequencies. Other VAD/SAD (speech activity detection) solutions have been developed, but the current solution used in this work was chosen as it is easily available to the research community (Kadir, 2008; Sadjadi and Hansen, 2013). A longitudinal study on the proposed strategies may investigate the performance of FACE and VFACE in real-world environments for CI users, as the results discussed here were assessed in an acute manner.

V. CONCLUSION

In this study, speech intelligibility was investigated between standard ACE and two alternate proposed signal processing strategies, FACE and VFACE using channel selection prioritizing formant frequencies. Significant improvements in speech intelligibility were observed in only one of the two low-level SNR conditions 5 dB SNR babble noise ($p < 0.05$). An increased selection of channels associated with the second formant frequency were found for FACE and VFACE strategies. No significant difference ($p > 0.05$) was found using the proposed channel selection technique for the reverberation conditions. Results from this study indicate that the combination of noise and reverberation may present difficulties in robust channel selection using a standard energy-based "*n-maxima*" criteria, yielding selection of noise-dominant bands. The innovation of the proposed channel selection used in FACE and VFACE strategies offer promise to aid in speech segregation between acoustically quiet and noisy conditions. Additional techniques, however, need to be explored further in order to address the compounding problem of reverberation and noisy reverberant environments.

ACKNOWLEDGMENTS

This work was supported by Grant No. R01 DC010494-01A from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

Ali, H., Ammula, S., and Hansen, J. H. L. (**2017**). "Subjective evaluation with UT–Dallas research interface for cochlear implant users," in *Proceedings of the Conference on Implantable Auditory Prostheses (CIAP 2017)*, July 16–21, Lake Tahoe, CA, p. 211.

Ali, H., Hong, F., Hansen, J. H. L., and Tobey, E. (**2014**). "Improving channel selection of sound coding algorithms in cochlear implants," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'14)*, May 4–9, Florence, Italy, pp. 905–909.

Ali, H., Hong, F., Wang, J., and Hansen, J. H. L. (**2016**). "Mobile research interface for cochlear implants," in *Proceedings of the International*

Conference on Cochlear Implants (CI2016), May 11–14, Toronto, Canada, pp. 27.
Alku, P., Pohjalainen, J., Laukkanen, A. M., and Story, B. H. (**2013**). "Formant frequency estimation of high-pitched vowels using weighted linear prediction," J. Acoust. Soc. Am. **134**, 1295–1313.
Assmann, P. F., and Summerfield, Q. (**2004**). "The perception of speech under adverse conditions," in *Handbook of Audiology Research*, edited by W. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, New York), pp. 231–308.
Blamey, P. J., Dowell, R. C., Clark, G. M., and Seligman, P. M. (**1987**). "Acoustic parameters measured by a formant-estimating speech processor for a multiple-channel cochlear implant," J. Acoust. Soc. Am. **82**, 38–47.
Blamey, P. J., Dowell, R. C., Tong, Y. C., Brown, A. M., Luscombe, S. M., and Clark, G. M. (**1984**). "Speech processing studies using an acoustic model of a multiple-channel cochlear implant," J. Acoust. Soc. Am. **76**, 104–110.
Brown, C. A., and Bacon, S. P. (**2011**). "Fundamental frequency and speech intelligibility in background noise," Hear. Res. **266**, 52–59.
Chen, B., and Loizou, P. C. (**2004**). "Formant frequency estimation in noise," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)*, May 17–21, Montreal, Canada, pp. I–581–584.
Choi, C. T. M., and Lee, Y. H. (**2012**). "A review of stimulating strategies for cochlear implants," in *Cochlear Implant Research Updates*, edited by C. Umat, and R. A. Tange (InTech Europe, Rijeka, Croatia), pp. 77–90.
Clay, K. M. S., and Brown, C. J. (**2007**). "Adaptation of the electrically evoked compound action potential (ECAP) recorded from Nucleus CI24 cochlear implant users," Ear Hear. **28**, 850–861.
Deller, J. R., Hansen, J. H. L., and Proakis, J. G. (**2000**). *Discrete-Time Processing of Speech Signals* (Institution of Electrical and Electronics Engineers Inc., New York), pp. 117–137, 336–338, 373–380.
Deng, L., and O'Shaughnessy, D. (**2003**). *Speech Processing: A Dynamic and Optimization–Oriented Approach* (Marcel Dekker Inc., New York), pp. 53–61.
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.
Dowell, R. C., Seligman, P. M., Blamey, P. J., and Clark, G. M. (**1987**). "Evaluation of a two-formant speech-processing strategy for a multichannel cochlear prosthesis," Ann. Otol. Rhinol. Laryngol. **96**, 132–143.
El-Jaroudi, A., and Makhoul, J. (**1991**). "Discrete all-pole modeling," IEEE Trans. Signal Process. **39**, 411–423.
Firszt, J. B., Holden, L. K., Reeder, R. M., and Skinner, M. W. (**2009**). "Speech recognition in cochlear implant recipients: Comparison of standard HiRes and HiRes 120 sound processing," Otol. Neurotol. **30**, 146–152.
Flanagan, J. L. (**1955**). "A difference limen for vowel formant frequency," J. Acoust. Soc. Am. **27**, 613–617.
Flanagan, J. L. (**1956a**). "Automatic extraction of formant frequencies from continuous speech," J. Acoust. Soc. Am. **28**, 110–118.
Flanagan, J. L. (**1956b**). "Band width and channel capacity necessary to transmit the formant information of speech," J. Acoust. Soc. Am. **28**, 592–597.
Fu, Q., Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustics and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.
Furuya, K., and Kataoka, A. (**2007**). "Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction," IEEE Trans. Audio Speech Lang. Process. **15**, 1579–1591.
Geurts, L., and Wouters, J. (**1999**). "Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants," J. Acoust. Soc. Am. **105**, 2476–2484.
Hawks, J. W. (**1994**). "Difference limens for formant patterns of vowel sounds," J. Acoust. Soc. Am. **95**, 1074–1084.
Hazrati, O., and Loizou, P. C. (**2012**). "Tackling the combined effects of reverberation and masking noise using ideal channel selection," J. Speech, Lang. Hear. Res. **55**, 500–510.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.
Hu, Y., and Loizou, P. C. (**2008**). "A new sound coding strategy for suppressing noise in cochlear implants," J. Acoust. Soc. Am. **124**, 498–509.
IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**(3), 227–246.
Kadir, H. (**2008**). "Speech compression using Linear Predictive Coding," https://www.mathworks.com/matlabcentral/fileexchange/13529-speech-compression-using-linear-predictive-coding (Last viewed January 29, 2007).
Kewley-Port, D., Burkle, T. Z., and Lee, J. H. (**2007**). "Contribution of consonant versus vowel information to sentence intelligibility for young normal–hearing and elderly hearing-impaired listeners," J. Acoust. Soc. Am. **122**, 2365–2375.
Koch, D. B., Osberger, M. J., Segel, P., and Kessler, D. (**2004**). "HiResolution and conventional sound processing in the HiResolution Bionic ear: Using appropriate outcome measures to assess speech recognition ability," Audiol. Neurootol. **9**, 214–223.
Kokkinakis, K., and Loizou, P. C. (**2011**). "The impact of reverberant self-masking and overlap-masking effects on speech intelligibility by cochlear implant listeners (L)," J. Acoust. Soc. Am. **130**, 1099–1102.
Loizou, P. C. (**1998**). "Mimicking the human ear," IEEE Signal Process. Mag. **15**, 101–130.
Loizou, P. C. (**1999**). "Signal-processing techniques for cochlear implants," IEEE Eng. Med. Biol. Mag. **18**, 34–46.
Loizou, P. C. (**2007**). *Speech Enhancement Theory and Practice* (Taylor & Francis, London), pp. 69–93.
Markel, J. D., and Gray, A. H. (**1976**). "Speech synthesis structures," in *Linear Prediction of Speech*, edited by J. D. Markel and A. H. Gray (Springer-Verlag, Berlin, Germany), pp. 92–128.
McDermott, H. J., McKay, C. M., and Vandali, A. E. (**1992**). "A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant," J. Acoust. Soc. Am. **91**, 3367–3371.
Neuman, A., Wroblewski, M., Hajicek, J., and Rubinstein, A. (**2010**). "Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults," Ear Hear. **31**, 336–344.
Neuman, A., Wroblewski, M., Hajicek, J., and Rubinstein, A. (**2012**). "Measuring speech recognition in children with cochlear implants in a virtual classroom," J. Speech, Lang. Hear. Res. **55**, 532–541.
Nogueira, W., Uchner, A., Lenarz, T., and Edler, B. (**2005**). "A psychoacoustic 'NofM'-type speech coding strategy for cochlear implants," EURASIP J. Appl. Signal Process. **18**, 3044–3059.
O'Shaughnessy, D. (**2008**). "Formant estimation and tracking," in *Springer Handbook of Speech Processing*, edited by J. Benesty, N. M. Sondhi, and Y. Huang (Springer-Verlag, Berlin, Germany), pp. 213–227.
Parikh, G., and Loizou, P. C. (**2005**). "The influence of noise on vowel and consonant cues," J. Acoust. Soc. Am. **118**, 3874–3888.
Patrick, J. F., and Clark, G. M. (**1991**). "The Nucleus 22-channel cochlear implant system," Ear Hear. **12**, 3S–9S.
Peterson, G., and Barney, H. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Peterson, G. E., and Lehiste, I. (**1962**). "Revised CNC lists for auditory tests," J. Speech Hear. Disord. **27**, 62–70.
Rabiner, L. R., Schafer, R. W., and Rader, C. M. (**1969**). "The chirp z-transform algorithm," IEEE Trans. Audio Electroacoust. **17**, 86–92.
Rubinstein, J. T. (**2004**). "How cochlear implants encode speech," Curr. Opin. Otolaryngol. Head Neck Surg. **12**, 444–448.
Sadjadi, S. O., and Hansen, J. H. L. (**2013**). "Unsupervised speech activity detection using voicing measures and perceptual spectral flux," IEEE Signal Process. Lett. **20**, 197–200.
Schafer, R. W., and Rabiner, L. R. (**1970**). "System for automatic formant analysis of voiced speech," J. Acoust. Soc. Am. **47**, 634–648.
Seligman, P. M., and McDermott, H. J. (**1995**). "Architecture of the Spectra 22 speech processor," Otol. Rhinol. Laryngol. **104**, 139–141.
Seligman, P. M., Patrick, J. F., Tong, Y. C., Clark, G. M., Dowell, R. C., and Crosby, P. A. (**1984**). "A signal processor for a multiple-electrode hearing prosthesis," Acta Otolaryngol. Suppl. **413**, 135–139.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Skinner, M. W., Clark, G. M., Whitford, L. a, Seligman, P. M., Staller, J. S., Shipp, D. B., Shallop, J. K., Everingham, C., Menapace, C. M., Arndt, P. L., Antogenelli, T., Brimacombe, J. A., Pijl, S., Daniels, P., George, C. R., McDermott, H. J., and Beiter, A. L. (**1994**). "Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system," Am. J. Otol. **15**, 15–27.
Skinner, M. W., Holden, L. K., Holden, T. A., Dowell, R. C., Seligman, P. M., Brimacombe, J. A., and Beiter, A. L. (**1991**). "Performance of post-

linguistically deaf adults with the wearable speech processor (WSP III) and mini speech processor (MSP) of the nucleus multi-electrode cochlear implant," Ear Hear. **12**, 3–22.
Snell, R. C., and Milinazzo, F. (**1993**). "Formant estimation from LPC analysis data," IEEE Trans. Speech Audio Process. **1**, 129–134.
Spahr, A. J., Dorman, M. F., Litvak, L. N., Van Wie, S., Gifford, R. H., Loizou, P. C., Loiselle, L. M., Oakes, T., and Cook, S. (**2012**). "Development and validation of the AzBio sentence lists," Ear Hear. **33**, 112–117.
Tao, Z., Zhao, H., Gu, J., Tan, X., and Wu, J. (**2008**). "Speech feature extraction of cochlear implants on the basis of auditory perception wavelet transform," in *Proceedings of the International Conference on Audiology, Language, and Image Processing (ICALIP'08)*, July 7–9, Shanghai, China, pp. 80–86.
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**, 1897–1899.
Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M. (**2000**). "Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system," Ear Hear. **21**, 608–624.
Wilson, B. S., and Dorman, M. F. (**2008**). "Cochlear implants: Current designs and future possibilities," J. Rehabil. Res. Dev. **45**, 695–730.
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.
Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., and Zerbi, M. (**1993**). "Design and evaluation of a continuous interleaved sampling (CIS) processing strategy for multichannel cochlear implants," J. Rehabil. Res. Dev. **30**, 110–116.
Wouters, J., McDermott, H. J., and Francart, T. (**2015**). "Sound coding in cochlear implants: From electric pulses to hearing," IEEE Signal Process. Mag. **32**, 67–80.
Wouters, J., and Vanden Berghe, J. (**2001**). "Speech recognition in noise for cochlear implantees with a two microphone monaural adaptive noise reduction system," Ear Hear. **22**, 420–430.
Zapata, J. G., Carlos, J., Martín, D., and Vilda, P. G. (**2004**). "Fast formant estimation by complex analysis of LPC coefficients," in *Proceedings from the European Signal Processing Conference (EUSIPCO 2004)*, September 6–10, Vienna, Austria, pp. 737–740.