An analysis of the effects of electrical field interaction with an acoustic model of cochlear implants

Trudie Strydom and Johan J. Hanekom[a)]
Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, Gauteng 0002, South Africa

(Received 9 December 2009; revised 31 July 2010; accepted 26 October 2010)

Electrical field interaction caused by current spread in a cochlear implant was modeled in an explicit way in an acoustic model (the SPREAD model) presented to six listeners with normal hearing. The typical processing of cochlear implants was modeled more closely than in traditional acoustic models by careful selection of parameters related to current spread or parameters that could amplify the electrical field interactions caused by current spread. These parameters were the insertion depth, electrode spacing, electrical dynamic range, and dynamic range compression function. The hypothesis was that current spread could account for the asymptote in performance in speech intelligibility experiments observed at around seven stimulation channels in a number of cochlear implant studies. Speech intelligibility for sentences, vowels, and consonants at three noise levels (SNR of +15 dB, +10 dB, and +5 dB) was measured as a function of the number of spectral channels (4, 7, and 16). The SPREAD model appears to explain the asymptote in speech intelligibility at seven channels for all noise levels for all speech material used in this study. It is shown that the compressive amplitude mapping used in cochlear implants can have a detrimental effect on the number of effective channels.
 [DOI: 10.1121/1.3518761]

PACS number(s): 43.71.Es, 43.71.Ky, 43.66.Ts [MAH] Pages: 2213–2226

I. INTRODUCTION

Acoustic models are widely used to understand and explain aspects of speech intelligibility by cochlear implant (CI) listeners (Fu *et al.*, 1998; Loizou *et al.*, 2000a; Baskent and Shannon, 2003; Baskent, 2006). Most existing acoustic models have poor quantitative correspondence with implant data in quiet and noisy listening conditions and typically predict increases in speech intelligibility for all noise types and conditions when the number of stimulation channels (stimulation electrode pairs) is increased above eight (Fu *et al.*, 1998; Friesen *et al.*, 2001; Bingabr *et al.*, 2008), whereas studies with CI listeners show saturation of speech intelligibility at about eight channels (Fishman *et al.*, 1997; Friesen *et al.*, 2001; Fu and Nogaki, 2005). There are exceptions, however. A few studies with CI users did find significant increases in speech intelligibility for some listeners as the number of channels was increased above eight, some showing improvement up to 12 channels for individual subjects (Kiefer *et al.*, 1997) and up to 16 channels using optimizing strategies for individual subjects (Frijns *et al.*, 2003; Buechner *et al.*, 2006).

Studies by Friesen *et al.* (2001) and Baskent (2006) hypothesized that channel interactions, specifically electrical field interactions, reduce the effective number of information channels to around eight for most CI listeners. Two types of channel interactions may be present in CI listeners (Shannon, 1983), namely electrical current field summation peripheral to stimulation of the nerves and neural–perceptual interaction following stimulation. The electrical field interaction component is absent in normal hearing, limiting channel interactions to those on the neural–perceptual level. In CI listeners, however, the effects of electrical field interactions may be important contributors to the observed effects of channel interactions.

The present study investigated how electrical field interactions may underlie the observed saturation of speech intelligibility that appears to occur at around eight channels.

Studies of channel interactions in acoustic models may be broadly divided into studies with spectral smearing and explicit models. In two representative simulations of spectral smearing, widened noise bands (Boothroyd *et al.*, 1996) and a smearing matrix (Baer and Moore, 1993) were used to smear the spectrum of the original speech signal. Both approaches aimed to simulate the widened auditory filters typical of CI users. Boothroyd *et al.* (1996) found that a smearing bandwidth of 250 Hz had a small but significant effect on vowel recognition, that vowels were affected more by smearing than consonants were, and that consonant place of articulation was affected more than manner of articulation or voicing cues. Baer and Moore (1993) found that spectral smearing affected speech intelligibility minimally in quiet but substantially in noise. Both of these studies used widened filters as synthesis filter[1] but did not consider filter slopes as models of current decay, as Fu and Nogaki (2005) did. The latter modeled channel interactions by using varying filter slopes in the synthesis filters (−24 to −6 dB/octave), thereby providing varying amounts of filter overlap. The varying slopes can be seen as models of current decay. Comparing their acoustic model predictions to CI listener results, they commented that, on average, CI listeners had mean speech reception thresholds (SRTs) that were close to SRTs of acoustic simulation listeners with four-channel spectrally

a)Author to whom correspondence should be addressed. Electronic mail: johan.hanekom@up.ac.za

smeared speech, although all CI listeners had more than eight stimulating channels.

The effects of dynamic range compression were ignored in the above studies but were included in a study by Bingabr *et al*. (2008), who studied the effects of monopolar and bipolar stimulation using an acoustic model. They modeled the spread of excitation for the different modes of stimulation by adjusting both the slopes and widths of the synthesis filters, assuming a current decay of 4 dB/mm for monopolar stimulation and 8 dB/mm for bipolar stimulation as measured along the basilar membrane (BM). They also modeled a current decay of 1 dB/mm. Synthesis filter width was determined by the typical width of excitation along the BM. Experiments were conducted with 4, 8, and 16 channels, using Hearing in Noise Test (HINT) sentences (Nilsson *et al*., 1994) in quiet and at 10 dB signal-to-noise ratio (SNR), as well as consonant/nucleus/consonant (CNC) words (Minimum Speech Test Battery for Adult Cochlear Implant Users, House Ear Institute and Cochlear Corporation, 1996). There was a significant increase in speech intelligibility in quiet and in noise when the current decay was increased from 1 to 4 dB/mm. In noise, however, when the current decay was increased further to 8 dB/mm, the speech intelligibility performance dropped significantly for 4 and 8 stimulation channels. The authors found significant increases in performance from 4 to 8 channels and from 8 to 16 channels, indicating that no asymptote was found. Effects of dynamic range were simulated by adjusting the filter slopes in the acoustic domain according to the ratio between the acoustic dynamic range (50 dB) and the electrical dynamic range (15 dB in their study). They also included effects of electrical dynamic range by determining widths of excitation based on the electrical dynamic range and current decay but did not consider non-linear compression.

In a study by Throckmorton and Collins (2002), channel interactions, as measured through forward masking, pitch reversals, and non-discriminable electrodes, were modeled more explicitly. They explicitly included forward-masking effects by setting signal intensity to zero within calculated time frames. They constructed three models for forward masking, named best-case, intermediate, and worst-case masking models. These models effectively used varying filter slopes of the synthesis filters combined with explicit modeling of forward-masking effects. The best-case model included masking effects of the same channel only. The intermediate model included effects of neighboring channels, with closer channels contributing more to masking effects. The worst-case masking model included effects from all channels with equal weights. Performance dropped significantly for all speech material in the intermediate case (e.g., 15% in phoneme recognition) and the worst-case masking model (e.g., 30% in phoneme recognition). Their study did not investigate the effects of the number of channels.

Apart from those discussed above, two other aspects need to be included when modeling the influence of current decay in an acoustic model. First, since current decays spatially away from the electrode, it is important to include the correct spacing between electrodes in the model. This was recognized by Baskent and Shannon (2003, 2007) in their acoustic models of compression effects.

Second, because of current spread, dynamic range compression will influence the effective current delivered at targeted stimulation sites. This is because linear and non-linear dynamic range compressions decrease or distort the difference in intensity levels between channels in the electrical domain, respectively, where electrical field interactions occur. It is known that dynamic range compression has an influence on speech perception. Fu and Shannon (1998) studied effects of compression in normal-hearing and CI listeners using four electrodes and found optimal performance when normal loudness was preserved. Similarly, Loizou *et al*. (2000a) considered the effects of linear dynamic range compression in an acoustic model and found that all speech material was affected by dynamic range compression, with vowels affected most and consonant place of articulation also affected significantly. These findings were ascribed to reduced spectral contrast.

In the work reported here, the hypothesis that the asymptote in speech intelligibility is caused by electrical field interactions was investigated with an acoustic model using more noise levels and a wider range of speech materials than in previously reported studies. In addition, the approach to modeling electrical field interaction was more explicit than that of previous studies (Baer and Moore, 1993; Throckmorton and Collins, 2002; Fu and Nogaki, 2005; Bingabr *et al*., 2008). In the study by Bingabr *et al*., for example, current decay effects were modeled using appropriate filter parameters, but effects of the compression function and electrode spacing were ignored, which could have obscured some of the effects of current decay. The present model included realistic values for electrode spacing, reduced input and electrical dynamic ranges, and logarithmic compression to give a truer reflection of electrical field interaction effects in implant listeners, as these parameters all have an impact on the effective current delivered to a target neural population.

II. METHODS

A. Acoustic models

Two model variations were developed, the first one similar to that used in the study by Friesen *et al*., with the same filter cutoffs and envelope extraction mechanisms (Friesen *et al*., 2001). This model is referred to as the STANDARD model. To provide closer mimicking of actual implants, electrical field interaction was explicitly modeled in the second model (referred to as the SPREAD model), while the effects of compression of a limited input dynamic range into a limited electrical dynamic range using a suitable loudness growth function and limited insertion depth were carefully modeled. More detail is provided in Sec. III.

1. A consideration of models of current spread to be used in the SPREAD model

Different sources may be used to determine the extent of current spread, including psychophysics experiments with forward masking (Kwon and van den Honert, 2006), single nerve recordings (Kral *et al*., 1998), finite-element models (Hanekom, 2001), or ringer-bath experiments (Kral *et al*., 1998). Predictions of current spread from forward-masking

values are more suitable to models of forward masking, whereas single-nerve recordings are obtained from animal subjects, which may limit their suitability for modeling current spread in human subjects. The extent of current spread from neighboring electrodes is determined by the electrode configuration, the spreading constant of the medium, the distance from the stimulating electrode, and the geometry of the medium and the electrodes (Frijns *et al.*, 1995; Hanekom, 2001). Monopolar configurations typically have larger spread of excitation than bipolar configurations (Kral *et al.*, 1998; Hanekom, 2001). The spreading constant of the medium in CIs is determined by various components, including spreading constants of the perilymph, endolymph, spiral ganglion, and BM. All of these are typically included in the available finite-element models (Frijns *et al.*, 1995; Hanekom, 2001, 2005). The distance between the delivering electrode and the point of neural activation is important, with the geometry of the cochlea also playing a role. As an example, the spread of current is more in the basal turns of the cochlea, presumably due to the wider cochlear duct (Kral *et al.*, 1998) and/or the spiral shape of the cochlea, with the spiral radius larger in the basal region than in the apical region (Hanekom, 2001). The present SPREAD model therefore mostly used tuning curves from the finite-element model of Hanekom (2001) and the ringer-bath experiments of Kral *et al.* (1998). All of the above-mentioned aspects were included in the finite-element model of Hanekom, which showed average values of current decay as a function of distance (millimeter along BM) from the delivering electrode, which can be used in a model of current spread. The last two approaches typically found current decay of 7.5–10 dB/mm for bipolar stimulation.

2. Assumptions for the acoustic models

The primary assumption for the SPREAD model was the way in which electrical field interaction is modeled. Current spread from neighboring stimulation channels affects the effective current that is delivered at a target nerve fiber population and therefore distorts the temporal envelopes of the stimulation signals that are conveyed to the population.

The electrical currents from different electrodes were assumed to be in phase in the SPREAD model, which meant that current spread from different electrodes could simply be added to find accumulated current values at target nerve population sites. The present model is therefore a model of simultaneous analogue stimulation (SAS) processing (Mishra, 2000), which is a simultaneous stimulation strategy, where electrical field interaction caused by current spread is believed to be most detrimental to speech intelligibility. The majority of existing acoustic models (e.g., Friesen *et al.*, 2001; Fu and Nogaki, 2005; Bingabr *et al.*, 2008) implicitly assume simultaneous stimulation, since no modeling of timing effects related to interleaved stimulation of electrodes is included. Although the models implicitly assume simultaneous stimulation (typical of SAS), they extract envelopes as done in continuous interleaved sampling (CIS) processing (Loizou, 2006). The present SPREAD model used the same approach. As SAS processing uses bipolar stimulation (Mishra, 2000), a bipolar stimulation mode was assumed in the present model. Unimodal stimulation patterns were assumed, with maximum stimulation opposite the active stimulating electrode. It may be noted that Friesen *et al.* (2001) found no significant difference between results obtained with CI listeners using SAS, CIS, and SPEAK processing schemes.

The SPREAD model assumed an input dynamic range limited to 60 dB (Mishra, 2000) that is logarithmically compressed into an electrical output dynamic range of 11 dB, the latter being an average value found for electrical dynamic range from a number of studies (e.g., Kreft *et al.*, 2004).

The inclusion of realistic electrode spacing presented a potential problem in terms of matching the analysis range to the range covered by the electrodes, since the typical range which is covered by the analysis filters in the 4- and 7-electrode simulation (to be expanded on later) is 250–6800 Hz, which is the Clarion analysis filter range (Mishra, 2000), whereas the range covered by an array of 16 mm is typically 185–2476 Hz if an insertion depth of 30 mm is assumed (Greenwood, 1990). An insertion depth of 25 mm was therefore assumed in the model to ensure that the modeled electrode positions, covering a range of 25 mm (512 Hz) to 10 mm (5084 Hz), would be more closely centered on the analysis range. An insertion depth of 25 mm has been shown to give optimal speech intelligibility (Baskent and Shannon, 2005) and would also be a realistic model of actual implant depths.

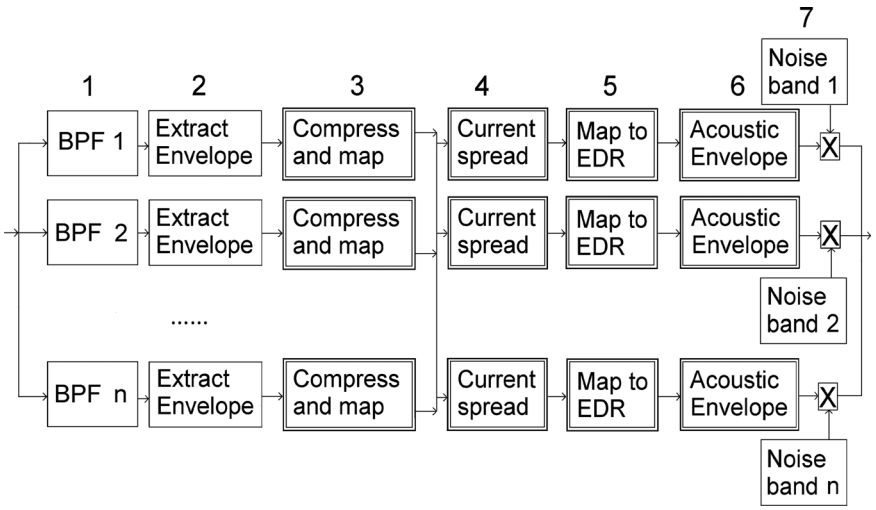

FIG. 1. Signal processing steps for the SPREAD and STANDARD model. Blocks with double lines are the additional steps for the SPREAD model. The acoustic envelope block is necessary to convert electrical current values from the previous step into acoustic intensity. EDR is the electrical dynamic range, which is assumed to be 11 dB in this study. The numbers in the figure are used to describe signal processing steps in the text. Noise bands are already bandpass filtered, using filters as shown in Table I.

A symmetrical current decay of 7 dB/mm was assumed for 4, 7, and 16 electrodes, even though current spread resulting from a bipolar pair of electrodes separated by 4 mm would be much larger than for a bipolar pair separated by 1 mm (Hanekom, 2001).

Noise bands were assumed to model the sound perceived by CI listeners. These appear to better approximate the sounds perceived by CI listeners than pure tones (Blamey *et al.*, 1984; Laneau *et al.*, 2006), although the study by Dorman *et al.* (1997) investigated the use of pure tones as synthesis signals, based on CI listeners reporting beep-like sounds from electrical stimulation. The latter study showed no significant differences in speech intelligibility for most speech material using pure tones or noise bands.

3. Signal processing for acoustic models

Figure 1 illustrates the signal processing steps for both models. The different stages of signal processing shown here will be explained below. Examples of outputs from the signal processing steps are shown in Fig. 2.

a. Steps 1 and 2: Filtering and envelope extraction. Speech material was processed using noise-band vocoder processing (Shannon *et al.*, 1995), which was augmented to include current spread in the cochlea as discussed in Sec. II 3 c). Speech-shaped noise was added to each speech token at the required SNR to allow comparison with the data by Friesen *et al.* (2001). All processing steps for filtering and envelope extraction were the same as for the acoustic model in the study by Friesen *et al.* (2001). The speech material was sampled at 44 100 Hz and filtered into a specified number of contiguous frequency channels using sixth-order Butterworth bandpass filters (the analysis filters). For 16 channels, the center frequencies were logarithmically spaced between 100 and 6000 Hz with the pass band of the first filter at 100 Hz and the stop band of the last filter at 6000 Hz. For 4 and 7 electrodes, the filter cutoffs were chosen according to the values used in the Clarion implant (Mishra, 2000). Table I shows the filter −3 dB cutoff frequencies for all filters. The filters overlapped at these frequencies. Envelopes of the filter outputs were extracted by half-wave rectification and low-pass filtering using third-order Butterworth filters with a cutoff frequency of 160 Hz for both models. The envelopes extracted at this stage are called acoustic temporal envelopes [shown in Figs. 2(a) and 2(f)], since they have not been mapped to electrical units yet.

b. Step 3: Compression. This step was included only in the SPREAD model to facilitate calculations with typical current levels as found in CIs. As such it may be seen as one of the steps used to model the electrical interface. The six highest maximum envelope values from the set of channel envelopes were determined for each speech token (sentence, vowel, or consonant). The average of these six maximum values was used as the saturation level for the input signal. A base level was selected at 60 dB down from this level to give a 60 dB input dynamic range. The envelopes were now fitted into this restricted range by setting values above saturation level to the value at saturation level (clipping) and by setting values below base level to zero. A logarithmic loudness growth function, as used in the Clarion implant (Mishra, 2000), was applied to this 60 dB range envelope to map this to an electrical dynamic range of 11 dB using assumed

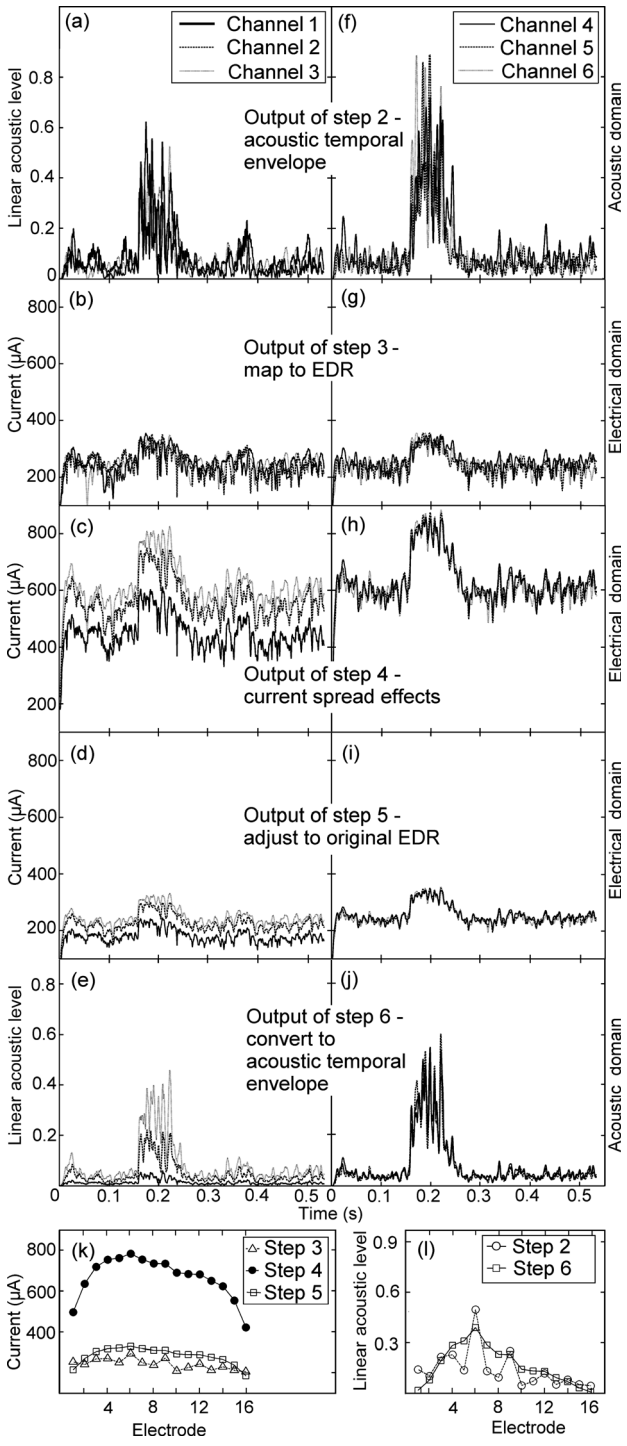

FIG. 2. Original envelope and processed envelope for the SPREAD model, 16-channel simulation, for the vowel /ɑ/ for channels 1, 2, and 3 (left panel) and channels 4, 5, and 6 (right panel). Note the different scales for the abscissa used for the different panels; (a)–(e) indicates the outputs for signal processing steps 2–6 in Fig. 1, respectively, and (f)–(j) are the corresponding signal processing outputs for channels 4–6. The panels for (b)–(e) and (g)–(i) indicate signal levels in microampere, whereas the panels (a), (f), (e), and (j) indicate linear acoustic level (normalized voltage units). (k) Outputs of steps 3, 4, and 5 for the SPREAD model at time 0.17 s. (l) Initial (step 2) and final spatial signal level profile (step 6) for SPREAD model at time 0.17 s.

TABLE I. Analysis and synthesis filter cutoff frequencies (−3 dB) for the different conditions.

Channels	Analysis and synthesis filters for STANDARD model	Synthesis filters for SPREAD model
4	250, 875, 1450, 2600, and 6800 Hz	334, 703, 1343, 2456, and 4390 Hz
7	250, 500, 875, 1150, 1450, 2000, 2600, and 6800 Hz	397, 606, 892, 1285, 1823, 2562, 3574, and 4963 Hz
16	100, 158, 228, 313, 417, 544, 698, 886, 1114, 1392, 1730, 2142, 2643, 3253, 3996, 4900, and 6000 Hz	449, 540, 645, 765, 903, 1061, 1242, 1451, 1690, 1965, 2281, 2644, 3060, 3537, 4086, 4716, and 5439 Hz

thresholds and comfort levels of implants of 100 μA (T-level) and 355 μA (C-level) respectively. Output for this step is shown in Figs. 2(b) and 2(g). Note that an inverse transformation (step 6 in Fig. 1) translates current values back to acoustic intensity envelope values.

c. Step 4: Current spread and electrical field interaction. This step still focuses on the electrical interface. Electrical currents, as determined from the previous step, contribute to current delivered at the target nerve populations of neighboring electrodes, thereby increasing the effective current delivered at all sites in the cochlea. The effective current at site i in the cochlea for a current decay of 7 dB/mm is given by

$$\text{Spread}(j,i) = 10^{-7nd/20} I(j), \tag{1}$$

$$I_{\text{eff}}(i) = \sum_{j=1}^{N} \text{Spread}(j,i), \tag{2}$$

where $I_{\text{eff}}(i)$ is the effective current at site i, N the number of electrodes, and Spread(j,i) is the magnitude of the spread of current from electrode j at site i. Spread(j,i) is the current delivered at site j by the electrode closest to site i. $I(j)$ is the current delivered at electrode j, d is the distance between two adjacent electrodes in millimeter for the specific model (e.g., 1 mm for the 16-channel model), and n is the number of electrode spaces between sites i and j. For example, if $i = j$, $n = 0$ and if i and j are two adjacent sites, $n = 1$. This approach assumes that the number of information channels is the same as the number of electrodes. Note that these effective current levels are found in microamperes in the model. The typical output of this signal processing step is shown in Figs. 2(c) and 2(h).

d. Steps 5 and 6: Interpreting the effective current effects. An acoustic temporal envelope was mapped to electrical current levels in step 3 (Fig. 1). In step 6, electrical current levels are converted back to linear acoustic output levels. The calculations for this need to be the inverse of the calculations in step 3. However, the effective current levels may now exceed the electrical comfort levels owing to electrical field interaction. To model the effect that this would have in an actual implant, the maximum of these current levels from all channels was taken as the new electrical perceptual comfort level. The new electrical threshold level was calculated at 11 dB down from the comfort level. The current values in all channels were then decreased using a linear mapping function to fit the effective electrical stimulation currents into the original electrical dynamic range [step 5 in Figs. 1 and 2(d) and 2(i)]. The inverse of the loudness growth function was applied to predict the normal-hearing loudness percepts that would be associated with these current levels (step 6). The output from this step was an acoustic temporal envelope (linear level units) [Figs. 2(e) and 2(j)].

e. Step 7: Synthesis signals. For both model variations, the synthesis signals were noise bands that were generated from white noise that was bandpass filtered using sixth-order Butterworth bandpass filters. For the STANDARD model, the noise bands had the same cutoff frequencies as those used in step 1. In the SPREAD model, which had a modeled insertion depth of 25 mm, the cutoff frequencies were calculated according to simulated electrode position, using Greenwood's equation (1990) and assuming an insertion depth of 25 mm, with electrodes spaced 1, 2.3, and 4 mm apart for the 16, 7, and 4 electrode conditions, respectively. The positions of the electrodes were assumed to determine the center frequencies of the filters, and the −3 dB cutoff frequencies were chosen to correspond to positions halfway between the electrode positions. This corresponds to the approach of other acoustic models (e.g., Shannon *et al.*, 1995; Baskent and Shannon, 2003). It should be noted that noise bands may implicitly represent some spread in current, as exemplified by the approach of Bingabr *et al.* (2008). The present SPREAD model therefore included both an explicit modeling of electrical field interaction as well as this unintended additional current spread. The choice of noise bands as synthesis signals thus introduced a potential error in the modeled effective current delivered at a specific site. An estimation of the magnitude of this error is made in the Results section and is illustrated in Fig. 9(a). The net effect is that the effective current decay changes to around 6 dB/mm for 16 channels, as opposed to the explicitly modeled 7 dB/mm.

f. Modulation of synthesis signals by envelope outputs. The envelope outputs from step 4 were used to modulate the synthesis signals obtained in step 5. An equalizing step ensured that the root-mean-square (rms) energy in each of the final modulated signals remained the same as the rms energy in the corresponding processed acoustic envelope from step 4 in Fig. 1. These modulated signals were added to arrive at the final output signal.

B. Experimental methods

1. Listeners

Six Afrikaans-speaking listeners, aged between 18 and 35 yr, participated in the study. All had normal hearing as determined by a hearing screening test, with all subjects

having thresholds better than 20 dB at frequencies ranging from 250 to 8000 Hz.

2. Speech material

Sentences, spoken by a female voice, were used in sentence recognition tests (Theunissen *et al.*, 2008). The sentences were of easy to moderate difficulty and had an average length of six words. The sentences were normed for equal difficulty and were grouped into lists of ten sentences each. List slopes covered a range of 2.37%/dB, with an average slope per list of 16.02%/dB and a standard deviation of 0.64%/dB across lists. This means that, when presented to listeners with normal hearing, word recognition improved by 16.02% with each decibel of increase in the SNR.

Fourteen medial consonants (b d g p t k m n f s ʃ v z j), spoken by a male and female voice (Pretorius *et al.*, 2006), were presented in the context a/consonant/a. Twelve medial vowels (ɑ ɑ: œ æ ɛ ɛ: u i y ə ɔ e:) spoken by a female and male voice (Pretorius *et al.*, 2006), in the context p/vowel/t, were presented to the same listeners.

3. Experiments

Two sets of experiments were conducted, one set for each model. Ceiling effects could obscure asymptote effects in quiet, so experiments were conducted in noise at +15, +10, and +5 dB SNR with 4, 7, and 16 channels, for a total of nine conditions for each set.

4. Procedure

Experiments were conducted in a double-walled sound booth. Processed speech material was presented in the free field using a Yamaha MS101 II loudspeaker (Yamaha Electronics Corp., Buena Park, CA). Listeners could adjust the volume to comfortable levels. These levels were found to be between 60 and 70 dB sound pressure level (SPL). They were seated 1 m from the loudspeaker, which was at ear level, facing it.

Sentences were presented in an order designed to produce maximal learning effects, with the easiest material first. Each condition consisted of ten sentences. Subjects had practiced with processed speech for at least 2 h before commencing with the sentence recognition experiments. A short additional practice session of ten sentences (which could be repeated), for a specific processing scheme, was also allowed before the commencement of each experiment. New sentences that had not been used in practice sessions were played back once when gathering experimental data. Subjects were encouraged to report any parts of sentences, even if it did not make sense. Subjects reported verbally what they had heard. Each correct word was scored.

Consonants and vowels were presented to listeners in random order using customized software (Geurts and Wouters, 2000) without any practice session. Twelve repetitions of each vowel or consonant (six male and six female) were presented. The software presented processed consonant or vowel material, and the listener had to select the correct consonant or vowel by clicking on the appropriate button on the screen. Vowel and consonant confusion matrices were constructed automatically by the software. The material was presented one condition at a time with the easiest material first to allow listeners maximum opportunity for adapting. Chance performance level for the vowel test was 8.3%, and the 95% confidence level was at 12.48% correct. Chance performance level for the consonant test was 7.14%, with the 95% confidence level at 11.1% correct. No feedback was given. Listeners tired easily, so rest periods of 5–10 min were allowed after three to four conditions. Experiments were conducted over several days for each subject. Scores for vowels and consonant were corrected for chance [similar to the Friesen *et al.* study (2001)] by using Eq. (3).

TABLE II. Categories used for feature analysis.

Consonants	p	t	k	b	d	m	n	s	ʃ	f	v	j	z	g
Voicing	0	0	0	1	1	1	1	0	0	0	1	1	1	1
Manner	1	1	1	2	2	3	3	4	4	4	5	5	4	2
Place	1	2	3	1	2	1	2	2	2	1	1	2	2	3

Vowels[a]	ɑ:	ɑ	æ	e:	ɛ	ɛ:	i	ə	œ	ɔ	u	y
F1	3	3	3	1	2	2	1	2	2	3	1	1
F2	2	2	2	3	3	3	3	2	2	1	1	3
Duration	2	2	2	2	1	2	1	1	1	1	1	1

[a]Classification of the vowel features duration, F1 and F2. For duration, category 1: <200 ms; category 2: >200 ms. For F1, category 1: <375 Hz; category 2: 375–500 Hz; category 3: >500 Hz. For F2, category 1: <1125 Hz; category 2: 1125–1875 Hz; category 3: >1875 Hz.

$$\text{Score}_{\text{corrected}} = 100\left(\frac{\text{Score} - \text{chance\_performance}}{100 - \text{chance\_performance}}\right). \quad (3)$$

Analysis of the confusion matrices for consonants using voicing, manner of articulation, and place of articulation features was done according to the method described in Miller and Nicely (1955). The categories for voicing, manner of articulation, and place of articulation are shown in Table II. Analysis of the confusion matrices for vowels was done assuming as cues formants F1, F2, and duration, as described by van Wieringen and Wouters (1999). In order to perform a feature information transmission analysis (FITA), the first formants (F1) and second formants (F2) were categorized as shown in Table II. Categories were chosen to correspond to filter cutoff frequencies used for 16 channels and to ensure that the F2s of the male and female utterances would belong to the same category. Categories for duration are the same as in the van Wieringen and Wouters study.

III. RESULTS

Results are shown in Figs. 3–9. Where the acoustic model results are compared to CI data (Figs. 3–6), the latter was always for bipolar stimulation. In each case, a two-way repeated measures analysis of variance (ANOVA) was used to determine if there were significant effects of number of electrodes or noise level. *Post-hoc* two-tailed paired *t*-tests were performed if significant effects were found in the ANOVA. The results of these *t*-tests are indicated on the graphs. Significant differences for each model are indicated

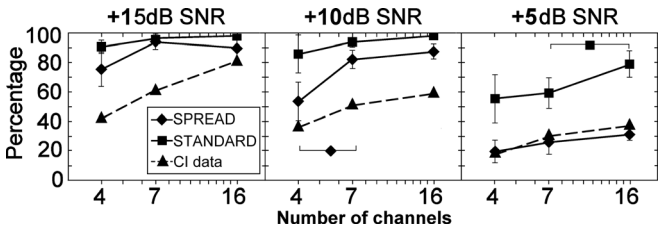

FIG. 3. Sentence intelligibility at three SNRs for 4, 7, and 16 channels. The CI data are from the study by Friesen *et al*. (2001). Error bars show ±1 standard deviation (SD). Significant differences between scores at 4 and 7 and between scores at 7 and 16 are indicated by the same symbols as the graph. The symbol ⊢◆⊣, for example, indicates significant difference between scores at the Holm–Bonferroni corrected 0.05 level for the SPREAD model.

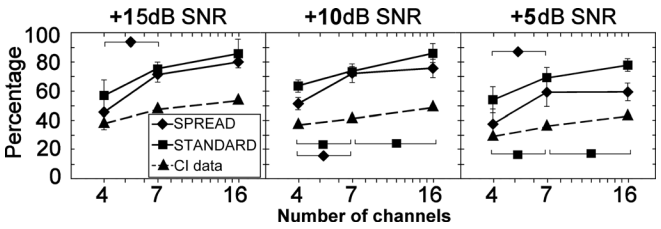

FIG. 4. Consonant intelligibility at 3 SNRs for 4, 7, and 16 channels, corrected for chance. The CI data are from the study by Friesen *et al*. (2001). Error bars indicate ±1 SD. Significant differences (using Holm–Bonferroni correction) are indicated by the same symbols as those used for the graph. Error bars for values from the present study indicate ±1 SD.

by the same character as the symbol used for the graph. Using Holm–Bonferroni correction (Holm, 1979), one symbol indicates significant difference at the corrected 0.05 level (which is typically corrected to between 0.05 and 0.0083 to maintain the family-wise type I error-level at the 0.05 level). Two symbols indicate significant differences at the corrected 0.001 level. For example, the symbol ⊢◆◆⊣ indicates a significant difference (at the corrected 0.001 level) in scores for the SPREAD model. In Figs. 3–7 significant differences are determined using the corrected 0.05 and 0.001 levels.

A. Sentence intelligibility

Figure 3 shows the results of the sentence intelligibility scores for both models, as well as one set of data from the study by Friesen *et al*. (2001). Clarion implant results are not reported for 16 electrodes in the study by Friesen *et al*. (2001), so results from the Nucleus implant are used as a substitute, since there were non-significant differences between results for CIS, SPEAK, and SAS stimulation in the study by Friesen *et al*. (2001). The figure indicates that the SPREAD model gives consistently lower values than the STANDARD model, except at the highest SNR of +15 dB. Sentence intelligibility appears to asymptote at seven channels for the SPREAD model at all noise levels. The asymptote could have been obscured by ceiling effects in the STANDARD model at +15 dB SNR, but ceiling effects appeared to be absent at +10 and +5 dB SNR. A statistical analysis was performed to test these observations.

For the STANDARD model, a two-way repeated measures ANOVA indicated a significant main effect of noise level [$F(2,45) = 20.5$, $p < 0.001$], a significant effect of number of electrodes [$F(2,45) = 18.6$, $p < 0.001$], and no significant interaction [$F(4,45) = 2.35$, $p = 0.07$]. In the SPREAD model, a two-way repeated measures ANOVA indicated a significant main effect of number of electrodes [$F(2,45) = 33.9$, $p < 0.001$] and noise level [$F(2,45) = 297.2$, $p < 0.001$] in the SPREAD model. There was significant interaction between noise and number of channels [$F(4,45) = 4.82$, $p < 0.05$]. Significant differences between scores are indicated on Fig. 3, using the symbols as discussed. Figure 3 shows that sentence intelligibility in both the SPREAD and STANDARD model asymptotes at seven channels for all noise levels. A one-way ANOVA, pooling data for all noise levels and for all numbers of electrodes, comparing results for the SPREAD and STANDARD models, showed a significant main effect of model [$F(1,107) = 12.7$, $p < 0.001$].

B. Consonant intelligibility

Results for consonant intelligibility are displayed in Fig. 4, together with one set of CI data (Friesen *et al*., 2001). Consonant recognition appears to display an asymptote at seven channels for all noise levels in the SPREAD model. The results for the SPREAD model are generally lower than those for the STANDARD model. The consonant intelligibility scores also do not appear to decline as steeply as the sentence intelligibility scores from +10 to +5 dB SNR. Statistical analysis was performed on the consonant intelligibility scores using a two-way repeated measures ANOVA, followed by post-hoc paired *t*-tests where significant effects were found. Significant differences between scores (Holm-Bonferroni corrected) are indicated in Fig. 4, using the symbols as discussed for sentences.

For the STANDARD model a two-way repeated measures ANOVA indicated a significant main effect of noise level [$F(2,45) = 17.86$, $p < 0.001$], significant main effect of number of electrodes [$F(2,45) = 69.31$, $p < 0.001$], and no significant interaction [$F(4,45) = 1.74$, $p = 0.16$]. For the SPREAD model, a two-way repeated measures ANOVA indicated a significant main effect of noise level [$F(2,45) = 17.86$, $p < 0.001$], significant main effect of number of electrodes [$F(2,45) = 69.31$, $p < 0.001$], and a non-significant interaction [$F(4,45) = 1.74$, $p = 0.16$]. A one-way ANOVA, pooling data for all noise levels and for all numbers of electrodes, comparing results for the SPREAD and STANDARD models, showed a significant main effect of model [$F(1,107) = 13.1$, $p < 0.001$].

The consonant feature percentage scores for voicing, manner, and place of articulation for both models are displayed in Fig. 5. Scores from implant listeners from the study by Friesen *et al*. are displayed for comparison.

The different feature scores for the two models were compared to determine if there were significant differences in scores and to determine if the trend of an asymptote at seven channels was also observed in the different features of consonants. Repeated measures ANOVAs were performed for each feature to determine if there were effects of number of channels and noise level. These ANOVAs for the STANDARD model indicated significant effects of number of channels

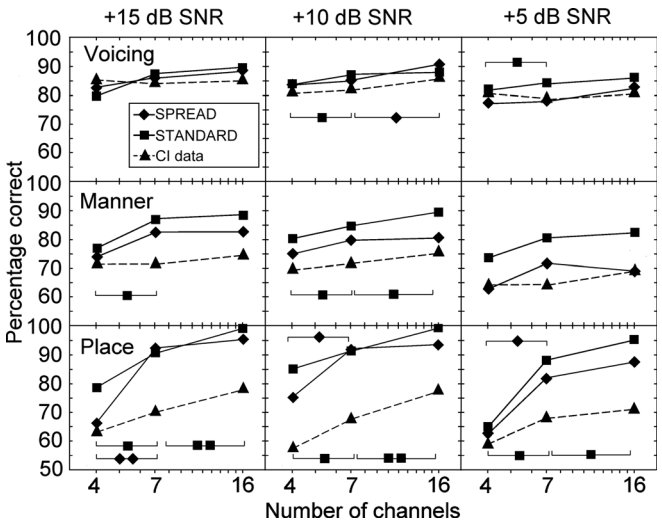

FIG. 5. Percentage correct for the features voicing, manner, and place of articulation for consonants. The CI data are from the study by Friesen *et al.* (2001). Significant differences (using Holm–Bonferroni correction) are indicated using the same symbols as for the model, as discussed in text.

[voicing: $F(2,45) = 7.33$, $p < 0.005$, manner: $F(2,45) = 13.35$, $p < 0.001$, place: $F(2,45) = 107.74$, $p < 0.001$] and noise level [manner: $F(2,45) = 4.65$, $p < 0.05$, place: $F(2,45) = 16.84$, $p < 0.001$], but no significant main effect of noise level for voicing [$F(2,45) = 0.69$, $p = 0.50$]. The ANOVAs for the SPREAD model indicated significant effects of number of channels [voicing: $F(2,45) = 6.85$, $p < 0.01$, manner: $F(2,45) = 11.45$, $p < 0.001$, place: $F(2,45) = 86.29$, $p < 0.001$] and noise level [voicing: $F(2,45) = 9.25$, $p < 0.001$, manner: $F(2,45) = 18.26$, $p < 0.001$, place: $F(2,45) = 8.23$, $p < 0.001$] for all features. The results in Fig. 5 indicate that all features asymptote at seven channels at all noise levels for the SPREAD model, except voicing at +10 dB SNR.

One-way ANOVAs were performed, pooling data for all noise levels and all numbers of channels, for each of the consonant features. There was no significant effect of model for voicing [$F(1,107) = 1.7$, $p = 0.19$], a significant main effect of model for manner [$F(1,107) = 19$, $p < 0.001$], and a significant main effect of model for place [$F(1,107) = 4.5$, $p < 0.05$].

In summary, consonant intelligibility also showed an asymptote at seven channels.

C. Vowel intelligibility

Results for vowel intelligibility are displayed in Fig. 6, together with one set of CI data (Friesen *et al.*, 2001). Vowel intelligibility displays an asymptote at seven channels (SPREAD model) for all noise levels, appearing to give slightly lower scores at 16 channels. The results for the SPREAD model are noticeably lower than those for the STANDARD model. The vowel intelligibility scores do not appear to decrease either as the SNR becomes poorer for the SPREAD model. Statistical analysis was performed on the vowel intelligibility scores using atwo-way repeated measures ANOVA, followed by paired *t*-tests where applicable. Similar to the consonant intelligibility scores, an analysis,

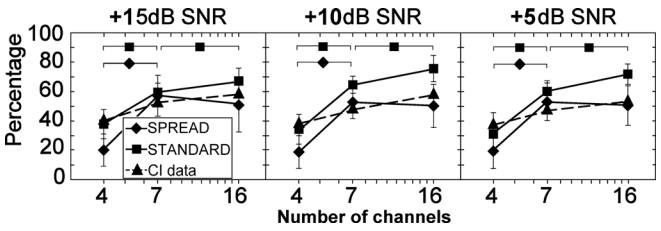

FIG. 6. Vowel intelligibility scores at three noise levels for 4, 7, and 16 channels, corrected for chance. The CI data are from the study by Friesen *et al.* (2001). Error bars indicate ±1 SD. Significant differences (using Holm–Bonferroni correction) are indicated using the same symbols as for the model, as discussed in text.

using post-hoc paired *t*-tests was also performed to determine if the results for the different models differed at 4, 7, and 16 channels. Significant differences between scores (Holm–Bonferroni corrected) are indicated in Fig. 6, using the symbols as discussed for sentence intelligibility.

For the STANDARD model a two-way repeated measures ANOVA indicated no significant main effect of noise level [$F(2,45) = 1.26$, $p = 0.29$], significant main effect of number of electrodes [$F(2,45) = 80.91$, $p < 0.001$], and no significant interaction [$F(4,45) = 0.99$, $p = 0.42$]. For the SPREAD model, a two-way repeated measures ANOVA indicated no significant main effect of noise level [$F(2,45) = 0.12$, $p = 0.88$], a significant main effect of number of electrodes [$F(2,45) = 36.97$, $p < 0.001$], and non-significant interaction [$F(4,45) = 0.05$, $p = 1.00$]. A one-way ANOVA, pooling data for all noise levels and for all numbers of electrodes, comparing results for the SPREAD and STANDARD models, showed a significant main effect of model [$F(1,107) = 15.6$, $p < 0.001$].

Results from all noise levels were pooled in the SPREAD and STANDARD model, since there was no statistically significant difference between scores at the different noise levels. The vowels with the lowest intelligibility scores were plʏlt, plʉlt, and pləlt for the SPREAD model for all numbers of electrodes. The vowel intelligibility for plɪlt (16 channels), plɛlt (seven channels), plɑlt, and plælt (four channels) was also very low. The vowel features F1, F2, and duration were analyzed. Results are displayed in Fig. 7. Single-factor ANOVAs were performed for each feature, after combining results from all noise levels. The ANOVAs for the STANDARD model indicated significant main effects of

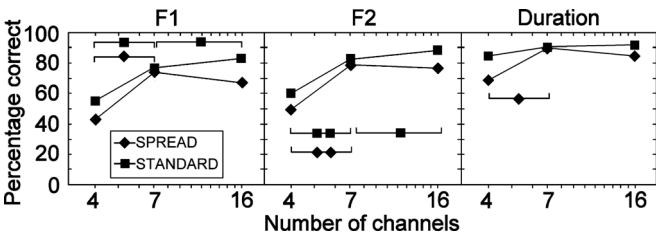

FIG. 7. Vowel feature percentages correct summarized over three noise levels. The study by Friesen *et al.* (2001) did not include a vowel FITA. Error bars indicate ±1 SD. Significant differences (using Holm–Bonferroni correction) are indicated using the same symbols as for the model, as discussed in text.

channel for F1 [$F(2,15) = 54.32$, $p < 0.001$] and F2 [$F(2,15) = 87.22$, $p < 0.001$] but not for duration [$F(2,15) = 3.62$, $p = 0.052$]. The ANOVAs for the SPREAD model indicated significant main effects of channel for F1 [$F(2,15) = 5.22$, $p < 0.05$] and F2 [$F(2,15) = 13.75$, $p < 0.001$] but not for duration [$F(2,15) = 2.35$, $p = 0.13$]. Paired t-tests were performed for the F1 and F2 cues to determine if there were significant differences between scores at 4 and 7 channels and between scores at 7 and 16 channels. Differences are indicated in the same way as with consonant features. The percentage correct for F1, F2, and duration cues for the models is displayed in Fig. 7. Figure 7 indicates that the SPREAD model displays asymptote at seven channels for F1, F2, and duration transmission. The STANDARD model does not display an asymptote but shows increases from 7 to 16 channels for F1 and F2 transmission, as well as for vowel recognition (Fig. 6).

One-way ANOVAs were performed, pooling data for all noise levels and all numbers of channels, for each of the vowel features. There was a significant main effect of model for F1 [$F(1,107) = 7.0$, $p < 0.01$], for F2 [$F(1,107) = 7.1$, $p < 0.05$], and for duration [$F(1,107) = 16.1$, $p < 0.001$].

D. Effect of modeled current decay

In an attempt to explain findings, the effects of electrical field interaction on the speech signal were investigated by considering typical outputs (Fig. 2) of the signal processing steps described in Fig. 1, considering power spectral densities (PSDs) of some of the vowels (Fig. 8) and studying the spatial signal level profile (after current spread from other electrodes had been added) [Fig. 9(a)]. Figure 9(a) also shows a comparison of the effects of different modeled values of current decay for a typical vowel.

Figure 2 shows that the signal temporal envelope is modified by current spread, by comparing Figs. 2(a)–2(e) and 2(f)–2(j). The changes are different for the low-frequency channels (channels 1, 2, and 3) from those for the mid-frequency channels (channels 4, 5, and 6). In this specific example, the intensities of channels 1 and 2 are reduced relative to channels 4, 5, and 6 in the SPREAD model. The intensity of channel 1 is reduced with respect to channels 2 and 3. Channels 4, 5, and 6 are also modified by current spread, but these changes appear less severe than those of the lower-frequency channels. Figures 2(b) and 2(g) indicate that the

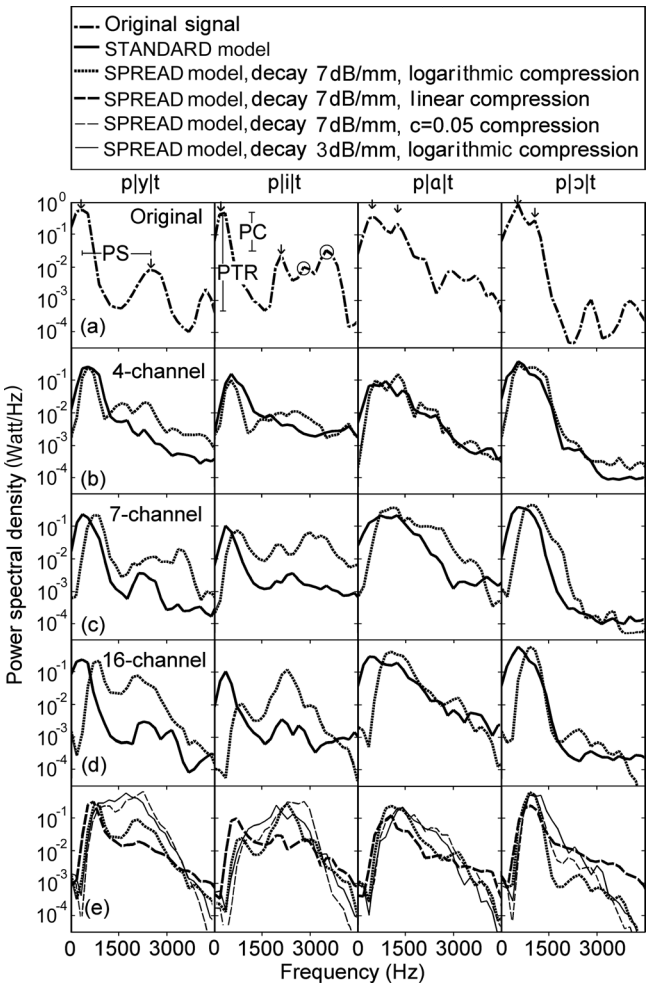

FIG. 8. PSDs for the vowels p|y|t, p|i|t, p|ɑ|t, and p|ɔ|t. Some traces are slightly displaced on the vertical axis for clarity. Arrows indicate approximate positions of the first two formants. (a) PSD of the unprocessed signal; (b) 4-channel simulation, STANDARD and SPREAD model; (c) 7-channel simulation, STANDARD and SPREAD model; (d) 16-channel simulation, STANDARD and SPREAD model; (e) PSDs of signal using 16-channel simulation with the SPREAD model with different compression functions and current decay: $c = 0.05$ compression (7 dB/mm current decay), linear compression (7 dB/mm current decay), and logarithmic compression (3 dB/mm current decay). The SPREAD model trace (16-channel, 7 dB/mm current decay, and logarithmic compression) is repeated in this panel to facilitate comparison with the other traces.

electrical field interaction could be influenced by the compression function, which reduces contrast between the signals. Figure 2(k), which is a snapshot in time of the spatial intensity profile over all the channels, shows that the compression

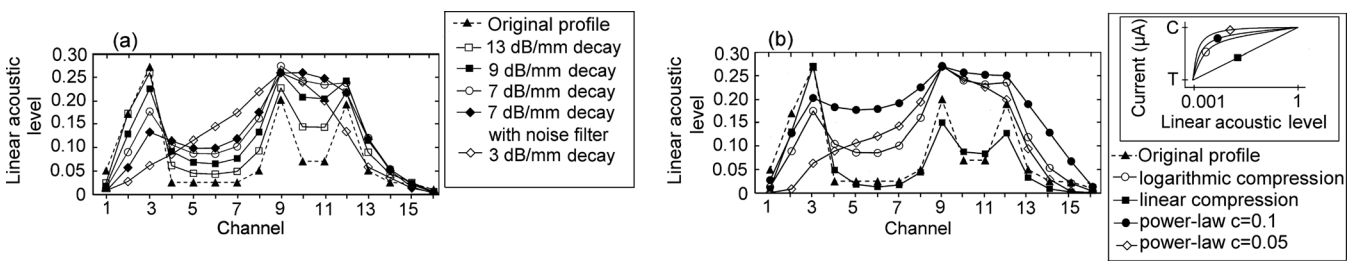

FIG. 9. Original spatial signal level profile (before processing) plotted along with the effective output signal level profiles (after processing with the SPREAD model) for a number of (a) values of current decay and (b) compression functions (with fixed current decay of 7 dB/mm). These represent a given time instant for the vowel p|i|t for the 16-channel SPREAD model.

function reduces contrast in the electrical domain, leading to reduction in contrast in the acoustic domain [Fig. 2(l)].

Figure 8 shows the PSDs of signals for the original signals and processed signals using the two acoustic models for the 4-, 7- and 16-channel conditions. There are visible changes to the PSDs in most cases, but some of the changes are less pronounced than others. The PSD for the vowels plylt and plilt appears minimally affected in the STANDARD model for 7 and 16 channels, but the spectral contrast is visibly changed in the SPREAD model. This effect is more severe at 16 channels and appears more severe for the vowel plilt in these examples.

Figure 9(a) provides a comparison of effective signal levels at different electrodes (i.e., a spatial signal level profile) at a given instant in time for electrodes separated by 1 mm. It shows that noise bands implicitly representing a current decay of 13 dB/mm (the average noise band filter slope) would minimally affect the effective spatial level profile. The error introduced by the use of noise bands is estimated to reduce the explicitly modeled current decay of 7 dB/mm to an effective current decay of around 6 dB/mm. [The trace for a current decay of 6 dB/mm is not shown in Fig. 9(a), as it coincides with the trace for the 7 dB/mm combined with the noise filter of 13 dB/mm.] Figure 9(a) also shows the effects of different values of current decay. It appears that current decay of around 13 dB/mm allows effective representation of the original envelope, with minimal effects on spectral contrast. At a current decay of 3 dB/mm, there is a severe degradation of the signal envelope, and the spectral peak at electrode 3 is lost.

E. Effect of different compression functions

The effects of the compression function were investigated by studying PSDs of vowels processed using a linear compression and power-law compression, combined with current decay of 7 dB/mm. Equation (4) was used to calculate the current for the power-law compression function (Fu and Shannon, 1998), with Eq. (5) giving the logarithmic compression function (Mishra, 2000)

$$I = T + k(s - T_a)^c, \tag{4}$$

$$I = A\log(s) + K, \tag{5}$$

where I is the current in microamperes; T is the electrical threshold; k, K, and A are constants; s is the linear acoustic signal intensity; T_a is the lower extreme of the acoustic dynamic range in linear units; and c is the power-law compression factor. The values of K, k, and A may be solved for from the boundary conditions. That is, if $s = C_a$, then I should equal C, where C is the electrical comfort level and C_a is the upper extreme of the acoustic dynamic range. Also, if $s = T_a$, I should equal T. For example, the values of K and A are 354.8 and 84.9, respectively, for $T_a = 0.001$ and a 60 dB input dynamic range mapped to an electrical dynamic range between 100 μA (T) and 355 μA (C). The value of k would be 254.8 for power-law compression with $c = 0.05$ under the same conditions as above.

Results are shown in Figs. 8(e) and 9(b). Figure 8(e) shows that power-law compression with a compression factor 0.05 yields PSDs similar to those obtained with 3 dB/mm current decay. Figure 9(b) shows that the spatial signal level profile obtained with a power-law compression factor of 0.05 is similar to that obtained with a current decay of 3 dB/mm. Both the power-law compression factor of 0.05 (combined with current decay of 7 dB/mm) and the 3 dB/mm current decay appear to cause decreases in peak-to-trough ratio (abbreviated as PTR in Fig. 8) for the vowels in Fig. 8(e). For plylt and plilt [Fig. 8(e)], the more compressive function ($c = 0.05$) causes loss of contrast between the two spectral peaks.

IV. DISCUSSION

A. Asymptote in speech intelligibility

Modeling the effects of current decay of 7 dB/mm, whilst fixing parameters for electrode spacing and dynamic range to suitable values, appears to explain the asymptote in speech intelligibility at seven channels at all noise levels for vowel, consonant, and sentence intelligibility.

1. Vowel intelligibility

The asymptote in vowel intelligibility at seven channels in the SPREAD model may be explained by the compromising of spectral cues that already emerges at seven channels [e.g., vowels plylt and plilt in Fig. 8(c)] and appears to worsen for some vowels at 16 channels [plylt and plilt in Fig. 8(d)]. A decrease in spectral contrast [formant peak contrast (PC) in Fig. 8] may be observed between F1 and F2 in Fig. 8, along with decreased PTRs for F1 and F2 (visible in both Figs. 8 and 9). Other spectral distortions include merging of F1 and F2 peaks [e.g., vowels plɑlt and plɔlt in Figs. 8(b)–(d)] and a slight shifting of the F1 peaks toward higher frequencies.

The movement of formant peaks is minimal, except in the case where the F1 and F2 peaks merge, where the shift may be more [e.g., plɑlt and plɔlt in Fig. 8(c)]. The slight movement of F1 is caused by the assumed insertion depth of 25 mm.

The decrease in PTR is caused by current spread, as shown in Figs. 9(a), 2(k), and 2(l). This decrease is evident in all vowels in Fig. 8 at 4, 7, and 16 channels. Loizou and Poroy (2001) found significant effects of spectral contrast for vowel recognition. Small separations in formant peaks [peak separation (PS), defined in Fig. 8], such as those observed in back vowels (e.g., plɔlt and plɑlt), typically result in merging of formant peaks when current spread is large enough or, equivalently, major decreases in PTR, as illustrated in Fig. 9(a) between electrodes 9 and 12. The merging of F1 and F2 peaks also appear in the STANDARD model at 4 and 7 channels [e.g., plɑlt at 4 and 7 channels, plɔlt at 4, 7, and 16 channels, Fig. 8(b)–8(d)]. In both models, this merging may also be caused by the bandpass filter widths, which could not provide fine enough resolution to separate the F1 and F2 peaks (e.g., plɑlt at four channels and plɔlt at seven channels). For the vowel plɑlt, however, the STANDARD model's bandpass filters at 16 channels allowed the separation of formant peaks, but in the SPREAD model these peaks were merged due to current spread.

Changes in spectral PC (in Fig. 8, e.g., for the vowel p|i|t) appear to be caused by current spread but in a more complex manner than for PTR. Current spread from strong higher frequency channels [examples encircled in Fig. 8(a) for the vowel p|i|t] appears to be a main cause thereof, since these channels would typically have much larger effects on the F2 channels than on the F1 channels, causing the F2 peak to become more dominant [as illustrated for p|i|t in Fig. 8(d)]. The separation between the peaks (PS) and relative magnitude of the peaks (PC) all contribute to this effect, as illustrated by comparing Fig. 8(a) and 8(d) for the vowels p|y|t and p|i|t. The compression function used could also play a role in this, since it typically decreases contrast in the electrical domain [Fig. 2(b)], making some channels more vulnerable to electrical field interaction resulting from current spread.

2. Consonant intelligibility

Consonant recognition and consonant feature intelligibility also showed asymptote at seven channels. The SPREAD model results in compromised spectral cues, as discussed for vowel intelligibility. These cues are compromised even at 4 and 7 channels, as illustrated in Fig. 8. The spectral cue changes appear relatively large (changing relative strengths of spectral channels and changes in PTRs) at the lowest frequency channels [comparing Figs. 2(a) and 2(e)] and somewhat smaller (changes mostly in terms of lowered PTRs) at the higher frequency channels [comparing Figs. 2(f) and 2(j)], where consonants are mainly coded. The SPREAD model also alters temporal envelope cues, as is evident in channel 1 when comparing Figs. 2(a) and 2(e), for example. This channel shows that the temporal modulations are changed both in depth and in shape for the time 0.3–0.5 s, which typically represent the |t| of the utterance p|ɑ|t. Although this is clearly visible for channel 1 in Fig. 2(e), the same trend may be observed at other channels. These changes in temporal modulations in the SPREAD model would amplify the noise at all noise levels.

Consonant intelligibility may be described by the features of voicing, manner, and place of articulation, the first two of which are mainly affected by temporal envelope cues and the last mainly by spectral cues (Xu *et al.*, 2005). It has been illustrated (Fu *et al.*, 1998; Friesen *et al.*, 2001; Fu and Nogaki, 2005) that spectral cues become more important as the SNR becomes poorer. This effect could have caused consonant and sentence intelligibility (Figs. 3–5) in the present study to drop substantially at +5 dB SNR. The same effect was not observed for vowel intelligibility in the present study, presumably since vowel intelligibility relies strongly on spectral cues at all noise levels (Xu and Zheng, 2007) and was already affected even at +15 dB SNR.

At seven channels (Figs. 4 and 5) in the SPREAD model, at +10 and +15 dB SNR, it appears as if listeners were able to utilize mostly salient temporal cues to reach a high level of consonant intelligibility, close to the no-spread condition of the STANDARD model. It is surprising that the place of articulation feature transmission was similar to that of the STANDARD model at seven channels at the better noise levels, considering the reliance of this feature's transmission on spectral cues (Xu *et al.*, 2005). It may be that the place of articulation feature relies more on transmission of second formant information (Miller and Nicely, 1955), which appears to be less affected by current spread than first formant information [comparing Figs. 2(f) and 2(j), channels 4–6]. At +5 dB SNR for seven channels, the STANDARD model afforded good intelligibility, likely due to salient spectral cues, which now dominated the recognition task, since the temporal cues would be compromised (by noise) at this noise level. In the SPREAD model at +5 dB SNR, both spectral and temporal cues are compromised, the first by electrical field interaction caused by current spread, and the second by noise, making the recognition task very challenging.

The asymptotic behavior of the results in Figs. 4 and 5 suggests that compromising of cues that affect consonant intelligibility becomes serious at 16 channels, when the simulated electrodes are closest together, offsetting the possible benefits of the additional spectral channels.

3. Sentence intelligibility

Sentence intelligibility in the SPREAD model appears to asymptote at seven channels at all noise levels. Sentence intelligibility in the present study appeared quite robust to the electrical field interaction caused by current spread (Fig. 3), most likely owing to the practice that the listeners had had. However, sentence intelligibility dropped significantly at high noise levels (Fig. 3, +5 dB SNR). Sentence intelligibility appears to be dominated increasingly by the limitations imposed by poor vowel intelligibility (and compromised spectral cues) as the SNR deteriorates, leading to an increasing deviation from the STANDARD model results (Fig. 3). When modest noise was present, listeners were able to overcome poor vowel intelligibility and were able to extract sufficient information, possibly relying more on temporal cues (that had not yet been affected to a great extent by noise), rather than the compromised spectral cues. However, as noise masked temporal cues increasingly at poorer SNRs, listeners were probably forced to rely more on the compromised spectral cues. This increased reliance on spectral cues, rather than temporal cues, at poor SNRs has been illustrated previously (Fu *et al.*, 1998; Fu and Nogaki, 2005).

The low scores at four channels in the SPREAD model for all speech material cannot be explained by insertion depth effects, since the mismatch between synthesis filter and analysis filter center frequencies is minimal at four channels. Also, when the electrode spacing is 4 mm, electrical field interaction should be minimal. There are, however, two aspects that could amplify the channel interaction caused by current spread. First, the analysis bandpass filters reduce the spectral contrast visibly, as illustrated in the STANDARD model in Fig. 8(b), compared to the spectral contrast of the original signal [Fig. 8(a)]. Second, the compression function would decrease this contrast still further, as may be seen in Figs. 2(a) and 2(b), which will amplify the electrode interaction. These combined effects lead to the visible decrease in spectral contrast when comparing the PSDs for the STANDARD and SPREAD models in Fig. 8(b). The effects of

decreased spectral contrast for speech intelligibility are more important at a lower number of spectral channels than at a higher number (Loizou and Poroy, 2001), which could explain the low score at four channels. The study by Bingabr *et al*. (2008) showed scores at four channels that were even lower than the present study SPREAD model scores.

B. Comparison with other acoustic models

The difference in speech material, filter cutoffs, and noise material complicated comparison with other acoustic models. The model by Bingabr *et al*. (2008), which modeled spread of excitation, and the model by Baskent and Shannon (2003), which modeled compression of the analysis range and insertion depth effects in quiet, yielded results quite close to the present study results. Conversely, results from the models by Fu and Nogaki (2005) and Boothroyd *et al*. (1996) differed substantially from the SPREAD model results, as well as from CI listener results, generally predicting much lower scores than those of CI listeners.

The model by Bingabr *et al*. (2008) did not demonstrate an asymptote at seven channels. Intelligibility improved up to 16 channels for both HINT sentences and CNC words. This model did include aspects of dynamic range by finding equivalent filter slopes in the acoustic domain for the assumed current decays, but possible effects of the nonlinear compression function were not considered. Their sentence intelligibility results were very close to the SPREAD model results, except at four channels at +10 dB SNR, where their results were much lower than the SPREAD model results. Although the results of the study by Bingabr *et al*. did not show the asymptote, they are quite close to the SPREAD model results, while using a simpler approach. This approach, however, cannot model effects of the compression function and does not provide as much flexibility in modeling the electrical interface or in the choice of the synthesis signal. The Baskent and Shannon model (2003), which modeled insertion depth and frequency range compression effects, did not investigate the asymptote at seven channels. It included implicitly the effect of current spread using noise-band vocoders. As results were only obtained for quiet conditions, it is uncertain how well the model would correspond to implant listener results in noisy conditions. The results of this model for a 5 mm compression of the analysis filter range into the synthesis filter range were very close to the SPREAD model results, indicating that frequency range compression and insertion depth effects, when combined with implicit modeling of current spread, could provide results similar to the SPREAD model in quiet. This model also yielded consonant intelligibility results that were substantially higher than implant listener results.

C. Comparison with CI listener results

Implant listener vowel intelligibility appears to be reasonably well modeled with the SPREAD acoustic model. Consonant intelligibility, however, appears to differ substantially. The first possible explanation of this could be a difference in speech material used in different studies. Some studies with implant listeners produced consonant intelligibility results of around 70% or better in quiet (e.g., Fu and Shannon, 1998; Loizou *et al*., 2000b; Shannon *et al*., 2002), while other studies reported CI listener consonant recognition scores of 60% or worse (e.g., Friesen *et al*., 2001; Zeng *et al*., 2002; Loizou *et al*., 2003). Another reasonable explanation for this inability of the SPREAD model to predict consonant intelligibility correctly may lie in the assumptions of the model or omissions in the model. Dynamic range, insertion depth, and current spread are all highly variable across CI listeners. Kral *et al*. (1998) showed that there may be greater spread of excitation in the basal regions of the cochlea, where many of the consonants are primarily encoded. The Boothroyd *et al*. model (1996) showed consistently lower results than the study by Friesen *et al*. (2001) for consonant intelligibility using 707 Hz for the synthesis filter widths. Although these values are much lower than those of CI listeners, it may be the clue to improving correspondence with CI listener results. Effects of nerve survival could influence consonant intelligibility in CI listeners, but that alone cannot account for the substantially lower scores of the CI listeners in the Friesen *et al*. study, as illustrated in the study by Baskent (2006) with hearing-impaired listeners. Spectral asynchrony, variable thresholds in CI listeners, forward masking, and difference in modulation detection thresholds are also possible candidates for causing the lowered consonant intelligibility scores in CI listeners. Of these, only spectral asynchrony and forward masking have been modeled in acoustic models (e.g., Healy and Bacon, 2002; Throckmorton and Collins, 2002). Forward-masking effects, as modeled by Throckmorton and Collins (2002), yielded relatively high consonant intelligibility scores (75% for the worst-case masking model), which suggests that forward-masking effects are not the cause of lowered consonant intelligibility scores in implant listeners in the study by Friesen *et al*. (2001). Whitmal *et al*. (2007) have shown that narrowband Gaussian noise carriers yield substantially lower consonant intelligibility scores, presumably due to the higher modulation detection thresholds of these signals. The choice of synthesis signal could therefore be an important key to finding an acoustic model that yields results that are closer to results of CI listeners for a wider range of speech material.

D. Effect of modeled current decay

Figure 9(a) represents the spatial signal level profile at a single time instant for the vowel p|i|t, which illustrates how the slopes of current decay typically influence the relative strengths of the effective current at the neural populations closest to each electrode at a given time. Figures 2(k) and 2(l) show how current decay reduces the spectral contrast for the vowel p|ɑ|t. Similar effects are observed in the spectra of the signals, as shown in Fig. 8. The overall shape of the signal envelope appears to be preserved at values of current decay of around 7 dB/mm and higher, although the PTRs become smaller as the current decay values decrease (i.e., the amount of current spread increases). Loizou and Poroy (2001) found significant effects of spectral contrast for vowel recognition. It could be expected that a decrease in current decay would lead to a decrease in vowel intelligibility due to

the reduced contrast at lower decay values, observed in Fig. 9(a). The peaks at electrodes 9 and 12 only appear to become distinguishable at a current decay of 13 dB/mm. At this point, an increase in F2 and F3 transmissions and improved vowel recognition are expected for vowels that are spectrally similar to this one (e.g., p|i|t and p|y|t). These effects would vary across vowels that have different formant patterns. The signal envelope for the vowel p|ɔ|t, for example, retains its shape and is minimally altered by the current spread. This is confirmed by its PSD in Fig. 8(d).

E. Effect of the compression function

The compression function appears to influence spectral contrast in a manner similar to current decay. Note, for example, the similarity in traces between the logarithmic compression function trace for a −3 dB/mm current decay [Fig. 9(a)] and the trace for the power-law compression with a compression factor of 0.05 combined with −7 dB/mm current decay [Fig. 9(b)]. Figure 8(e) also shows similarity in the PSDs of power-law compression ($c = 0.05$) and current decay of −3 dB/mm. It appears therefore as if more compressive functions exacerbate electrical field interaction caused by current spread. Linear compression for the vowel p|i|t appears to minimize electrical field interaction [Fig. 9(b)] and also to preserve the spectral peaks for p|i|t [Fig. 8(d)] (although with reduced PTR), but the PSDs for linear compression [Fig. 8(e)] for the other vowels suggest that linear compression ($c = 1$) presents other problems, for example, failure to suppress high-frequency noise components, as evidenced from the high-frequency tail in the linear compression traces. Also, maintaining normal loudness growth in CI listeners requires the use of non-linear compression functions for optimal perception (Fu and Shannon, 2000).

V. CONCLUSION

(1) The approach used in the present study provides a more flexible way of modeling the electrical field interaction caused by current spread in an acoustic CI model, when compared to simpler approaches used in earlier studies (e.g., Bingabr *et al.*, 2008). Specifically, whereas the use of noise bands as synthesis signals may be used to model current spread, the present approach allowed a separation between the choice of synthesis signal and the way in which electrical field interaction resulting from current spread is modeled.

(2) This approach facilitated the finding that non-linear dynamic range compression of the signal exacerbates the electrical field interaction caused by current spread. Thus, the effective number of information channels may be reduced by using compressive mapping in CI processing, with more compressive functions being more detrimental.

(3) The SPREAD acoustic model, which explicitly modeled electrical field interaction caused by current spread, along with appropriate assumptions about dynamic range compression, electrode spacing, and insertion depth, was able to explain the asymptote in speech intelligibility at seven channels at all the noise levels for all speech material used in this study. The asymptote appears to arise from current spread that compromised spectral cues.

(4) It follows that improving the selectivity of stimulation (e.g., by improved electrode designs or improved stimulation paradigms) and thereby decreasing electrical field interaction has the potential to improve CI performance.

(5) Furthermore, careful design of the compressive mapping function may reduce electrical field interaction. However, retaining normal loudness growth is an opposing challenge.

(6) The SPREAD model results for consonant and sentence intelligibility; however, it did not correspond quantitatively to the selected set of CI listener results. Consonant and sentence intelligibility appeared to be more robust against electrical field interaction than vowels, except at +5 dB SNR, where sentence intelligibility for the SPREAD model was quite close to that of implant listeners.

ACKNOWLEDGMENTS

This study was made possible in part by grants from the National Research Foundation. The authors would also like to thank listeners for their commitment and time spent on the experiments.

[1]In an acoustic model, the analysis filters are those used to analyze the input signal into contiguous frequency bands, while the synthesis filters are used to define the widths of noise bands that are used in acoustic models that simulate current spread with band limited noise. Generally, these differ from the analysis filters.

Baer, T., and Moore, B. C. J. (**1993**). "Effects of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am. **94**, 1229–1241.
Baskent, D. (**2006**). "Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels," J. Acoust. Soc. Am. **120**, 2908–2925.
Baskent, D., and Shannon, R. V. (**2003**). "Speech recognition under conditions of frequency-place compression and expansion," J. Acoust. Soc. Am. **113**, 2064–2076.
Baskent, D., and Shannon, R. V. (**2005**). "Interactions between cochlear implant electrode insertion depth and frequency-place mapping," J. Acoust. Soc. Am. **117**, 1405–1416.
Baskent, D., and Shannon, R. V. (**2007**). "Combined effects of frequency compression-expansion and shift on speech recognition," Ear Hear. **28**, 277–289.
Bingabr, M., Espinoza-Varas, B., and Loizou, P. C. (**2008**). "Simulating the effect of spread of excitation in cochlear implants," Hear. Res. **241**, 73–79.
Blamey, P. J., Dowell, R. C., Tong, Y. C., and Clark, G. M. (**1984**). "An acoustic model of a multiple-channel cochlear implant," J. Acoust. Soc. Am. **76**, 97–103.
Boothroyd, A., Mulhearn, B., Gong, J., and Ostroff, J. (**1996**). "Effects of spectral smearing on phoneme and word recognition," J. Acoust. Soc. Am. **100**, 1807–1818.
Buechner, A., Frohne-Buechner, C., Gaertner, L., Lesinski-Schiedat, A., Battmer, R. D., and Lenarz, T. (**2006**). "Evaluation of Advanced Bionics high resolution mode," Int. J. Audiol. **45**, 407–416.
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.
Fishman, K. E., Shannon, R. V., and Slattery, W. H. (**1997**). "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor," J. Speech. Lang. Hear. R. **40**, 1201–1215.
Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.
Frijns, J. H. M., de Snoo, S. L., and Schoonhoven, R. (**1995**). "Potential distributions and neural excitation patterns in a rotationally symmetric model of the electrically stimulated cochlea," Hear. Res. **87**, 170–186.

Frijns, J. H. M., Klop, W. M. C., Bonnet, R. M., and Briaire, J. J. (**2003**). "Optimizing the number of electrodes with high-rate stimulation of the Clarion CII cochlear implant," Acta Oto-Laryngol. **123**, 138–142.

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

Fu, Q. J., and Shannon, R. V. (**1998**). "Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **104**, 2570–2577.

Fu, Q. J., and Shannon, R. V. (**2000**). "Effects of dynamic range and amplitude mapping on phoneme recognition in Nucleus-22 cochlear implant users," Ear Hear. **21**, 227.

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

Geurts, L., and Wouters, J. (**2000**). "A concept for a research tool for experiments with cochlear implant users," J. Acoust. Soc. Am. **108**, 2949–2956.

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

Hanekom, T. (**2001**). "Three-dimensional spiraling finite element model of the electrically stimulated cochlea," Ear Hear. **22**, 300–315.

Hanekom, T. (**2005**). "Modelling encapsulation tissue around cochlear implant electrodes," Med. Biol. Eng. Comput. **43**, 47–55.

Healy, E. W., and Bacon, S. P. (**2002**). "Across-frequency comparison of temporal speech information by listeners with normal and impaired hearing," J. Speech Lang. Hear. Res. **45**, 1262–1275.

Holm, S. (**1979**). "A simple sequentially rejective multiple test procedure," Scand. J. Stat. **6**, 65–70.

Kiefer, J., Ilberg, C., Rupprecht, V., Hubnet-Egener, J., Baumgartner, W., Gstottner, W., Forgasi, K., and Stephan, K. (**1997**). "Optimized speech understanding with the CIS speech coding strategy in cochlear implants: The effect of variations in stimulus rate and number of channels," in *Vth International Cochlear Implant Conference*, New York, NY, pp. 1009–1020.

Kral, A., Hartmann, R., Mortazavi, D., and Klinke, R. (**1998**). "Spatial resolution of cochlear implants: The electrical field and excitation of auditory afferents," Hear. Res. **121**, 11–28.

Kreft, H. A., Donaldson, G. S., and Nelson, D. A. (**2004**). "Effects of pulse rate on threshold and dynamic range in Clarion cochlear-implant users (L)," J. Acoust. Soc. Am. **115**, 1885–1888.

Kwon, B. J., and van den Honert, C. (**2006**). "Effect of electrode configuration on psychophysical forward masking in cochlear implant listeners," J. Acoust. Soc. Am. **119**, 2994–3002.

Laneau, J., Wouters, J., and Moonen, M. (**2006**). "Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants," J. Acoust. Soc. Am. **119**, 491–506.

Loizou, P. C. (**2006**). "Speech processing in vocoder-centric cochlear implants," Adv. Otorhinolaryngol. **64**, 109–143.

Loizou, P. C., Dorman, M., and Fitzke, J. (**2000a**). "The effect of reduced dynamic range on speech understanding: Implications for patients with cochlear implants," Ear Hear. **21**, 25–31.

Loizou, P. C., and Poroy, O. (**2001**). "Minimum spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners," J. Acoust. Soc. Am. **110**, 1619–1637.

Loizou, P. C., Poroy, O., and Dorman, M. (**2000b**). "The effect of parametric variations of cochlear implant processors on speech understanding," J. Acoust. Soc. Am. **108**, 790–802.

Loizou, P. C., Stickney, G., Mishra, L., and Assmann, P. (**2003**). "Comparison of speech processing strategies used in the Clarion implant processor," Ear Hear. **24**, 12–19.

Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Mishra, L. N. (**2000**). "Analysis of speech processing strategies for the Clarion implant processor," Master of Science dissertation, University of Texas, Dallas.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

Pretorius, L. L., Hanekom, J. J., Van Wieringen, A., and Wouters, J. (**2006**). "n Analitiese tegniek om die foneem-herkenningsvermoë van Suid-Afrikaanse koglêre inplantingsgebruikers te bepaal (Analytical technique to determine the phoneme-recognition ability of South African cochlear implant users)," Die Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie **25**, 195–207.

Shannon, R. V. (**1983**). "Multi-channel electrical stimulation of the auditory nerve in man. II. Channel interaction," Hear. Res. **12**, 1–16.

Shannon, R. V., Galvin, J. J., III, and Baskent, D. (**2002**). "Holes in hearing," J. Assoc. Res. Otolaryngol. **3**, 185–199.

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

Theunissen, M., Swanepoel, D., and Hanekom, J. J. (**2008**). "A comparison of list compilation methods in a test of sentence recognition in noise," in *XXIXth International Congress of Audiology in Hong Kong*, Hong Kong Convention and Exhibition Centre, Wanchai, Hong Kong.

Throckmorton, C. S., and Collins, L. M. (**2002**). "The effect of channel interactions on speech recognition in cochlear implant subjects: Predictions from an acoustic model," J. Acoust. Soc. Am. **112**, 285–296.

van Wieringen, A., and Wouters, J. (**1999**). "Natural vowel and consonant recognition by Laura cochlear implantees," Ear Hear. **20**, 89–103.

Whitmal, N. A., III, Poissant, S. F., Freyman, R. L., and Helfer, K. S. (**2007**). "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," J. Acoust. Soc. Am. **122**, 2376–2388.

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

Xu, L., and Zheng, Y. (**2007**). "Spectral and temporal cues for phoneme recognition in noise," J. Acoust. Soc. Am. **122**, 1758–1764.

Zeng, F. G., Grant, G., Niparko, J., Galvin, J., Shannon, R., Opie, J., and Segel, P. (**2002**). "Speech dynamic range and its effect on cochlear implant performance," J. Acoust. Soc. Am. **111**, 377–386.