Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech

Michael K. Qin[a] and Andrew J. Oxenham[b]
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and Harvard-MIT Division of Health Sciences and Technology, Speech and Hearing Bioscience and Technology Program, Cambridge, Massachusetts 02139

(Received 9 April 2005; revised 26 January 2006; accepted 29 January 2006)

This study investigated the benefits of adding unprocessed low-frequency information to acoustic simulations of cochlear-implant processing in normal-hearing listeners. Implant processing was simulated using an eight-channel noise-excited envelope vocoder, and low-frequency information was added by replacing the lower frequency channels of the processor with a low-pass-filtered version of the original stimulus. Experiment 1 measured sentence-level speech reception as a function of target-to-masker ratio, with either steady-state speech-shaped noise or single-talker maskers. Experiment 2 measured listeners' ability to identify two vowels presented simultaneously, as a function of the *F*0 difference between the two vowels. In both experiments low-frequency information was added below either 300 or 600 Hz. The introduction of the additional low-frequency information led to substantial and significant improvements in performance in both experiments, with a greater improvement observed for the higher (600 Hz) than for the lower (300 Hz) cutoff frequency. However, performance never equaled performance in the unprocessed conditions. The results confirm other recent demonstrations that added low-frequency information can provide significant benefits in intelligibility, which may at least in part be attributed to improvements in *F*0 representation. The findings provide further support for efforts to make use of residual acoustic hearing in cochlear-implant users.
[DOI: 10.1121/1.2178719]

PACS number(s): 43.71.Bp, 43.71.Es, 43.71.Ky [PFA] Pages: 2417–2426

I. INTRODUCTION

Despite many significant advances made in the development of cochlear implants (e.g., Dorman, 2000; Zeng, 2004), even the most successful cochlear-implant users do not hear as well as normal-hearing listeners, particularly in noise. Fu *et al.* (1998) and Friesen *et al.* (2001) reported that implant users require higher target-to-masker ratios in broadband noise to achieve levels of speech reception performance comparable to normal-hearing listeners. More recent findings suggest that the differences in performance between normal-hearing listeners and implant users (real and simulated) are especially pronounced for speech in the presence of more complex, fluctuating backgrounds, such as modulated noise or single-talker interference (Nelson *et al.*, 2003; Qin and Oxenham, 2003; Stickney *et al.*, 2004).

Part of the difficulty experienced by cochlear-implant users in fluctuating backgrounds may reflect an impaired ability to perceptually segregate the fluctuations of the target from those of the masker. Voice pitch, or the fundamental frequency (*F*0) of voicing, has long been thought to be an important cue in the perceptual segregation of speech sources (e.g., Bregman, 1990; Darwin and Carlyon, 1995). While, in principle, voice pitch information is available to implant users via envelope modulations (McKay *et al.*, 1994; Wilson, 1997; Geurts and Wouters, 2001; Green *et al.*, 2002; Moore, 2003), the pitch salience associated with envelope periodicity is known to be less robust than that associated with resolved lower-order harmonics in normal-hearing listeners (e.g., Burns and Viemeister, 1976; 1981; Shackleton and Carlyon, 1994; Kaernbach and Bering, 2001; Bernstein and Oxenham, 2003). Qin and Oxenham (2003) suggested that the difference in speech reception performance between (real and simulated) implant users and normal-hearing listeners in fluctuating maskers can be attributed in part to the loss of pitch as a segregation cue on the part of the implant users. They tested this idea more directly in a later study (Qin and Oxenham, 2005) and found that, under envelope-vocoder processing, large *F*0 differences, which were readily discriminated in a sequential task, could not be used to improve the recognition of two simultaneously presented vowels. A possible explanation for this deficit may be that the auditory system cannot use envelope-periodicity cues to extract the *F*0's of two concurrent sounds, and instead (in normal hearing) relies on low-order resolved harmonics as the primary segregation cue (Carlyon, 1996; Deeks and Carlyon, 2004). Listeners' lack of ability to use envelope-based pitch cues to segregate sources may therefore account for at least part of the difficulties encountered by cochlear-implant users in complex (fluctuating) backgrounds.

A relatively recent trend is to consider patients with some low-frequency residual hearing as candidates for cochlear implantation. Some of these individuals are able to

[a]Electronic mail: qin@nsmrl.navy.mil
[b]Current address: Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455. Electronic mail: oxenham@umn.edu

use an implant in one ear while using an acoustic hearing aid in the other ear to form a hybrid electric-plus-acoustic system (EAS). While the residual acoustic hearing present in these implant users is unlikely to contribute directly to speech intelligibility, the additional low-frequency temporal fine-structure cue in acoustic hearing may provide sufficient information to aid in source segregation. In fact, a recent study by Kong *et al.* (2005) has shown that speech reception in the presence of interference improved with combined electric and acoustic hearing compared to either alone.

In another variant of an EAS application, short-insertion cochlear implants have been developed, which stimulate only the basal end of the cochlea, leaving the apical (low-frequency) end sufficiently intact to still preserve low-frequency residual hearing (von Ilberg *et al.*, 1999; Gantz and Turner, 2003). A study by Turner *et al.* (2004) showed that three patients with short-insertion cochlear implants and amplified residual hearing performed better than any of their patients with traditional cochlear implants for speech recognition in a background of competing speech, but not in steady noise. The studies of Turner *et al.* (2004) and Kong *et al.* (2005) suggest that residual acoustic hearing can lead to substantial improvements in performance over cochlear-implant stimulation alone.

In addition to testing implant patients, Turner *et al.* (2004) also tested normal-hearing listeners using a noise-excited envelope vocoder (EV) (e.g., Shannon *et al.*, 1995), which is a type of channel vocoder (Dudley, 1939) designed to simulate aspects of cochlear-implant processing. They investigated the effect of introducing unprocessed low-frequency information on listeners' ability to recognize two-syllable words (spondees) in a background of either steady-state noise or two-talker babble. Using a 16-channel EV and steady-state noise interference, they found no effect of processing: EV processing did not degrade performance, relative to the unprocessed condition, and adding back unprocessed acoustic information below 500 Hz did not improve performance. In contrast, performance in two-talker babble was degraded by 16-channel EV processing, and was partially restored by reintroducing the unprocessed signal below 500 Hz.

These results reinforce the idea that low-frequency fine-structure information might be particularly useful in complex, fluctuating backgrounds, where perceptual segregation is necessary (e.g., Qin and Oxenham, 2003). However, some questions remain. First, how much low-frequency information is required to observe an improvement in speech reception? Turner *et al.* (2004) used a cutoff frequency of 500 Hz, which provided a good approximation to the residual hearing of their implant patients. Information below 500 Hz provides *F*0 information, but also provides information about the first formant of many vowels. Is a similar benefit observed if the additional information is limited to even lower frequencies? Second, does the additional low-frequency information aid performance in situations where dynamic fluctuations and dynamic grouping cues are not present?

To address the two questions above, the current study used two low-pass cutoff frequencies (300 and 600 Hz) to simulate varying extents of residual hearing, and to probe the limits of improvement due to additional low-frequency information. Experiment 1 measured speech-reception accuracy for sentences as a function of target-to-masker ratio in the presence of speech-shaped steady-state noise or a competing single talker. Experiment 2 measured vowel identification for synthetic vowels presented either alone or simultaneously in pairs, as a function of the *F*0 difference between the two vowels. In both experiments the processing conditions included traditional EV processing and two conditions in which the lowest two or three EV frequency channels were replaced by a low-pass-filtered version of the original stimulus.

II. EXPERIMENT 1: SPEECH RECEPTION IN THE PRESENCE OF INTERFERENCE

A. Methods

1. Participants

Eighteen normal-hearing (audiometric thresholds between 125 and 8000 Hz were $<$20 dB HL) native speakers of American English participated in this study. Their ages ranged from 18 to 28.

2. Stimuli

All stimuli in this study were composed of a target sentence presented in the presence of a masker. The stimulus tokens were processed prior to each experiment. The targets were H.I.N.T. sentences (Nilsson *et al.*, 1994) spoken by a male talker. The maskers were either a different male competing talker or speech-shaped noise. The targets and maskers were combined at various target-to-masker ratios (TMRs) prior to any processing. TMRs were computed based on the token-length root-mean-square amplitudes of the signals. The maskers began and ended 500 ms before and after the target, respectively. They were gated on and off with 250-ms raised-cosine ramps.

The H.I.N.T. sentence corpus consists of 260 phonetically balanced high-context sentences of low-to-moderate difficulty. Each sentence contains four to seven keywords. The competing-talker maskers were excerpts from the audio book "Timeline" (novel by M. Crichton) read by Stephen Lang (as used in Qin and Oxenham, 2003). The competing-talker masker had a mean *F*0 (mean=111.4 Hz, s.d. =27.06 Hz) similar to that of the target talker (mean =110.8 Hz, s.d.=24.15 Hz), as estimated by the YIN program (de Cheveigné and Kawahara, 2002). To avoid long silent intervals in the masking speech (e.g., sentence-level pauses), the competing-talker maskers were automatically preprocessed to remove silent intervals longer than 100 ms. The competing-talker maskers and speech-shaped noise maskers were spectrally shaped to match the long-term power spectrum of the H.I.N.T. sentences. The maskers were then subdivided into nonoverlapping segments to be presented in each trial.

3. Stimulus processing

The experimental stimuli for each listener were digitally generated, processed, and stored on disk prior to each experi-

TABLE I. Filter cutoffs for the noise-excited vocoders (3 dB down points).

	EV_{1-8}		$LF_{300}+EV_{3-8}$		$LF_{600}+EV_{4-8}$	
Channel number	Low (kHz)	High (kHz)	Low (kHz)	High (kHz)	Low (kHz)	High (kHz)
1	0.080	0.221	Unprocessed (0.080–0.300)		Unprocessed (0.080–0.600)	
2	0.221	0.426				
3	0.426	0.724	0.426	0.724		
4	0.724	1.158	0.724	1.158	0.724	1.158
5	1.158	1.790	1.158	1.790	1.158	1.790
6	1.790	2.710	1.790	2.710	1.790	2.710
7	2.710	4.050	2.710	4.050	2.710	4.050
8	4.050	6.000	4.050	6.000	4.050	6.000

ment. Stimulus processing was performed using MATLAB (Mathworks, Natick, MA) in the following manner.

The experimental stimuli were presented in four processing conditions. In all conditions, the target level was fixed at 65 dB SPL and the masker levels were varied to meet the desired TMR. In the first processing condition (unprocessed), the stimuli were filtered between 80 Hz and 6 kHz, but were otherwise left unchanged. The second processing condition (EV_{1-8}), designed to simulate the effects of envelope-vocoder implant processing, used an 8-channel noise-excited vocoder (Qin and Oxenham, 2005). The input stimulus was first bandpass filtered (sixth-order Butterworth filters) into eight contiguous frequency channels between 80 and 6000 Hz (see Table I). The entire frequency range was divided equally in terms of the Cam scale (Glasberg and Moore, 1990; Hartmann, 1997). The envelopes of the signals were extracted by half-wave rectification and low-pass filtering (using a second-order Butterworth filter) at 300 Hz, or half the bandpass filter bandwidth, whichever was lower. The 300-Hz cutoff frequency was chosen to preserve as far as possible the *F*0 cues in the envelope. The envelopes were then used to amplitude modulate independent white-noise carriers. The same bandpass filters that were used to filter the original stimuli were then used to filter the amplitude-modulated noises. Finally, the modulated narrow-band noises were summed and scaled to have the same level as the original stimuli.

The last two processing conditions were designed to simulate "electric plus acoustic" systems (EAS). The unprocessed stimuli were low-pass filtered using a sixth-order Butterworth filter with 3-dB down points at 300 and 600 Hz (LF_{300} and LF_{600} conditions). To simulate the effects of EAS with residual hearing below 300 Hz ($LF_{300}+EV_{3-8}$), LF_{300} was paired together with EV_{3-8}, consisting of the upper six channels of the eight-channel vocoder. The first formant frequency of most vowels exceeds 300 Hz (Hillenbrand *et al.*, 1995), so by limiting information to below that frequency, most direct speech information should be eliminated. To simulate the effects of an EAS with residual hearing below 600 Hz ($LF_{600}+EV_{4-8}$), LF_{600} was paired with EV_{4-8}, consisting of the upper five channels of the eight-channel vocoder simulation. The 600-Hz cutoff was selected to incorporate the first formant frequency of most vowels, and to include the most dominant harmonics in the formation of pitch percepts (Moore *et al.*, 1985; Dai, 2000). The low-pass filter cutoff frequencies were somewhat below the lower cutoff frequencies of the EV, leaving small (roughly 1 critical band) spectral gaps between the two (unprocessed and processed) regions (see Table I). This was done to limit the low-frequency information to the desired region, while maintaining the same channel configuration that was used in the full EV_{1-8} condition.

4. Procedure

The 18 participants were divided into two groups of nine. The speech reception of each group was measured under only one of the masker types (competing talker or speech-shaped noise). The experiment involved the participants listening to sentences in a background masker and entering as much as they could of the target sentence via a computer keyboard. No feedback was given.

Prior to the experiment session, listeners were given practice performing the experimental tasks as well as given exposure to the processed stimuli. The target sentences used in the training came from the IEEE corpus (IEEE, 1969), whereas the target sentences in the experiment sessions came from the H.I.N.T. corpus. While the maskers in the training and experiment session came from the same corpus, care was taken to ensure that the same masker token was never repeated. During the training session, the listener was exposed to a total of 35 stimulus tokens (five lists, with seven sentences per list), in each of the four processing conditions. In each processing condition, the target sentences were presented at a TMR in the midrange of the experimental TMRs (see Table II). The listeners were instructed to enter their responses via a computer keyboard, as in the experiment. No feedback was given.

In the experiment session, speech reception was measured for each listener under all four processing conditions (unprocessed, EV_{1-8}, $LF_{300}+EV_{3-8}$, $LF_{600}\_EV_{4-8}$), at 5 TMRs (see Table II), in the presence of one masker type. The TMRs for each processing condition and masker type were determined in an earlier pilot study, using two to three listeners. The TMRs were chosen to minimize floor and ceiling effects in the psychometric function. For each listener and condition, the target sentence lists were chosen at random, without replacement, from among the 25 lists of H.I.N.T.

TABLE II. List of conditions tested.

Processing condition	Masker type	Target-to-masker ratio (dB)
Unprocessed	Competing talker	[−20,−15,−10,−5,0]
	Steady-state noise	[−10,−7,−5,−3,0]
NEV_{1-8}	Competing talker	[−5,0,5,10,15]
	Steady-state noise	[−5,−1,2,6,10]
$LPF_{300}+NEV_{3-8}$	Competing talker	[−10,−5,0,5,10]
	Steady-state noise	[−5,−1,2,6,10]
$LPF_{600}+NEV_{4-8}$	Competing talker	[−10,−6,−3,1,5]
	Steady-state noise	[−10,−6,−3,1,5]

sentences. This was done to ensure that no target sentence was presented more than once to any given listener. Data were collected using one list (i.e., ten sentences) for each TMR in each condition. For each listener, the different conditions were run in random order.

The stimuli were converted to the analog domain using a soundcard (LynxStudio, LynxOne) at 16-bit resolution with a sampling rate of 22 050 Hz. They were then passed through a headphone buffer (TDT HB6) and presented diotically via Sennheiser HD580 headphones to the listener seated in a double-walled sound-attenuating booth.

Listener responses were scored offline by the experimenter. Obvious misspellings of the correct word were considered correct. Each listener's responses to the ten sentences at a given TMR, under a given masker condition, were grouped together to produce a percent-correct score that was based on the number of correct keywords.

B. Results and discussion

1. Fits to the psychometric functions

The percent-correct scores as a function of TMR under a given masker condition for each listener were fitted to a two-parameter sigmoid model (a cumulative Gaussian function),

$$\text{Percent correct} = \frac{100}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\text{TMR}} \exp\left(\frac{-(x-\text{SRT})^2}{2\sigma^2}\right)dx, \tag{4.1}$$

where x is the integration variable, SRT is the speech reception threshold in dB at which 50% of words were correctly identified, σ is related to the slope of the function, and TMR is the target-to-masker ratio (dB). The two-parameter model assumes that listeners' peak reception performance is 100%. The two-parameter model provided generally good fits to the data. The individual standard-errors-of-fit had a mean of 2.65% with a standard deviation of 2.52% (median of 1.87% and a worst case of 12.93%).

2. Speech-reception thresholds (SRT)

The mean SRT values and standard errors of means were derived from the SRT values of individual model fits. In general, performance across listeners was consistent, so only the mean SRT values as a function of masker condition and processing condition are plotted in Fig. 1. A higher SRT value implies a condition more detrimental to speech reception. Figure 1 shows that in the unprocessed conditions (right-most bars), the steady-state noise masker was a more effective masker than the competing-talker masker. However, with eight-channel envelope-vocoder processing (EV_{1-8}) the reverse was true (i.e., competing talker was the more effective masker), consistent with the findings of Qin and Oxenham (2003). When unprocessed low-frequency information (LF) was added to the envelope-vocoder processed speech, the SRT associated with both maskers decreased, indicating a speech reception benefit.

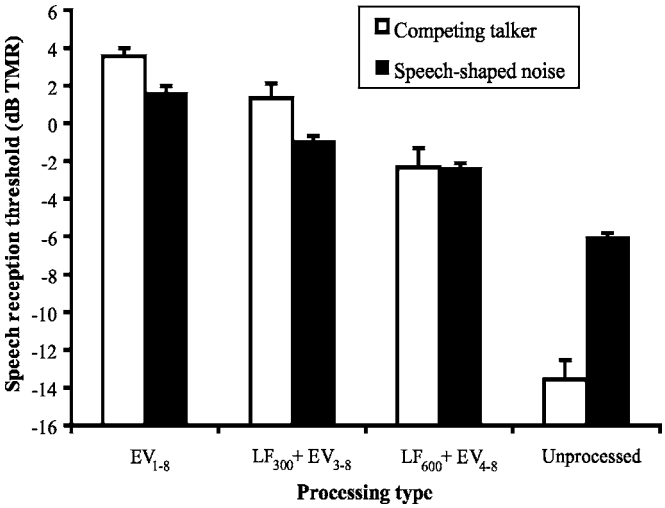

FIG. 1. Group mean speech reception threshold (SRT) values for the two types of background interference. The error bars denote ±1 standard error of the mean.

A two-way mixed-design analysis of variance (ANOVA) was performed on the data for the EV_{1-8}, $LF_{300}+EV_{3-8}$, and $LF_{600}+EV_{4-8}$ conditions and for both noise and single-talker maskers. The statistical significance of the findings was determined with SRT as the dependent variable, masker type as the between-subjects factor, and processing condition as the within-subjects factor. The ANOVA showed significant main effects of both processing condition ($F_{2,32}=57.16$, $p<0.05$) and masker type ($F_{1,16}=5.18$, $p<0.05$). There was also a significant interaction between masker type and processing condition ($F_{2,32}=3.49$, $p<0.05$). This may reflect the fact that adding low-frequency information seems to have a somewhat greater effect for the single-talker masker than for the steady-state noise masker. In particular, the SRT difference between the competing-talker and steady-state maskers, which was present in the EV_{1-8} condition, was not found in the $LF_{600}+EV_{4-8}$ condition. Note, however, that even with the 600-Hz low-pass information, listeners were still not able to benefit from the single-talker masker fluctuations as they were in the unprocessed condition.

Fisher's *post hoc* least-significance difference (LSD) test ($\alpha=0.05$) revealed some differences between individual conditions. In the presence of speech-shaped noise, SRTs were higher in the EV_{1-8} condition than in the $LF_{300}+EV_{3-8}$ condition, which in turn was higher than in the $LF_{600}+EV_{4-8}$ condition. Thus, increasing the amount of low-frequency information produced a systematic improvement in performance. Similar improvements were observed with the single-talker interferer, although the difference in SRT between the

EV_{1-8} condition and the $LF_{300}+EV_{3-8}$ condition failed to reach significance in the *post hoc* tests.

The larger effect of adding low-frequency information below 600 Hz for the competing talker than for the steady-state noise is consistent with the results of Turner *et al.* (2004). However, in contrast to their findings, we also observed a substantial improvement in the steady-state noise condition. This difference may be due to a number of factors. Most importantly, performance in their 16-channel EV condition was the same as in the unprocessed condition, leaving no room for improvement by reintroducing low-frequency information. In contrast, we used an eight-channel EV, which more accurately simulates the abilities of the best current implant users (Friesen *et al.*, 2001), and which also provided a substantial decrease in performance relative to the unprocessed condition. Also, their closed-set 12-alternative spondee identification task may have required fewer cues to perform correctly than our sentence task, leading to high performance in the EV condition. In any case, our results suggest that low-frequency residual hearing may be beneficial in a variety of backgrounds, including steady-state noise.

In summary, the addition of unprocessed low-frequency information improved performance in speech-shaped noise at both cutoff frequencies (300 and 600 Hz). The $LF_{600}+EV_{4-8}$ condition performance also produced a significant improvement compared to the EV-alone condition in the presence of competing talkers. Adding unprocessed information below 600 Hz to envelope-vocoded high-frequency information improved speech reception threshold by about 6 dB in the presence of a competing talker and by about 4 dB in the presence of speech-shaped noise compared to the EV-only condition. Although the addition of low-frequency information did not return speech reception performance to normal levels, it was nevertheless a significant improvement when compared with envelope-vocoder alone.

III. EXPERIMENT 2: CONCURRENT-VOWEL IDENTIFICATION

A. Rationale

Experiment 1 showed that low-frequency information can improve performance in a sentence recognition task under both steady-state noise and single-talker interference. The cues that are provided by the low-frequency information are primarily *F*0 and, particularly with the 600-Hz cutoff, first-formant information. Particularly in the case of the single-talker interference, amplitude fluctuations in the target and the interferer lead to many time instances where one or the other dominate. The addition of coherently fluctuating low-frequency *F*0 information may help listeners identify these instances and thus improve performance. Experiment 2 used a concurrent-vowel paradigm (Scheffers, 1983), which rules out these time-varying aspects and their accompanying perceptual segregation cues, such as onset asynchronies and different dynamic frequency and amplitude trajectories. The concurrent-vowel paradigm, despite its artificial nature, also eliminates possible semantic and grammatical context cues that might have influenced performance in our sentence recognition task. It has been a popular task to probe the role of *F*0, and other potential segregation cues (such as interaural time differences) in separating sounds in normal (Brokx and Nooteboom, 1982; Assmann and Summerfield, 1990; Summerfield and Assmann, 1991; Culling and Darwin, 1993; Assmann and Summerfield, 1994; Culling and Darwin, 1994; Darwin and Carlyon, 1995; de Cheveigné *et al.*, 1997; Bird and Darwin, 1998) and impaired (Arehart *et al.*, 1997) hearing.

TABLE III. Formant frequencies (Hz) for vowels. Values enclosed in parentheses represent formant bandwidths (Hz).

	Formant			
Vowel	*F*1 (60)	*F*2 (90)	*F*3 (150)	*F*4 (200)
/i/	342	2322	3000	3657
/A/	768	1333	2522	3687
/u/	378	997	2343	3357
/ɛ/	580	1799	2605	3677
/ɝ/	474	1379	1710	3334

By using concurrent vowels that differed only in *F*0, we were able to assess the extent to which reintroducing low-frequency acoustic information aided listeners in their ability to use *F*0 differences to segregate two simultaneous sounds. As in experiment 1, two levels of residual hearing (<300 Hz and <600 Hz) were tested.

B. Methods

1. Participants

Six native speakers of American English (audiometric thresholds between 125 and 8000 Hz were <20 dB HL) were paid for their participation in this experiment. Their ages ranged from 19 to 26.

2. Stimuli

Five American English vowels (/i/ as in *heed*, /A/ as in *hod*, /u/ as in *hood*, /ɛ/ as in *head*, /ɝ/ as in *herd*) were synthesized using an implementation of Klatt's cascade synthesizer (Klatt, 1980). They were generated at a sampling frequency of 20 kHz, with 16-bit quantization. The formant frequencies (see Table III) used to synthesize the vowels were based on the estimates of Hillenbrand *et al.* (1995) for an average male talker. The vowels were chosen because of their positions in the $F1-F2$ space; because their natural duration characteristics (House, 1960; 1961) are similar to the stimulus durations used in this experiment (i.e., 200 ms); and because they can all be reasonably approximated by time-invariant formant patterns. Each vowel was generated with seven different *F*0's ranging from 0 to 14 semitones above 100 Hz (100, 105.9, 112.2, 126.0, 158.7, 200.0, and 224.5 Hz).

The concurrent-vowel pairs were constructed by summing two single vowels with equal rms levels, with their onsets and offsets aligned, and with their pitch periods in phase at the onset of the stimulus. No vowel was paired with itself to generate the concurrent-vowel pairs. Each concurrent-vowel token was constructed using one vowel

with an $F0$ of 100 Hz and the other with an $F0$ of 100 Hz $+\Delta F0$, where the $\Delta F0$ ranged from 0 to 14 semitones. This yielded a total of 140 concurrent-vowel stimuli (20 vowel pairs $\times 7$ $\Delta F0$'s). Each stimulus had a total duration of 200 ms and was gated on and off with 25-ms raised-cosine ramps. The stimuli were presented at an overall rms level of 70 dB SPL.

3. Stimulus processing

All stimulus tokens were digitally generated, processed, and stored on computer disk prior to the experiments. The experimental stimuli were presented in seven conditions (unprocessed, LF_{300}, LF_{600}, EV_{3-8}, EV_{4-8}, $LF_{300}+EV_{3-8}$, and $LF_{600}+EV_{4-8}$). The additional low-frequency-only conditions (LF_{300} and LF_{600}) were added to test how much information could be gleaned from those frequencies in isolation. Given that the frequency of the first formant was always above 300 Hz but mostly below 600 Hz, it could be predicted that no identification would be possible in the LF_{300} condition, but that some degree of performance might be possible in the LF_{600} condition.

4. Procedure

Performance in single-vowel and concurrent-vowel identification tasks was measured using a forced-choice paradigm. Listeners were instructed to identify the vowels heard by selecting visual icons associated with the vowels. In the single-vowel identification task, listeners were instructed to identify the vowel heard by selecting from five different choices. In the concurrent-vowel identification task, listeners were instructed to identify both of the constituent vowels, and the response was only marked correct if both vowels were correctly identified. The responses were entered via a computer keyboard and mouse inside the booth. No feedback was provided.

Each listener took part in six 2-h sessions. Three sessions incorporated the unprocessed, LF_{300}, EV_{3-8}, and $LF_{300}+EV_{3-8}$ conditions, and the three other sessions incorporated the unprocessed, LF_{600}, EV_{4-8}, and $LF_{600}+EV_{4-8}$ conditions. The 300-Hz and 600-Hz sessions were interleaved, with the order randomized across subjects.

Each experiment session was subdivided into eight blocks, with each block involving a single condition (unprocessed, LF, EV, and LF+EV). The first four blocks measured single-vowel identification and the next four blocks measured concurrent-vowel identification. The order of the blocks was randomized from session to session.

Within a given block, the stimulus tokens were presented in random order. Within each single-vowel identification block, a total of 70 stimulus tokens was presented (7 $F0$'s $\times 5$ vowels $\times 2$ repetitions). For each listener, this translates to a total of 30 trials (5 vowels $\times 2$ repetitions $\times 3$ sessions) at each $F0$ under each processing condition. Within each concurrent-vowel identification block, a total of 140 stimulus tokens was presented (7 $\Delta F0$'s $\times 20$ vowel pairs). For each listener, this translates to a total of 60 trials (20 vowel pairs $\times 3$ sessions) at each $F0$ under each processing condition. In the unprocessed condition listeners were exposed to twice as many trials at each $F0$, because the unprocessed condition was presented in each of the six sessions. The six sessions took place over a span of 2–3 weeks, depending on the availability of the participants.

Prior to each experiment session, every listener was given practice performing the experimental tasks. As in the actual experiment, listeners were instructed to enter their responses via the computer keyboard, and no feedback was provided. On average, listeners were exposed to 40–80 stimulus tokens prior to data gathering.

C. Results and discussion

Figure 2 shows the identification accuracy, in percent correct, as a function of the $F0$ in the single-vowel identification task (dotted lines) and as a function of the $F0$ of the upper vowel in the concurrent-vowel identification task (symbols and solid lines), in units of semitones above 100 Hz. The unprocessed conditions are shown in Fig. 2(a), the LF conditions in Fig. 2(b), the EV conditions in Fig. 2(c), and the LF+EV conditions in Fig. 2(d). To investigate trends in the data, repeated-measures ANOVAs were conducted. All scores were arcsine transformed prior to statistical analysis.

In the unprocessed conditions [Fig. 2(a)] a single-factor repeated-measures ANOVA was conducted with $\Delta F0$ as the within-subject factor. The analysis revealed that the effect of $\Delta F0$ was statistically significant ($F_{6,30}=11.87$, $p<0.05$), as illustrated by the average improvement in performance from 78% to 95% as $\Delta F0$ increases from 0 to 2 semitones. Above this $\Delta F0$, performance plateaus until the $\Delta F0$ equals 12 semitones (one octave), consistent with previous studies (Brokx and Nooteboom, 1982; Culling and Darwin, 1993; Qin and Oxenham, 2005). When the $\Delta F0$ equals one octave, the harmonics of the two constituent vowels become inseparable, leading to a drop in identification performance. At $\Delta F0$ of 14 semitones, the identification performance seems to improve somewhat, although the performance difference between 12 and 14 semitones was not statistically significant (Fisher's LSD, $p>0.05$).

In the LF conditions [Fig. 2(b)], identification performance is greatly reduced, as predicted by the articulation index (AI) and the speech intelligibility index (SII) (ANSI, 1997). When a repeated-measures ANOVA with two within-subject factors ($\Delta F0$ and low-pass cutoff) was conducted, a statistically significant difference was found between LF_{300} and LF_{600} ($F_{1,5}=11.03$, $p<0.05$), but there was no statistically significant effect of $\Delta F0$ ($F_{6,30}=1.67$, $p>0.1$) and no interaction ($F_{6,30}<1$, n.s.). In the LF_{300} condition, performance for both single and double vowels was reduced to around chance (20% and 10%, respectively) and no benefits of $F0$ differences were seen. In the LF_{600} condition, although identification of both single and concurrent vowels improved, still no benefit of $F0$ differences was observed.

In the EV conditions [Fig. 2(c)], while single-vowel identification was generally high, concurrent-vowel identification was modest. When a repeated-measures ANOVA with two within-subject factors ($\Delta F0$ and low-pass cutoff) was conducted, no statistically significant difference was found between EV_{3-8} and EV_{4-8} ($F_{1,5}=1.67$, $p>0.1$). In addition,

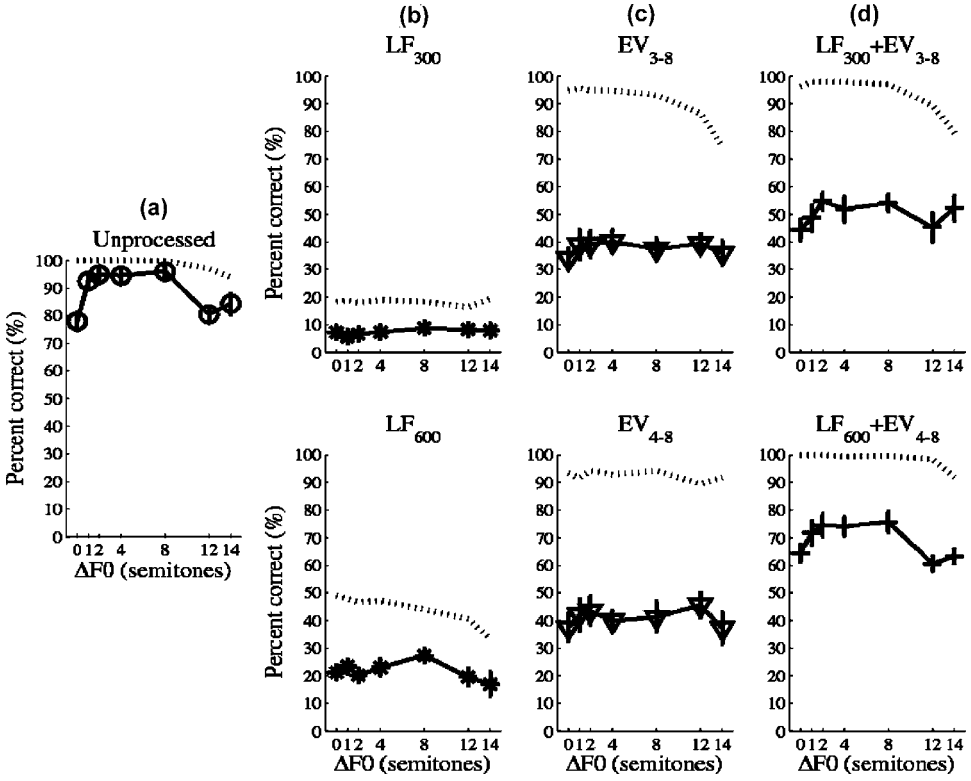

FIG. 2. Dotted lines show the percent of correct responses as a function of the $F0$ in the single-vowel identification task (dotted lines), where $F0$ is described in terms of the number of semitones from 100 Hz. Solid lines show the percent of correct responses as a function of the $\Delta F0$ (in semitones) between constituent vowels in the concurrent-vowel identification task, where the lower $F0$ was always 100 Hz. The error bars denote ±1 standard error of the mean. The unprocessed conditions are shown in the left panel (a), next the low-pass filtered (LF) conditions (b), then the envelope-vocoder (EV) conditions (c), finally the LF+EV conditions in the right panels (d). The top row of figures is associated with the 300-Hz low-pass sessions, and the bottom row of figures is associated with the 600-Hz low-pass sessions.

no effect of $\Delta F0$ ($F_{6,30}=1.67$, $p>0.1$), and no interaction ($F_{6,30}<1$, n.s.), was seen. This suggests that vowel identification performance does not improve as a function of $F0$ difference in either of the EV conditions, consistent with observations from our previous study using both 8- and 24-channel EV processing (Qin and Oxenham, 2005). A direct comparison of the EV_{4-8} condition in this experiment and the EV_{1-8} condition in Qin and Oxenham (2005) suggests only a small improvement due to the addition of the three lowest vocoder channels. In particular, the averaged percent correct for the eight subjects in that study, pooled across $\Delta F0$'s, was 45.3%, compared with 41.0% in the present study.

In the LF+EV conditions [Fig. 2(d)], identification performance improved compared to both LF-only and EV-only conditions, and overall performance in the $LF_{600}+EV_{4-8}$ condition was higher than that in the $LF_{300}+EV_{3-8}$ condition. A repeated-measures ANOVA with two within-subject factors ($\Delta F0$ and low-pass cutoff) showed significant main effects of cutoff ($F_{1,5}=23.71$, $p<0.05$) and $\Delta F0$ ($F_{6,30}=6.73$, $p<0.05$). These effects remained significant when the 12- and 14-semitone conditions were excluded, suggesting that the effect is truly due to the improvement in performance with increasing $\Delta F0$ at the lower $\Delta F0$'s. The interaction between cutoff and $\Delta F0$ was also found to be significant ($F_{6,30}=2.59$, $p<0.05$). However, when the data from the 12- and 14-semitone $\Delta F0$ conditions were excluded from the analysis, no significant interaction remained between cutoff and $\Delta F0$ ($F_{4,20}<1$, n.s.), indicating that the benefits of increasing $\Delta F0$ in the $LF_{300}+EV_{3-8}$ and $LF_{600}+EV_{4-8}$ conditions were not statistically different for moderate $\Delta F0$ between 0 and 8 semitones. Overall, the pattern of results in the $LF_{300}+EV_{3-8}$ and $LF_{600}+EV_{4-8}$ conditions is similar to that found in the unprocessed condition, although overall performance remains lower even in the $LF_{600}+EV_{4-8}$ condition, as was also found in experiment 1.

It has been suggested that part of the $\Delta F0$ benefit observed in concurrent-vowel identification at small values of $\Delta F0$ (1 semitone or less) may arise from "beating" between adjacent components with concurrent vowels (Assmann and Summerfield, 1994; Culling and Darwin, 1994; Darwin and Carlyon, 1995; Bird and Darwin, 1998). At least two types of interaction exist. The first involves two individual low-order harmonics which interact, producing a pattern of amplitude and frequency modulation; the second involves interactions between groups of unresolved harmonics that produce so-called pitch-period asynchronies (e.g., Assmann and Summerfield, 1994). Both types of interaction are present to some extent in our stimuli, with the interaction between individual harmonics present in the LF region, and only the pitch-pulse asynchronies present in the EV region (albeit somewhat masked by the inherent fluctuations of the noise carriers). It is interesting to note that the pitch-pulse asynchronies seem not to be sufficient on their own to improve vowel identification, given that the EV-alone conditions showed no benefit of differences in $F0$ (see also Qin and Oxenham, 2005). In any case, most of our conditions involved $\Delta F0$'s that were greater than those thought to be influenced by slowly varying beats.

In summary, reintroducing low-frequency information led to an overall improvement in the identification of concurrent vowels, and also led to a positive effect of $F0$ difference between the two vowels that was not present in either the LF-alone or the EV-alone conditions.

IV. DISCUSSION

The results from the two experiments can be summarized as follows. In experiment 1, speech reception for sentences in both steady-state noise and single-talker interference improved as (unprocessed) low-frequency information was added to EV simulations of cochlear-implant processing. In experiment 2, the addition of low-frequency information improved overall identification of two concurrently presented vowels, and led to a benefit of *F*0 differences between the two vowels, even when the cutoff frequency was as low as 300 Hz.

Our results are broadly consistent with those of Turner *et al.* (2004) in showing an improvement due to unprocessed low-frequency information. Our findings also extend the previous results in the following ways. First, sentence recognition can improve with additional low-frequency information even in steady-state noise. Second, improvements can be observed with a low-pass cutoff frequency as low as 300 Hz. Third, benefits of low-frequency information are found with steady-state vowel sounds, where dynamic grouping cues, and "glimpsing" of the target, are not available. This provides further support for approaches that attempt to combine acoustic with electric stimulation in cochlear-implant users with some residual hearing (Kong *et al.*, 2004; Turner *et al.*, 2004). However, it is important to note that our simulations involve "ideal" residual hearing, with no hearing loss and accompanying effects, such as broadened auditory filters. These conditions are unlikely to hold in real EAS users. On the other hand, real EAS users will have time to adjust to their processing scheme, and are perhaps more likely to perceptually fuse the two modes of presentation after prolonged exposure. In any case, care should be taken when interpreting the current findings in terms of potential benefits for implant patients. With that caveat in mind, some possible implications of our work can be examined.

A. Frequency extent of residual hearing necessary for tangible benefits in speech reception

Given the potential variability in the amount of residual hearing available across the population of cochlear implant candidates, one aim of the current study was to investigate the frequency extent of residual hearing necessary to show tangible speech reception benefits. Our findings suggest that even an extremely limited range (<300 Hz) of residual hearing may be beneficial to the reception of speech in the presence of interference. Both experiment 1 and 2 showed that when unprocessed information below 300 Hz was added to envelope-vocoder processing, significant improvements in speech identification could be observed. In experiment 1, the SRT decreased by 2.5 dB in steady-state noise. This 2.5-dB improvement in SRT translates to an improvement in intelligibility of about 20 percentage points for the sentence material used in our tests. In experiment 2, concurrent-vowel identification improved beyond that of the vocoder-only condition. In addition, listeners exhibited $\Delta F0$ benefits that were not observed in the vocoder-only conditions. While the addition of low-frequency information did not return speech reception performance to normal levels in either experiment, the improvement was nevertheless significant when compared with envelope-vocoder alone. The current findings, taken together with the positive results from real EAS users (Tyler *et al.*, 2002; Ching *et al.*, 2004; Turner *et al.*, 2004; Kong *et al.*, 2005), lead us to be cautiously optimistic about the ability of combined electric and acoustic stimulation to enhance the perceptual segregation of speech.

B. Role of *F*0 representation in improving performance

As stated in the Introduction, previous researchers (Turner *et al.*, 2004; Kong *et al.*, 2005) have suggested that speech reception benefits of residual hearing may be attributable to an improvement in *F*0 representation. However, the cochlear-implant subjects in their studies had residual hearing up to frequencies as high as 1 kHz. With this level of residual acoustic hearing, speech-reception benefits could be attributed at least in part to increased spectral resolution and more accurate formant frequency information, rather than to improvements in *F*0 representation alone. Even in the acoustic simulations of Turner *et al.* (2004) the cutoff frequency of 500 Hz was probably too high to rule out the improved representation of the first formant in the LF region. The current simulation experiment examined the effect of adding unprocessed information, low-pass filtered at 300 Hz. Because the lower cutoff frequency of the first noise band was at 426 Hz, and because all the stimuli were filtered with sixth-order Butterworth filters, the crossover frequency between the unprocessed low-frequency stimuli and the first noise band was at around 360 Hz, at which point the filters were roughly 10 dB below their peak value. As the lower-frequency noise band would likely mask any information beyond this crossover frequency, it is reasonable to assume that only reduced (and spectrally distorted) unprocessed low-frequency fine-structure information was available beyond 300 Hz, and none was available beyond 360 Hz. This frequency range is thought to contain very little speech information (ANSI, 1997) and, in the case of our vowels, would not have been sufficient to fully represent the spectral peak of any of the first formants in our experiment. This is supported by the fact that our listeners performed at chance in the single-vowel identification task when listening only to the unprocessed stimuli, low-pass filtered at 300 Hz. Thus, it is reasonable to conclude that an improvement in *F*0 representation aided performance for both the vowels and the sentences.

With the higher cutoff frequency of 600 Hz, part of the improvement in performance was likely due to improved spectral representation of the first formant. This is suggested in the results of experiment 2 by the fact that single- and double-vowel identification was no longer at chance in the LF_{600} condition. In contrast, the benefits of increasing $\Delta F0$ in concurrent-vowel identification results were similar for both the 300- and 600-Hz cutoff frequencies, supporting the notion that the addition of unprocessed low-frequency information improved *F*0 representation and thus improved speech segregation abilities.

For most conditions in experiment 2, the *F*0's were sufficiently far apart (2 semitones or more) to provide at least

one or two resolved harmonics for each vowel. In the 1-semitone condition that might not be the case, given that 1 semitone corresponds to 6%, which is smaller than the limit usually ascribed to peripheral resolvability (e.g., Plomp, 1964; Bernstein and Oxenham, 2003). However, depending on the vowel pairing, the amplitude of the fundamental components of concurrent vowels can be significantly different. The amplitude of the component at the *F*0 varies as the frequency of the first formant varies. For example, the amplitude of *F*0 for the vowel /i/ is higher than that of the vowel /A/, because the /i/'s *F*1 (i.e., 342 Hz) is closer to the *F*0 than /A/'s *F*1 (i.e., 768 Hz). The difference in amplitude between fundamental components of vowels may aid in the resolvability of the dominant *F*0 component by rendering the other *F*0 less audible, thereby increasing the ability to segregate the concurrent vowels.

It is currently unclear exactly how *F*0 information from the low-frequency region is used to help perceptual segregation and speech recognition of EV-processed speech sounds. It is possible that the additional *F*0 information aids the grouping of channels in the envelope-vocoder processed region. In experiment 1 with the competing talker, this could be on a moment-by-moment basis, with the *F*0 information cuing which voice is dominant at a given time; with the steady-state masker, it may simply improve pitch contour perception, which has been shown to aid lip-reading (Rosen *et al.*, 1981) and may similarly aid in interpreting EV-processed information. In experiment 2, when two vowels are presented simultaneously at similar intensities, there can still be spectral regions that are dominated by the energy from one of the vowels. An improved representation of *F*0 through low-frequency information may aid the listener in grouping together different spectral regions, perhaps in conjunction with top-down or "template" mechanisms that make use of prior knowledge of vowel structures as well as the *F*0 cues in the envelope that alone are too weak a cue for perceptual segregation.

As mentioned above, the benefit of adding low-frequency information, even when it is low-pass filtered at 300 Hz, is reminiscent of a technique that presented the *F*0 of voiced speech to profoundly hearing-impaired listeners as a potential aid to visual speech cues (Rosen *et al.*, 1981; Faulkner *et al.*, 1992). This technique was originally proposed as a possible alternative to cochlear implantation, but the present results, along with encouraging previous simulation results (Turner *et al.*, 2004) and results in actual implant patients (Turner *et al.*, 2004; Kong *et al.*, 2005), suggest that such a scheme (either via *F*0 tracking, or simply presenting unprocessed low-frequency acoustic information) may be a valuable supplement to cochlear implant patients with some residual hearing.

V. SUMMARY

(1) In a sentence recognition task, adding unprocessed information below 300 or 600 Hz led to significant improvements in speech reception for envelope-vocoder-processed speech in both steady-state speech-shaped noise and single-talker interference.
(2) In a concurrent-vowel identification task, significant benefits of *F*0 differences between the two vowels were observed when low-frequency information was introduced, even when it was over a very limited range (<300 Hz).
(3) The improvements, particularly in the case of the 300-Hz cutoff frequency, can probably be attributed to an improvement in *F*0 representation through the presence of low-order harmonics.
(4) The findings presented here extend those of Turner *et al.* (2004) and Kong *et al.* (2005), providing further support for schemes involving a combination of electric and acoustic stimulation (EAS), and suggest that even very limited residual hearing has the potential to provide substantial speech recognition benefits, when amplified and combined with a cochlear implant.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIDCD Grant R01 DC 05216). We thank Joshua Bernstein, Louis Braida, Christophe Micheyl, Barbara Shinn-Cunningham, Andrea Simonson, and the reviewers for helpful comments on earlier versions of this manuscript.

ANSI (**1997**). S3.5-1997 "Methods for calculation of the speech intelligibility index," (American National Standards Institute, New York).
Arehart, K. H., King, C. A., and McLean-Mudgett, K. S. (**1997**). "Role of fundamental frequency differences in the perceptual separation of competing vowel sounds by listeners with normal hearing and listeners with hearing loss," J. Speech Lang. Hear. Res. **40**, 1434–1444.
Assmann, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.
Assmann, P. F., and Summerfield, Q. (**1994**). "The contribution of waveform interactions to the perception of concurrent vowels," J. Acoust. Soc. Am. **95**, 471–484.
Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?" J. Acoust. Soc. Am. **113**, 3323–3334.
Bird, J., and Darwin, C. J. (**1998**). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 263–269.
Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (Bradford Books, MIT Press, Cambridge, MA).
Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.
Burns, E. M., and Viemeister, N. F. (**1976**). "Nonspectral pitch," J. Acoust. Soc. Am. **60**, 863–869.
Burns, E. M., and Viemeister, N. F. (**1981**). "Played again SAM: Further observations on the pitch of amplitude-modulated noise," J. Acoust. Soc. Am. **70**, 1655–1660.
Carlyon, R. P. (**1996**). "Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker," J. Acoust. Soc. Am. **99**, 517–524.
Ching, T. Y. C., Incerti, P., and Hill, M. (**2004**). "Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears," Ear Hear. **25**, 9–21.
Culling, J. F., and Darwin, C. J. (**1993**). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by *F*0," J. Acoust. Soc. Am. **93**, 3454–3467.
Culling, J. F., and Darwin, C. J. (**1994**). "Perceptual and computational separation of simultaneous vowels: Cues arising from low-frequency beating," J. Acoust. Soc. Am. **95**, 1559–1569.
Dai, H. P. (**2000**). "On the relative influence of individual harmonics on pitch judgment," J. Acoust. Soc. Am. **107**, 953–959.
Darwin, C. J., and Carlyon, R. P. (**1995**). "Auditory Grouping," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).
de Cheveigné, A., and Kawahara, H. (**2002**). "YIN, a fundamental frequency

estimator for speech and music," J. Acoust. Soc. Am. **111**, 1917–1930.
de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (**1997**). "Concurrent vowel identification. I. Effects of relative amplitude and *F*0 difference," J. Acoust. Soc. Am. **101**, 2839–2847.
Deeks, J. M., and Carlyon, R. P. (**2004**). "Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation," J. Acoust. Soc. Am. **115**, 1736–1746.
Dorman, M. F. (**2000**). "Speech perception by adults," in *Cochlear Implants*, edited by S. Waltzman and N. Cohen (Thieme, New York), pp. 317–329.
Dudley, H. W. (**1939**). "Remaking speech," J. Acoust. Soc. Am. **11**, 169–177.
Faulkner, A., Ball, V., Rosen, S., Moore, B. C. J., and Fourcin, A. J. (in press). "Speech pattern hearing aids for the profoundly hearing impaired: Speech perception and auditory abilities," J. Acoust. Soc. Am..
Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.
Fu, Q. J., Shannon, R. V., and Wang, X. S. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.
Gantz, B. J., and Turner, C. W. (**2003**). "Combining acoustic and electrical hearing," Laryngoscope **113**, 1726–1730.
Geurts, L., and Wouters, J. (**2001**). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," J. Acoust. Soc. Am. **109**, 713–726.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Green, T., Faulkner, A., and Rosen, S. (**2002**). "Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants," J. Acoust. Soc. Am. **112**, 2155–2164.
Hartmann, W. M. (**1997**). *Signals, Sound, and Sensation* (Springer, New York).
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.
House, A. S. (**1960**). "Formant band widths and vowel preference," J. Speech Hear. Res. **3**, 3–8.
House, A. S. (**1961**). "On vowel duration in English," J. Acoust. Soc. Am. **33**, 1174–1178.
IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **AU-17**(3), 225–246.
Kaernbach, C., and Bering, C. (**2001**). "Exploring the temporal mechanism involved in the pitch of unresolved harmonics," J. Acoust. Soc. Am. **110**, 1039–1048.
Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.
Kong, Y. Y., Stickney, G. S., and Zeng, F. G. (**2005**). "Speech and melody recognition in binaurally combined acoustic and electric hearing," J. Acoust. Soc. Am. **117**, 1351–1361.
Kong, Y. Y., Cruz, R., Jones, J. A., and Zeng, F. G. (**2004**). "Music perception with temporal cues in acoustic and electric hearing," Ear Hear. **25**, 173–185.
McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). "Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees," J. Acoust. Soc. Am. **96**, 2664–2673.
Moore, B. C. (**2003**). "Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants," Otol. Neurotol. **24**, 243–254.
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). "Relative dominance of individual partials in determining the pitch of complex tones," J. Acoust. Soc. Am. **77**, 1853–1860.
Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.
Nilsson, M., Soli, S., and Sullivan, J. (**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.
Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.
Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.
Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on *F*0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.
Rosen, S. M., Fourcin, A. J., and Moore, B. C. J. (**1981**). "Voice pitch as an aid to lipreading," Nature (London) **291**, 150–152.
Scheffers, M. T. M. (**1983**). "Sifting vowels: Auditory pitch analysis and sound segregation," Groningen University, The Netherlands.
Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency-modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Stickney, G., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.
Summerfield, A. Q., and Assmann, P. F. (**1991**). "Perception of concurrent vowels: Effects of pitch-pulse asynchrony and harmonic misalignment," J. Acoust. Soc. Am. **89**, 1364–1377.
Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (**2004**). "Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing," J. Acoust. Soc. Am. **115**, 1729–1735.
Tyler, R. S., Parkinson, A. J., Wilson, B. S., Witt, S., Preece, J. P., and Noble, W. (**2002**). "Patients utilizing a hearing aid and a cochlear implant: Speech perception and localization," Ear Hear. **23**, 98–105.
von Ilberg, C., Kiefer, J., Tillein, J., Pfenningdorff, T., Hartmann, R., Sturzebecher, E., and Klinke, R. (**1999**). "Electric-acoustic stimulation of the auditory system—New technology for severe hearing loss," ORL **61**, 334–340.
Wilson, B. S. (**1997**). "The future of cochlear implants," Br. J. Audiol. **31**, 205–225.
Zeng, F.-G. (2004). "Trends in cochlear implants," Trends in Amplification **8**, 1–34.