Research Article

Use of Amplitude Modulation Cues Recovered from Frequency Modulation for Cochlear Implant Users When Original Speech Cues Are Severely Degraded

JONG HO WON,¹ HYUN JOON SHIM,² CHRISTIAN LORENZI,³ AND JAY T. RUBINSTEIN⁴

¹Department of Audiology and Speech Pathology, University of Tennessee Health Science Center, Knoxville, TN 37996, USA

²Department of Otolaryngology, Eulji Medical Center, School of Medicine, Eulji University, Seoul, 139-231, Republic of Korea

³Equipe Audition (CNRS, Université Paris Descartes), Institut d'Etude de la Cognition, Ecole Normale Supérieure, Paris Sciences et Lettres, Paris, 75005, France

⁴Virginia Merrill Bloedel Hearing Research Center, Department of Otolaryngology-Head and Neck Surgery, University of Washington, Seattle, WA 98195, USA

Received: 4 April 2013; Accepted: 20 January 2014; Online publication: 15 February 2014

ABSTRACT

Won et al. (*J Acoust Soc Am* 132:1113–1119, 2012) reported that cochlear implant (CI) speech processors generate amplitude-modulation (AM) cues recovered from broadband speech frequency modulation (FM) and that CI users can use these cues for speech identification in quiet. The present study was designed to extend this finding for a wide range of listening conditions, where the original speech cues were severely degraded by manipulating either the acoustic signals or the speech processor. The manipulation of the acoustic signals included the presentation of background noise, simulation of reverberation, and amplitude compression. The manipulation of the speech processor included changing the input dynamic range and the number of channels. For each of these conditions, multiple levels of speech degradation were tested. Speech identification was measured for CI users and compared for stimuli having both AM and FM information (intact condition) or FM information only (FM condition). Each manipulation degraded speech identification performance for both intact and FM conditions. Performance for the intact and FM conditions became similar for stimuli having the most severe degradations. Identification performance generally overlapped for the intact and

FM conditions. Moreover, identification performance for the FM condition was better than chance performance even at the maximum level of distortion. Finally, significant correlations were found between speech identification scores for the intact and FM conditions. Altogether, these results suggest that despite poor frequency selectivity, CI users can make efficient use of AM cues recovered from speech FM in difficult listening situations.

Keywords: speech intelligibility, cochlear implants, amplitude modulation, frequency modulation, envelope reconstruction

INTRODUCTION

Today, clinically available cochlear implant (CI) systems deliver primarily amplitude-modulation (AM) cues, also referred to as acoustic “temporal envelope” cues. With this design, robust speech identification in quiet is possible with four to six frequency channels (Dorman et al. 1989; Fishman et al. 1997; Friesen et al. 2001). However, a serious limitation of this design is that AM cues are highly susceptible to noise. This imposes a rapid deterioration in speech identification performance with increasing levels of background noise (Fu et al. 1998; Friesen et al. 2001; Fu and Nogaki 2005). An alternative method of delivering

Correspondence to: Hyun Joon Shim · Department of Otolaryngology, Eulji Medical Center, School of Medicine · Eulji University · Seoul, 139-231, Republic of Korea. Telephone: +82-2-9708276; fax: +82-2-9708275; email: eardoc11@naver.com

speech information is the use of frequency modulation (FM), often referred to as “acoustic temporal fine structure” (Qin and Oxenham 2003; Zeng et al. 2005; Drennan et al. 2008). FM is attractive because it is less susceptible to background noise (Torick and Keller 1985).

When discussing the limitations of current CI systems, it is generally stated that FM information cannot be transmitted through CIs (e.g., Wilson and Dorman 2008). This may be true regarding the temporal auditory mechanism utilizing phase-locking information conveyed by the responses of auditory nerve fibers, because current CI systems do not evoke neural phase-locking cues. Previous studies, however, demonstrated that the place mechanism utilizing the so-called “FM-to-AM conversion” process is another way of encoding FM information (e.g., Zwicker 1962; Moore and Sek 1996; Saberi and Hafter 1995; Ghitza 2001; Heinz and Swaminathan 2009; Swaminathan and Heinz 2012). Here, the FM-to-AM conversion occurs when the differential attenuation of audio filtering gives rise to the conversion of the frequency excursions of FM into the dynamic variations of the output levels of the filters. This FM-to-AM conversion (or “recovery” of AM from FM) occurs for any system that involves bandpass filtering processes, including the cochlear filters in the peripheral auditory system or digital filters implemented in CIs. However, compared to the normal auditory system, poorer FM-to-AM conversion is expected for CIs because CI filters are much wider and shallower than the normal auditory filters. For example, if the speech processor is mapped with an eight-channel continuous interleaved sampling (CIS) strategy, the bandwidth of the digital filters is approximately three times wider than the bandwidth of the normal auditory filters. Consistent with this prediction, Lorenzi et al. (2012) showed that the broadening of cochlear filters typically associated with sensorineural hearing loss has a detrimental effect on the ability to use the AM speech cues recovered from the broadband FM speech cues, when cochlear filters are broadened by a factor greater than two.

A recent study (Won et al. 2012) focused on the possibility of encoding and using FM speech information for CI users through the FM-to-AM conversion mechanism. They tested the hypotheses that: (1) when broadband FM signals are presented, the CI speech processor generates AM cues recovered from broadband FM as a result of the audio-filtering process; and (2) CI users can use the recovered AM cues for robust speech identification. To test these hypotheses, consonant identification was measured for CI users using speech stimuli presented in quiet and processed with FM vocoders (e.g., Zeng et al. 2005; Gilbert and Lorenzi 2006; Sheft et al. 2008). The FM vocoders preserved the FM speech cues and

discarded the AM cues in each frequency analysis band. FM vocoders were created using four different band numbers: 1, 2, 4, and 8 bands. When a one-band FM vocoder was used, the electrode outputs showed salient AM cues, whereas an eight-band FM vocoder produced extremely weak AM cues. Consistent with these electrode outputs, consonant identification by CI users improved systematically as the number of vocoder band decreased from 8 to 1. For the FM-1 band vocoder, consonant identification scores for CI subjects ranged from 26 to 74 rationalized arcsine units (RAUs), demonstrating that FM speech information may be transmitted via the FM-to-AM conversion mechanism for commercially available implant systems. The previous study by Won et al. (2012) focused on testing the ability of CI users to use AM cues recovered from speech FM in a quiet listening condition. The present study extended this line of research to determine if the AM cues recovered from speech FM convey important phonetic information for CI users when they are communicating in an acoustic environment that degrades the original speech cues.

When communicating with people, CI users receive modulation information of the target speech, but these cues are degraded by multiple sources. Two general sources are identified which could primarily degrade the speech modulation cues: (1) acoustic distortions, and (2) distortions as a result of implant processing. With respect to the acoustic distortions, three different effects were tested: the presence of background noise, reverberation, and amplitude compression. With respect to the distortions that result from implant processing, the effects of different settings of input dynamic range (IDR) and the number of channels were tested.

These five potential sources of distortion were chosen for the present study because they are known to affect the transmission of modulation cues that are crucial for speech understanding for CI users [e.g., background noise, Fu and Nogaki (2005); reverberation, Kokkinakis and Loizou (2011); amplitude compression, Zeng (2004); IDR, Spahr et al. (2007); number of channels, Friesen et al. (2001)]. In addition, multiple levels of distortion were tested for each of these five sources of distortion. For each distortion factor, speech identification was measured for CI users with stimuli having FM information alone (hereafter referred to as FM condition). In addition, speech identification was measured with stimuli having both AM and FM information (hereafter referred to as “intact” condition) to provide an indication of the maximum performance of CI users for each testing condition.

The present study tested three hypotheses: (1) for most of the levels of the distortions tested here,

speech identification should be better than chance performance in the FM condition; (2) when the level of distortion is low, identification performance should be always better for the intact condition than the FM condition, but as the level of distortion increases, performance for the intact and FM conditions should become similar, (3) the ability to use AM cues recovered from speech FM should contribute to the variability in speech identification performance in listening conditions where the original speech AM cues are severely degraded. The second hypothesis would be supported (a) if a significant interaction were found between the level of distortion and speech processing type (intact or FM); and (b) if speech identification performance for the intact and FM conditions became similar as the level of the distortion increased. The third hypothesis would be supported if a significant correlation were found between performance for the intact and FM conditions.

MATERIALS AND METHODS COMMON TO ALL EXPERIMENTS

All testing sessions for five experiments were conducted in a double-walled sound-attenuating (IAC) booth. The size of the booth was 2.74 m×1.98 m×1.98 m (length×width×height) and its reverberation time was 0.1 s. Stimuli were presented to subjects via a Crown D45 amplifier and a free-standing speaker (B&W DM303) located directly in front of the subject at a distance of 1 m. The speaker exceeded ANSI standards for speech audiometry, varying ± 2 dB from 100 to 20,000 Hz. The phase response of the speaker was smooth across frequencies, varying $\pm 30^\circ$ from 150 to 20,000 Hz and $\pm 45^\circ$ below 150 Hz.

Speech Processing

Intact Condition

Original speech signals sampled at 44,100 Hz were passed through a single wideband bandpass filter (12th order FIR). The low and high cutoff frequencies for the bandpass filter were matched to those in the speech processors for each subject (see Table 1). There was no further processing; therefore, original speech AM and FM information were passed through the speech processor. The resulting electrode outputs reflect the AM of the input speech as well as the AM recovered from the FM of the input speech.

FM Condition

Figure 1 shows the conceptual diagram for encoding the signals for the FM condition. The same wideband bandpass filter (Table 1) from the intact condition

was used to filter the original speech signals to produce a broadband speech signal. The Hilbert transform was used to decompose the filtered signal. Then the Hilbert FM (i.e., the cosine of the angle of the analytic signal) was multiplied by the root mean square power of the filtered signal to create the final signals (i.e., FM speech). As a result, in the acoustic domain, only FM cues were preserved while AM cues were discarded for the FM condition. Then the FM cues are passed through the bandpass filters of the speech processor. The resulting electrode outputs reflect the AM recovered from the FM of the input speech. This recovered AM information is finally encoded in the auditory nerve.

For experiments 1–3, speech signals were first submitted to three forms of distortions (adding background noise, simulations of reverberation, and amplitude compression), and then the distorted signals were processed using the intact and FM speech processing described above. For experiments 4 and 5, the original speech signals (without any acoustic distortion) were processed with the intact and FM speech processing, and then subjects were tested with different settings of the speech processor. Except for the IDR testing in experiment 4, a clinically default IDR setting was used for all subjects (30 dB for Cochlear devices; 60 dB for Advanced Bionics devices; and 55 dB for MED-EL devices).

Consonant Identification

Consonant identification was used for experiments 1, 2, 4, and 5. Sixteen bisyllables (vowel–consonant–vowels, VCV) read by three different speakers (one male and two females) were used in a context of /aCa/ (/aba/, /ada/, /afa/, /aga/, /a a/, /aka/, /ala/, /ama/, /ana/, /apa/, /ara/, /asa/, /aʃa/, /ata/, /ava/, /aza/) (Shannon et al. 1999). Before the actual testing, the subjects listened to each consonant more than three times by selecting the button on a computer screen. The practice session used 16 bisyllables read by another female speaker. For the actual testing, a MATLAB (The Mathworks) graphical user interface running on a Mac G5 computer was used to present sound stimuli to listeners. The 16 bisyllables were displayed as virtual buttons on a computer screen and subjects indicated their answer by clicking on the appropriate button using a computer mouse. A single-interval, 16-alternative procedure was used to measure consonant identification. Each trial presented the entire set of VCV stimuli (16 consonants×3 speakers) in random order for a single stimulus version (intact or FM) at a single testing condition (e.g., at a single SNR value). For each testing condition, two trials of consonant identification were run and averaged for individual subjects. Throughout this paper, percent

TABLE 1

Low and high cutoff frequencies (in Hertz) for the intact and FM speech processing for subjects grouped by their CI devices

CI devices	Low cutoff	High cutoff
Advanced bionics	250	8,700
Cochlear	188	7,938
MED-EL	70	8,500

correct scores for consonant identification were corrected for chance level and then transformed to rationalized arcsine units (RAUs). The rationalized arcsine transform (Studebaker 1985) was applied to the proportions of correct responses in order to make the identification scores follow a normal distribution. Therefore, a RAU value of 0 represents performance at chance level. For brevity, speech identification scores are referred to as “speech scores” in this paper.

Pulse-Train Output Measurements

For each experiment, the pulse-train outputs of CI processors for the intact and FM conditions were measured. This was done using the speech processor programmed with a CIS map for one of the CI subjects who participated in this study. Then the electrical pulse trains were taken from an “implant-in-a-box” (Cochlear, Ltd; Advanced Bionics Corp) and recorded using an oscilloscope (Tektronix MSO3014).

EXPERIMENT 1: EFFECTS OF BACKGROUND NOISE

Subjects

For all five experiments reported in this paper, all CI subjects were native speakers of American English and had at least 1 year of experience with their CIs. Individual subject information is listed in Table 2. All experimental procedures followed the regulations set by the National Institutes of Health and were approved by the University of Washington’s Human Subject Institutional Review Board. In experiment 1, 13 postlingually deafened adult CI users participated, but four of them were bilateral users. Bilateral implant users were tested with each CI, therefore, data for 17 implanted ears are reported for experiment 1. All subjects were fitted with an eight-channel CIS strategy to control for the effect of speech processing strategy. It was also desirable to test CI subjects with the same number of channels because the degree of AM recovery varies as a function of the number of analysis bandpass filters. The automatic gain control (AGC)

was turned on during testing as the subjects use the AGC in their everyday life.

Testing Procedure

The 16 bisyllables were presented in steady-state, speech-shaped noise. The steady noise maskers were spectrally shaped to have the same long-term average power spectrum across bisyllables. In all trials, maskers were gated on and off with 50-ms linear ramps. The onset of the target bisyllable was 500 ms after the onset of the background noise. The offset of the background noise was 50 ms after the offset of the target consonant. Consonant identification was measured for seven different signal-to-noise ratios (SNRs) including +9, +6, +3, 0, -3, -6, and -9 dB SNR, allowing a close examination of the contribution of

FIG. 1. Block diagram shows how the acoustic signals for the FM condition are encoded in the auditory nerve. See text for details.

TABLE 2
Subject demographics

Subject	Age (years)	Duration ^a of hearing loss (years)	Duration of CI use (years)	Etiology	Implant type	Experiment participated ^b
S12 (L)	55	0	5	Connexin 26	MedEl Combi40+	1
S12 (R)	55	0	8	Connexin 26	MedEl Combi40+	1
S48	72	10	6	Unknown	HiRes90K	1,2,3,4
S52	81	0	5	Noise exposure	HiRes90K	1,2,3,4
S71	73	15	3.5	Genetic	HiRes90K	1
S72 (L)	34	0	3.5	Atypical Meniere's	HiRes90K	1,2,3,4
S72 (R)	34	0	6	Atypical Meniere's	HiRes90K	1,2,3,4
S80	61	2	2	Otosclerosis	HiRes90K	1,2,3
S84	48	26	2.5	Genetic	HiRes90K	1,2,3,4
S89	63	2	1.3	Genetic	Nucleus 5	1,2,5
S91 (L)	24	0	3	EVA	Nucleus Freedom	1,2,5
S91 (R)	24	0	9	EVA	Nucleus Freedom	1,2,5
S94	82	10	4.5	Unknown	Nucleus Freedom	1,2,5
S106	66	0	4	Unknown	Nucleus Freedom	1,2,5
S109	79	5	2.5	Noise exposure	Nucleus 5	1,2,5
S110 (L)	48	7	8	Hereditary	HiRes90K	1,2,3,4
S110 (R)	48	7	17	Hereditary	HiRes90K	1,2,3,4

All CI subjects were mapped and tested with a CIS strategy.

EVA: enlarged vestibular aqueduct

^aDuration of severe to profound hearing loss is based on subjects' self-report of the number of years they were unable to understand people on the telephone prior to implantation.

^b1: Effect of background noise. 2: Effect of reverberation. 3: Effect of amplitude compression. 4: Effect of IDR. 5: Effect of number of channels

recovered AM cues to speech intelligibility as a function of SNR. The noise masker was first added to bisyllables at an appropriate SNR; then, the mixture of the bisyllable and masker was processed with either the intact or FM speech processing. The target bisyllables were presented at 65 dB SPL. The subjects were instructed not to change the volume and sensitivity settings during the test. Figure 2 shows the waveform of the first vowel/a/of the/aba/utterance processed by the intact and FM speech processing in quiet and in the presence of background noise.

Results

Figure 3 shows the resulting electrode outputs for electrode 5 (assigned to the frequency channel between 1,387 and 1,958 Hz) in response to the bisyllable/aba/processed by the intact and FM speech processing. In a quiet situation, the electrode output shows a clear within-band AM pattern for the intact condition. With the addition of background noise (e.g., at an SNR of +9 dB), the AM fluctuations of the utterance/aba/are smeared. For the FM condition, the acoustic input sound has a flat temporal modulation (see Fig. 2), but as a result of the implant's filtering process, the gross AM cues are partially recovered in the electrode outputs.

Figure 4A shows speech scores for individual subjects as a function of SNR. A wide range of performance was observed at each SNR. At the level of individual subjects, speech scores for the FM and intact conditions strongly overlapped particularly at

low SNRs. With the FM information alone, the majority of subjects showed speech scores greater than chance. Figure 4B shows averaged scores across subjects as a function of SNR. Speech scores for both condition increased as SNR increased as expected. Speech scores for the intact and FM conditions were compared using a 2×7 repeated measures analysis of variance (ANOVA), with the main factors of two speech processing methods and seven SNR values, and the interaction between the two factors. Identification scores for both speech processing methods decreased as SNR decreased [$F(6,96)=52.8$, $p<0.001$]. The effect of speech processing on identification scores was significant [$F(1,16)=39.3$, $p=0.001$], and there was a significant interaction between the two factors [$F(6,96)=13.6$, $p<0.001$].

Overall, the intact score was greater than the FM score, but the difference in performance between the two condition decreased as SNR decreased. To systematically analyze this pattern, a post hoc paired samples *t* test was performed to compare all intact and FM speech scores in pairs for each SNR value. These multiple comparisons for the main factor of speech processing type within the same SNR value can identify the specific SNR values where speech scores for the intact and FM conditions differed significantly. For these multiple *t* tests, the criterion for statistical significance was divided by 7 (i.e., $\alpha=0.007$) to correct for the number of comparisons. In Figure 4B, the testing conditions, where the difference in speech scores for the intact and FM conditions was statistically

FIG. 2. Waveform of the first vowel /a/ of the /aba/ utterance for three different background noise conditions. The *left panel* shows the waveform processed by the intact processing. The *right panel* shows the waveform processed by the FM processing.

significant, are shown by the asterisks under the scores. This comparison analysis reveals that speech scores differed significantly between the intact and FM conditions at high SNRs (≥ 0 dB), but when the SNR was lower, speech scores became similar for the two conditions. It should be noted that speech scores for both intact and FM conditions at -9 dB SNR were significantly greater than chance level [one-sample t test, $t(16)=6.20$, $p<0.001$ for the intact condition; $t(16)=6.05$, $p<0.001$ for the FM condition]. The significant interaction between the speech processing type (intact vs. FM) and the level of the distortion (seven SNRs) along with the statistically similar performance at the maximum distortion level, support the second hypothesis formulated above (i.e., the use of recovered AM cues from speech FM information is a robust speech perception mechanism for CI users in very noisy listening environments).

The relationship between identification scores for the intact and FM conditions was examined to test the third hypothesis. In Figure 4C, speech scores for the intact conditions are plotted as a function of the speech scores for the FM condition. Speech scores at seven different SNRs are shown by different symbols. To analyze the relationship, Pearson correlation was com-

puted between the scores for the intact and FM conditions. For this computation, all data points in Figure 4C were included. A highly significant correlation was found between the intact and FM conditions ($r=0.83$, $p<0.001$), supporting the third hypothesis. This indicates that CI subjects who showed high speech scores for the intact condition were those who showed high speech scores for the FM condition as well.

Discussion

The Role of FM-to-AM Conversion for Speech Identification in Background Noise. With the latest implant devices, CI users often achieve high levels of open-set speech identification in a quiet listening condition (for a recent review, see Wilson and Dorman 2008). However, speech identification performance decreases markedly for CI users as the level of background noise increases. A number of previous studies have attempted to clarify the origin of this abnormal susceptibility to noise. From the acoustic point of view, the presence of background noise distorts the representation of AM and FM cues for target speech signals (e.g., Steeneken and Houtgast 1980; Dubbelboer and Houtgast 2007). More specifically, the presence of background noise has the effect of reducing

FIG. 3. The electrode outputs in response to the utterance /aba/ for three different background noise conditions. The *left panel* shows the resulting outputs when the waveform /aba/ was processed by the intact processing. The *right panel* shows the resulting outputs when the waveform /aba/ was processed by the FM processing. Duration of 1.25 s is shown. Advanced Bionics Platinum sound processor was programmed with a CIS map that was used for Subject 72's left ear. Outputs

for electrode 5 were recorded from an "implant-in-a-box" (Advanced Bionics Corp). Electrode 5 was assigned to the frequency range of 1,387–1,958 Hz. AGC was turned on and the IDR was set to 60 dB. For noise conditions, the onset of noise was 0.5 ms before the onset of the target consonant, and the offset of noise was 50 ms after the end of the target consonant.

FIG. 4. **A** Consonant identification scores for individual subjects, corrected for chance and transformed to rationalized arcsine units (RAU), shown as a function of signal-to-noise ratio (SNR). The dotted red lines represent the scores for the intact condition. The solid blue lines show the scores for the FM condition. **B** Average consonant identification scores across subjects. The asterisks under the scores show the

significant difference in the performance ($p < 0.007$) by the intact and FM conditions based on multiple paired samples t tests. Error bars represent one standard error across subjects. **C** Scatter plot of consonant identification scores for individual subjects. Symbols represent SNR conditions. The Pearson correlation coefficient is included in the figure.

the AM depth of the target speech; moreover, if the level of background noise is higher than the level of target speech, the AM fluctuations of the background noise are dominant in the mixture waveform of the target speech and noise, and interfere with the auditory processing of the AM cues of the target speech.

From the signal processing stand point, it is generally accepted that two main factors contribute to reduced speech identification in noise for CI users: (1) limited spectral resolution including the negative effect of channel interaction (e.g., Fu et al. 1998; Friesen et al. 2001; Fu and Nogaki 2005; Won et al. 2007; Verschuur 2009), and (2) inability to deliver FM information through current CI speech processors (e.g., Smith et al. 2002; Qin and Oxenham 2003; Stickney et al. 2005; Zeng et al. 2005; Gnansia et al. 2010). The latter point should be stated more precisely because speech FM information can be transmitted in the form of recovered AM cues through the CI speech processor and CI users are able to exploit the recovered AM cues for speech identification (Won et al. 2012). The current experimental results confirm this mechanism by demonstrating that: (1) speech scores for the intact and FM conditions become very similar each other at low SNRs; (2) it was possible to achieve high levels of speech identification with speech FM information alone; (3) a significant correlation was found between the intact and FM conditions; and (4) speech scores for the FM condition were always significantly better than chance across all SNRs.

EXPERIMENT 2: EFFECTS OF REVERBERATION

Subjects

Eleven subjects participated in experiment 2. Three of the 11 subjects were bilateral users and they were tested with each implant separately. Therefore, data

for 14 implanted ears are reported. Subjects were fitted with an eight-channel CIS strategy. Their processor's AGC was active.

Simulations of Reverberation

The 16 bisyllables were used to evaluate the effects of reverberation. To simulate the effects of reverberation, bisyllables were convolved with a room impulse response (McGovern 2003). The room impulse response simulated a rectangular reverberant room with dimensions 20 m×20 m×5 m (length×width×height), reflecting a large size hall, with a microphone location of 6.5 m×6.5 m×1.3 m (which is equivalent to the location of head) and a sound source location of 10 m×10 m×1.3 m. The distance between the sound source and the listener was 5 m. Three reverberation times were implemented: 0, 0.6, and 1.0 s. Here, the reverberation time (RT_{60}) indicates the time required for the average sound in a room to decrease by 60 dB below its original level. The reverberation time for the testing booth was 0.1 s; hence, the effect of reverberation in the stimuli was dominated by the effect of simulation, not by the sound booth used for testing. Kokkinakis et al. (2011) demonstrated that CI users show significantly reduced sentence identification performance under a reverberation effect simulated with $RT_{60}=0.6$ and 1.0 s. Bisyllables were first simulated with the reverberation effect, and then submitted to the intact or FM speech processing. Stimuli were presented at 65 dB SPL.

Results

Figure 5 shows the resulting electrode outputs in response to the waveform of the/aba/utterance processed by the intact and the FM processing,

respectively. The left panels show that the original AM cues were significantly distorted when a RT_{60} of 1.0 s was used. The right panels reveal that, at least for $RT_{60}=0.6$, gross AM cues are partially recovered in the electrode pulse-train output.

Figure 6A shows speech scores for individual subjects as a function of RT_{60} . The intact condition is shown as the dotted red lines and the FM condition is shown as the solid blue lines. At each RT_{60} , the distribution of speech scores for the intact and FM conditions overlapped. Except for one subject for the intact condition at 1 s of RT_{60} , all subjects showed speech scores greater than chance for both conditions. Figure 6B shows averaged speech scores across subjects. Between $RT_{60}=0$ and 0.6 s, performance changed little, but between $RT_{60}=0.6$ and 1 s, scores dropped markedly, indicating that CI users are susceptible to the effect of reverberation even in a quiet listening condition. A 2×3 repeated measures ANOVA showed significant effects of speech processing type [$F(1,13)=55.9$, $p<0.001$], reverberation time [$F(2,26)=185.5$, $p<0.001$], and their interaction [$F(2,26)=84.9$, $p<0.001$].

As with experiment 1 (Fig. 4B), speech scores became similar as the level of distortion increased. The multiple comparison using a post hoc paired samples t test was done to confirm this pattern ($\alpha=0.017$ after the Bonferroni correction). When the reverberation effect was not strong ($RT_{60} \leq 0.6$), speech scores for the intact condition were statically higher than speech scores for the FM condition. In Figure 6B, the asterisks indicate such conditions. However, at $RT_{60}=1.0$, where the effect of reverberation was severe, speech scores for the intact and FM conditions became statistically similar. In addition, identification scores for the intact and FM conditions were above chance at the maximum distortion level ($RT_{60}=1.0$ s) [$t(13)=10.18$, $p<0.001$ for the intact condition; $t(13)=12.97$, $p<0.001$ for the FM condition]. These data support the second hypothesis that the recovered AM cues from acoustic FM information provide important speech information when CI users communicate in highly reverberant environments. Figure 6C shows a scatter plot of speech scores for the intact and FM conditions. Speech scores at three different reverberation times are shown by different symbols. For the data collected from 14 implanted ears, a significant correlation ($r=0.79$, $p<0.001$) was found between the intact and FM identification scores, supporting the third hypothesis.

Discussion

Factors Limiting Speech Identification Performance in Reverberant Environments and the Role of FM-to-AM Conversion in Such Acoustic Conditions. Even in quiet, understanding speech could be challenging for CI users if they have a conversation in reverberant spaces

such as classrooms. The acceptable limit of RT_{60} for classrooms by the ANSI standard is 0.6 s (ANSI 2002), at which our CI subjects showed no deterioration compared to the performance at $RT_{60}=0$. However, RT_{60} values for churches and museums generally exceed 1 s and concert halls often show RT_{60} values greater than 1.5 or 2.0 s (White and White 1980). In general, if the space is surrounded by materials that are highly reflective (e.g., concrete, tile, wood flooring, bricks, and glass), RT_{60} values are high. The effects of reverberation on speech identification for CI users has recently received extensive attention (e.g., Poissant et al. 2006; Whitmal and Poissant 2009; Kokkinakis and Loizou 2011; Kokkinakis et al. 2011; Zheng et al. 2011; Tillery et al. 2012). In the present study, when the effect of reverberation was introduced, speech identification performance remained constant until $RT_{60}=0.6$, and then decreased significantly at $RT_{60}=1.0$. This negative effect of reverberation is largely consistent with those previous studies, although the effect size is different across studies due to the use of different speech materials and testing conditions. These previous studies along with the present study suggest three possible mechanisms for the effects of reverberation on speech recognition in CI users.

First, limited spectral resolution degrades speech identification performance for CI users in reverberant listening environments. Previous studies by Poissant et al. (2006) and Whitmal and Poissant (2009) controlled the spectral resolution of normal-hearing listeners and showed a systematic degradation in speech identification as the number of vocoder channels decreased.

Second, reverberation smears out the AM cues of the target speech (see Brown and Palomaki 2006, for a review). When reverberation is created, a large number of reflected sound waveforms are built up and combined with the original sound wave; the reflected waveforms mask the segments of the original sound wave, and thus, the AM cues of the target speech are distorted. As a result, CI users get confused when important speech cues such as formants are presented (Kokkinakis and Loizou 2011; Kokkinakis et al. 2011). Compared to normal-hearing listeners, CI users have been shown to be highly susceptible to the effect of reverberation because their primary source of speech information is based on AM cues (Poissant et al. 2006; Kokkinakis et al. 2011).

Third, the degradation of speech identification for CI users in reverberant environments is related to the limited use of FM cues in speech signals. Tillery et al. (2012) tested the idea that residual low-frequency hearing would help CI users to understand speech in a reverberant environment. Using simulations of electric-acoustic stimulation, Tillery et al. demon-

FIG. 5. The electrode outputs in response to the utterance /aba/ for three different reverberation times. The *left* and *right* panels show the resulting outputs for the intact and FM speech processing, respectively. Duration of 1.35 s is shown. Electrode 5 was used for recording, which was assigned to the frequency range of 1,387–1,958 Hz. The operation of AGC was turned on and the IDR was set to 60 dB.

strated that FM cues in the low-frequency region improved speech intelligibility in reverberant environments. However, it should be noted that the mechanisms used to encode FM cues differ for the study by Tillery et al. and for the present study. In Tillery et al., normal-hearing listeners had access to both the purely temporal information (i.e., phase locked information conveyed by the firing patterns of auditory nerve fibers) as well as the recovered AM information below 500 Hz. Moreover, additional spectral information in the low-frequency region was also available for these normal-hearing subjects listening to the simulations of the electro-acoustic stimulation. However, only the recovered AM information was available for CI users in the present study. When speech was degraded by severe reverberation ($RT_{60}=1.0$ in Fig. 6A and B), speech identification scores for the intact and FM processing became similar; and at the same time, the FM speech scores were significantly better than chance. In addition, high levels of speech identification for the FM condition were observed for some individual subjects. Altogether, it is evident that the FM-to-AM conversion plays an important role for better speech identification outcomes in reverberant environments for CI users.

EXPERIMENT 3: EFFECTS OF AMPLITUDE COMPRESSION

People with normal-hearing have a wide dynamic range of about 120 dB (Moore 2012), but CI users have an extremely narrow dynamic range of only 10–20 dB with electric stimulation of the auditory nerve (Bilger 1977; Fourcin et al. 1979). The dynamic range of typical speech sounds is between 30 and 60 dB (Zeng et al. 2002). The speech processor must fit speech sounds into this narrow dynamic range for CI users. To do this, for all clinically available CIs, two forms of amplitude compression are employed. The first form, often called as automatic gain control (AGC), manages the overall input sound level. The AGC adjusts the microphone gain dynamically as the level of the input sound varies. The gain adjustment happens during a predetermined time window, which is often referred to as compression speed. The second step occurs later in the speech coding process. After the speech processor detects AM information, the AM signals in each channel get compressed and mapped to electric amplitudes. The compression processing is required to compensate for reduced dynamic range in CI users, but a consequence of amplitude compression is signal distortion. More importantly, amplitude

FIG. 6. **A** Consonant identification scores for individual subjects as a function of reverberation time (RT_{60}). The *dotted red* lines represent the scores for the intact condition. The *solid blue* lines show the scores for the FM condition. **B** Averaged consonant identification scores across subjects. The *asterisks* under the scores show the significant difference in the performance ($p<0.017$) by the intact and

FM conditions based on multiple paired samples *t* tests. *Error bars* represent one standard error across subjects. **C** Scatter plot of consonant identification scores for individual subjects. *Symbols* represent three different reverberation time conditions. The Pearson correlation coefficient is shown.

compression has its greatest distorting effect on AM cues of the acoustic speech signals (e.g., Plomp 1988; Jenstad and Souza 2005, 2007; Souza et al. 2006). Therefore, amplitude compression is expected to reduce the AM fluctuations of the acoustic signals, and subsequently, the AM cues in the electrode outputs may be diminished. Experiment 3 is designed to evaluate the effect of amplitude compression.

Subjects

Six subjects were tested and two of them were bilateral implant users. Therefore, a total of 8 implanted ears were tested. Subjects were fitted with an eight-channel CIS strategy. AGC was turned off to avoid any confound effect resulting from the interaction between the operation of AGC implemented in the speech processor and the compression effects simulated in the acoustic stimuli. The volume and sensitivity settings were set to 12 o'clock and the subjects were instructed not to change these settings during the test.

CNC Word Identification Test

Consonant-nucleus vowel-consonant (CNC) words were used to evaluate the effects of amplitude compression. Our pilot study showed that CNC word identification was more sensitive than consonant identification in showing the effect of amplitude compression. Open-set CNC monosyllabic words (Peterson and Lehiste 1962) were presented through a loudspeaker. A CNC word list was randomly chosen out of ten lists for each subject. The subjects were instructed to type the word that they heard and the experimenter scored performance offline. Each CNC word consists of three phonemes and a total phoneme score was calculated on a phoneme-by-phoneme basis after 50 presentations of CNC words (i.e., 150 phonemes). The phonemic scoring was used because it provides a more precise measure of the correct reception of the acoustic cues of speech than the whole-word scoring (Gelfand 2009).

Simulations of Amplitude Compression

CNC words were processed offline with compression simulation software (GennEM Version 1.0; Armstrong 1997). The GennEM software is a single-channel compression simulator that operates in the same manner as compression circuits in hearing aids. The compression threshold was fixed at 45 dB SPL. Attack and release time was fixed at 6 and 12 ms, respectively. No gain was applied. Three different compression ratios (CRs) were tested: 1:1, 5:1, and 10:1. CNC words were first processed with the GennEM software to

simulate the effects of acoustic compression. Those signals were then subsequently processed using either the intact or FM speech processing. Stimuli were presented through a loud speaker. The overall presentation levels were 65, 53, and 50 dB SPL for a CR of 1:1, 5:1, and 10:1, respectively.

Results

Figure 7 shows the resulting electrode outputs in response to the waveform of the /name/ utterance processed by the intact and FM processing, respectively. When a CR of 1:1 was used, clear AM fluctuations were shown for both intact and FM conditions. The electrode outputs for CR = 5:1 and for CR = 10:1 indicate that higher CRs reduced the overall pulse amplitudes as well as degraded the AM cues more. The outputs for the FM condition indicate that the AM cues are partially recovered as a result of implant filtering.

Figure 8A shows phonemic identification scores for individual subjects as a function of compression ratio (CR). Unlike the previous two experiments, there was less overlap between the distributions of identification scores for intact and FM conditions. The previous two experiments used a closed-set bisyllable recognition test, whereas this experiment used an open-set CNC word recognition test. This methodological difference likely affects the pattern of the speech intelligibility for two different speech processing methods as a function of the level of the distortion. In addition, previous studies reported large training effects for FM speech (i.e., speech stimuli containing Hilbert phase information alone, equivalent to the FM condition in this paper) when normal-hearing subjects were repeatedly exposed to FM speech stimuli up to 15 (Hopkins et al. 2010) or 17 sessions (Lorenzi et al. 2006). In comparison, the CI subjects in the present study received fewer training sessions for the CNC words processed by the FM speech processing. It is possible that an improvement for the FM condition would have been observed if more training sessions were given to the CI subjects.

Figure 8B shows averaged phonemic identification scores across subjects. As CR increased, the identification scores decreased, indicating that amplitude compression distorts important speech cues for CI users. This was confirmed by a 2×3 repeated measures ANOVA with the main factors of two speech processing methods and 3 CRs. The effect of speech processing [$F(1,7)=88.8$, $p<0.001$] and CR [$F(2,14)=21.9$, $p<0.001$] was significant. Although the interaction between the two factors reached significance [$F(2,14)=5.8$, $p=0.014$], speech scores for the intact condition were significantly higher than speech scores for the FM condition (paired samples t test, $p<0.001$)

FIG. 7. The electrode outputs in response to a CNC word /name/ for three different compression ratios. The *left* and *right* panels show the resulting outputs for the intact and FM speech processing,

respectively. Duration of 0.62 s is shown. Electrode 5 was used. The operation of the AGC was disabled and the IDR was set to 60 dB.

at all three CR values. Thus the second hypothesis is not fully supported by experiment 3. Nevertheless, at CR = 10:1, speech scores for the FM condition were significantly greater than chance [$t(7)=5.48$, $p=0.001$]. Figure 8C shows the relationship between speech identification for the intact and FM conditions. CNC phoneme identification scores for three different CRs are shown by different symbols. A significant positive correlation was found between intact and FM speech identification scores ($r=0.81$, $p<0.001$), indicating that subjects with higher FM speech scores were those with higher intact speech scores. This supports the third hypothesis.

Discussion

Consequences of Amplitude Compression. Unlike the previous two experiments, experiment 3 did not support the second hypothesis. That is, at the maximum distortion level (here, CR = 10:1), speech scores for the intact condition were still significantly greater than speech scores for the FM condition. Given that the effect of acoustic amplitude compression is clearly observed in the speech scores for the intact condition, it is possible that the speech scores for the FM condition were far too low even at CR = 1:1 in the first place. Or, the data may simply suggest that as long as both AM and FM information are provided (i.e., intact condition), the neural representation of speech AM cues might be preserved fairly well at the level of the auditory nerve so CI subjects can still show a certain level of speech identification (greater than 40 %). However, it should be noted that the magnitude of change in speech scores from CR = 1:1 to CR = 10:1 was different for the intact and FM conditions. When CR changed from 1:1 to 10:1, the decrease in speech scores was larger in the intact condition (29.1 %) than in the FM condition (17.2 %). This difference in the magnitude of change possibly suggests that the FM speech information is more robust to the effect of acoustic

amplitude compression and the large drop in the intact condition may be due to the degradation of AM information. In addition, a significant correlation was found between the intact and FM conditions, supporting the third hypothesis. Taken together, amplitude compression indeed affects the transmission of AM and FM speech information for CI users, but the exact mechanism underlying the relationship between the effects of compression in the acoustic signals (or electric signals) and the neural representation of modulation remains to be determined.

EXPERIMENT 4: EFFECTS OF INPUT DYNAMIC RANGE (IDR) SETTINGS

Subjects

Five Advanced Bionics device users were tested with a laboratory Platinum Speech Processor. Two of the five subjects were bilateral implant users and each ear was tested independently. Hence, the data for seven implanted ears are reported. Subjects were fitted with an eight-channel CIS strategy.

Testing Procedure

All commercially available CIs employ a variable IDR although the design and its operation vary across different implant systems. The speech processor encodes the peak acoustic amplitude to be mapped to the most comfortable level. The IDR value indicates the range of input levels below this peak acoustic amplitude to be presented between the most comfortable level and threshold level. Soft portions of the acoustic sounds below the IDR are mapped to sub-threshold levels and subsequently, degradation of AM cues occurs. For the Advanced Bionics devices, encoding the range of acoustic input levels depends on interactions among multiple factors: sensitivity setting, the AGC, IDR setting, and the patient's

FIG. 8. **A** CNC phoneme scores for individual subjects as a function of compression ratio (CR). The dotted red lines represent the scores for the intact condition. The solid blue lines show the scores for the FM condition. **B** Averaged CNC phoneme scores across 8 implanted ears. The asterisks under the scores show the significant difference in the performance ($p < 0.016$) by the intact and FM

conditions based on multiple paired samples t tests. Error bars represent one standard error across subjects. **C** Scatter plot of CNC phoneme identification scores for individual subjects. Symbols represent different compression ratios. The Pearson correlation coefficient is shown.

electric dynamic range. Throughout the experiment, the sensitivity setting was set to 12 o'clock for all subjects and AGC was turned off to avoid any confounding effect. Four values of IDR were tested: 30, 40, 50, and 60 dB. Bisyllables were presented in quiet at 55 dB SPL. This level was chosen because Spahr et al. (2007) demonstrated that the effect of increasing IDR was most prominent on sentence identification presented at this soft level. The speech processor manipulation was achieved using SoundWave™ software.

Results

Figure 9 shows the resulting electrode outputs in response to the waveform of the/aba/utterance for four different IDR settings. When a smaller IDR was used, some low amplitude portions of the speech were mapped to sub-threshold levels, leading to a less detailed representation of the speech AM cues. This effect can be seen by comparing the outputs for an IDR of 30 vs. 60 dB. Figure 9 also indicates that AM cues are partially recovered following implant filtering for each IDR value.

Figure 10A shows speech scores for individual subjects as a function of IDR. The distribution of speech scores for the FM condition strongly overlapped with the distribution of speech scores for the intact condition. At 30 and 40 dB IDR, the maximum performance achieved by the intact and FM conditions appears to be quite similar. Except for one subject at 30 dB IDR, individual subjects' scores for the FM condition were greater than chance. Figure 10B shows averaged speech scores across subjects. Results of a repeated measures ANOVA, with speech processing and IDR setting as main factors, showed a significant overall increase in speech scores

from 30 to 60 dB of IDR [$F(3,18)=30.6$, $p < 0.001$] and a significant effect of speech processing method [$F(1,6)=37.3$, $p < 0.001$]. The interaction between the speech processing type and IDR setting reached significance [$F(3,18)=6.1$, $p = 0.005$].

As with experiments 1 and 2, the intact and FM conditions showed similar speech scores as the level of distortion increased. To confirm this pattern, multiple paired samples t tests were performed (α level = 0.0125 after the Bonferonni correction). For IDR=50 and 60 dB, speech scores were significantly higher for the intact condition than the FM condition, and this is indicated by the asterisks in Figure 10B. However, when speech AM cues were further degraded by smaller IDRs, speech scores did not differ between the intact and FM conditions ($p = 0.26$ and 0.04 for 30 and 40 dB IDR, respectively). For IDR=30 dB, identification scores were significantly higher than chance for the FM condition [$t(6)=5.0$, $p = 0.002$]. Thus, the second hypothesis is supported. Figure 10C shows the relationship between speech identification for the intact and FM conditions. Consonant identification scores for four different IDRs are shown by different symbols. A significant positive correlation ($r = 0.78$, $p < 0.001$) was found between the intact and FM conditions, supporting the third hypothesis.

Discussion

AM Degradations Caused by the Conversion of Acoustic to Electric Amplitudes. In the present study, IDR settings of 30, 40, 50, and 60 dB were tested using Advanced Bionics devices, which have a default IDR setting of 60 dB. Performance for the intact condition increased monotonically as the IDR increased from 30 dB to

FIG. 9. The electrode outputs in response to the utterance /aba/ for four different IDR settings. The *left* and *right* panels show the resulting outputs for the intact and FM speech processing, respectively. Pulse amplitudes

over duration of 0.88 s are shown. Outputs for electrode 5 are shown. The operation of AGC was disabled for these measurements.

60 dB (Fig. 10B), which is consistent with a previous report (Zeng et al. 2002). A change in the IDR does not change the acoustic peak amplitudes that are mapped to the subjects' most comfortable levels. Instead, the audibility of soft signals improved as the IDR increased from 30 to 60 dB. For example, if an IDR setting of X dB is used, the acoustic peak amplitudes are mapped to the electric current levels that evoke the CI user's most comfortable loudness. This acoustic peak amplitude serves as a reference level, and the speech processor takes a range of X dB of the acoustic inputs below the reference level. Thus, some 'portions' of the acoustic sounds are taken, compressed logarithmically, and mapped between the user's threshold and most comfortable levels (i.e., the electric dynamic range). Therefore, compared to the AM distortion caused by noise, reverberation or amplitude compression, the operation of IDR imposes a different pattern of AM distortion. That is, the low amplitude acoustic sounds are discarded first, and then the remaining portions are compressed. Of course, if a narrower IDR is used, the AM distortion gets worse. Even so, the speech score for the FM condition at an IDR of 30 dB was significantly greater than chance ($p=0.002$). In addition, at IDR of 30 and 40 dB, the speech scores for the FM condition were not statistically

different from the speech scores for the intact condition. These data suggest that FM recovered as AM cues may provide salient speech information when the original AM cues are degraded with smaller IDR values. This is further supported by the significant correlation found between speech identification scores for the intact and FM conditions ($r=0.78$, $p<0.001$).

EXPERIMENT 5: EFFECTS OF NUMBER OF CHANNELS

Subjects

Five Cochlear device users were tested. One of them was a bilateral user. Six implanted ears were tested with a CIS strategy using a laboratory Freedom speech processor. Throughout this experiment, IDR was set to a default value for Cochlear devices (30 dB) and the operation of AGC was turned on.

Testing Procedure

Three different numbers of channels were tested including 4, 8, and 12. The frequency-to-electrode allocation in maps was determined by the default

FIG. 10. **A** Consonant identification scores for individual subjects as a function of input dynamic range (IDR). The *dotted red lines* represent the scores for the intact condition. The *solid blue lines* show the scores for the FM condition. **B** Averaged consonant identification scores across subjects. The *asterisks* under the scores show the significant difference in the performance ($p<0.0125$) by the

intact and FM conditions based on multiple paired samples t tests. *Error bars* represent one standard error across subjects. **C** Scatter plot of consonant identification scores for individual subjects. *Symbols* represent four different IDR settings. The Pearson correlation coefficient is included in the figure.

settings used in the CIS strategy for each of these channel number conditions. As the number of channels varied, the total bandwidth of 188–7,938 Hz for the Freedom processor remained the same, and the frequency range associated with each electrode changed accordingly. The bandwidths of the filters become narrower as the number of channels increased. Therefore, when a wideband FM speech signal is passed through the speech processor, more AM recovery will occur as the number of channels increases. This point is tested here. The pulse rate per channel was held constant throughout the testing conditions. Consonant identification was used as a measure of speech intelligibility. Original bisyllables (without any acoustic manipulation) were processed with the intact and FM speech processing. The bisyllables were presented at 65 dB SPL. Among the three testing conditions, the least amount of AM recovery was expected with the 4 channel condition.

Results

Figure 11 shows the resulting electrode outputs in response to the waveform of the/aba/utterance for three different settings of number of channels. When 4 channels were used, the electrode output showed a clipped pattern of broadband AM (resulting from the limited IDR and fast-acting AGC of the Cochlear device). Likewise, there was little AM observed in the four-channel FM condition. However, as more channels were used, the broadband signal was filtered through narrower filters. As a result, more AM recovery occurred for the FM condition.

Figure 12A shows individual subjects' speech scores as a function of the number of channels. The distributions of FM and intact speech scores strongly overlapped. All individual subjects showed speech scores greater than chance even for the four-channel condition. Figure 12B shows averaged scores across subjects. For both intact and FM conditions, the speech scores leveled off at 8 channels, which was most likely due to the effect of channel interaction (Friesen et al. 2001). A two-way repeated-measures ANOVA was conducted with speech processing type and the number of channels as main factors. These analyses showed a significant main effect of speech processing [$F(1,5)=14.2$, $p=0.013$] and the number of channels [$F(2,10)=16.2$, $p<0.001$]. However, the interaction between the two factors did not reach significance at the 0.05 level [$F(2,10)=0.67$, $p=0.53$].

A post hoc multiple comparison t test was used to compare speech scores between the intact and FM conditions at each channel number. For this comparison, the significance level at 0.016 (0.05 divided by 3) was used after the Bonferroni correction. For 8 and 12 channels, identification scores for the intact

condition were significantly higher than for the FM condition [$t(5)=5.2$, $p=0.003$ for 8 channels; $t(5)=4.6$, $p=0.006$ for 12 channels]. This is indicated by the asterisks in Figure 12B. Visually, speech scores at four channels appears to be different, but the difference in speech scores between the intact and FM conditions was not significant [$t(5)=2.3$, $p=0.071$]. Note that it is the four-channel condition where more AM degradation occurs compared to the 8 or 12 channel conditions. Taken together, the interaction between the speech processing type (intact vs. FM) and the level of distortion was not significant, but the speech scores for the intact and FM conditions at 4 channels were statistically the same. Thus the second hypothesis was partly supported.

Figure 12C shows the scatter plot of speech scores for the intact and FM conditions. Speech scores for three different numbers of channels are shown by different symbols. Speech scores for the FM condition were significantly correlated with speech scores for the intact condition ($r=0.68$, $p=0.002$), supporting the third hypothesis.

Discussion

Interpretation of the Effects of the Number of Channels from the Perspective of FM-to-AM Conversion. CIs process sound through a set of bandpass filters which divide the acoustic waveform into a certain number of channels. The number of active electrodes can be changed through clinical mapping software. If a certain electrode evokes facial nerve stimulation, pain, dizziness, or non-auditory sensation, the electrode is typically turned off. When turning off electrodes, the frequency allocation for each electrode gets redistributed and the bandwidth for each electrode changes. The results of previous studies investigating the effects of number of channels were generally interpreted in terms of limited frequency resolution (e.g., Dorman et al. 1997; Loizou et al. 1999; Friesen et al. 2001) and limitations of multichannel stimulation caused by channel interaction (e.g., Bingabr et al. 2008; Strydom and Hanekom 2011). A narrower bandwidth is desirable (i.e., more processing channels) because a more accurate analysis of the frequency spectrum is made possible.

The current study examined the effect of number of channels with a different conceptual framework, focusing on temporal processing instead of spectral (i.e., spatially encoded) processing. That is, if the number of active electrodes is changed, the bandwidth of bandpass filters designated for each electrode will be changed; as a result, the degree of AM recovery will change. The results support this view. As more channels were used, AM cues recovered from

FIG. 11. The electrode outputs in response to the utterance /aba/ for three different numbers of channels. The *left* and *right* panels show the resulting outputs for the intact and FM speech processing, respectively. Pulse amplitudes over duration of 0.88 s are shown. A Cochlear Freedom speech processor was programmed with CIS maps that were used for Subject 109. An “implant-in-a-box” (Cochlear, Ltd) was used to measure the electrode outputs from each channel. Outputs for electrode #18 are shown. As the number

of active electrodes increased, the bandwidth for electrode #18 decreased. For example, the four-channel condition sets a frequency range of 188–1,063 Hz for electrode 18. For the 8- and 12-channel conditions, a frequency range of 688–1,063 Hz and 813–1,063 Hz is processed and presented through electrode 18, respectively. The AGC was turned on while making these measurements. IDR was set to 30 dB.

speech FM cues became more salient and the speech intelligibility for the FM condition improved.

GENERAL DISCUSSION AND CONCLUSIONS

A previous study (Won et al. 2012) demonstrated that, in a quiet listening condition, CI users can understand speech using AM cues recovered from speech FM. The present study investigated this mechanism for listening conditions where speech information was severely degraded. Degradations of speech information were achieved in five different ways. Addition of background noise, simulation of the effects of reverberation, and amplitude compression were three examples of acoustic manipulations used to degrade speech information. Changing the IDR settings and

number of channels were two examples of speech processor manipulations, which also degrade the transmission of speech cues. For these (adverse) listening conditions, speech identification was measured in CI users using speech stimuli that were processed to preserve FM speech cues only. The speech identification tests showed:

- (1) High levels of speech identification performance were observed in some CI users for the FM condition where speech identification was solely based on AM cues recovered from speech FM cues.
- (2) At the maximum distortion levels, speech scores for the FM condition were significantly greater than chance level for all five tests.
- (3) As the level of distortion increased, speech identification performance for the intact and FM conditions decreased.

FIG. 12. **A** Consonant identification scores for individual subjects as a function of number of channels. The *dotted red lines* represent the scores for the intact condition. The *solid blue lines* show the scores for the FM condition. **B** Averaged consonant identification scores across subjects. The *asterisks* under the scores show the significant difference in the performance ($p < 0.016$) by the intact and

FM conditions based on multiple paired samples *t* tests. *Error bars* represent one standard error across subjects. **C** Scatter plot of consonant identification scores for individual subjects. *Symbols* represent different conditions for the number of channels. The Pearson correlation coefficient is shown in the figure.

- (4) A significant interaction between the level of distortion and speech processing type (intact vs. FM) was shown for most tests.
- (5) Speech identification performance for the intact and FM conditions became similar for testing conditions where severe AM degradations were employed.
- (6) Significant correlations between speech identification performance for the intact and FM conditions were found for most tests.

Altogether, these results suggest that despite poorer-than-normal frequency selectivity, CI users are able to perceive and use AM cues recovered from the speech FM cues. Moreover, the ability of CI users to use these recovered AM cues plays an important role in clinical success for acoustic environments where original speech cues are severely distorted.

The present study employed the Hilbert transform for the FM speech processing. This method, however, amplifies low-level recording noise to the same level as the target speech (for detailed demonstration, see Hopkins et al. 2010). As a result, the amplified recording noise in the FM stimuli could have a negative effect on the speech intelligibility for the FM condition, whereas the intact condition does not present such effects. Without such negative effects of the amplified noise, CI subjects would have shown better performance for the FM condition. Nevertheless, the results showed that robust speech identification was possible with FM cues alone over a wide range of distortion levels.

The straightforward implication of the present study is to design and use narrower analysis filters in the speech processor to improve FM-to-AM conversion. A recent study by Drennan et al. (2013) has evaluated the effect of the analysis-filter shape on complex FM discrimination in recipients of the Nucleus devices. When the signals were passed through a model of normal cochlear filter (i.e., gammatone filters; Patterson et al. 1987), which is narrower in bandwidth than the analysis filters implemented in the clinical speech processor, CI subjects showed improved FM discrimination performance.

It is noteworthy that a measure of speech identification in quiet was used to evaluate the effect of reverberation, amplitude compression, IDR setting, and number of channels in the present study. However, in daily listening situations, CI users encounter multiple sources of envelope distortions. For example, CI users often find themselves surrounded by multiple people in a highly reverberant acoustic space. In addition, bisyllables were mainly used for the speech identification tests in the present study. Therefore future studies should investigate the role of FM-to-AM conversion for multiple sources of speech degradation using sentence materials.

ACKNOWLEDGEMENTS

We are grateful for the dedicated efforts of our CI subjects. This research was supported by NIH-NIDCD Grants R01-DC007525, P30-DC04661, and an educational fellowship from the Advanced Bionics Corporation. Hyun Joon Shim was supported by Eulji Medical Center, Korea. Christian Lorenzi was supported by a visiting scholarship awarded by the Virginia Merrill Bloedel Hearing Research Center. Christian Lorenzi was also supported by three grants from ANR: HEARFIN project, ANR-11-0001-02 PSL, and ANR-10-LABX-0087.

REFERENCES

- ANSI (2002) ANSI S12.60-2002: acoustical performance criteria, design requirements, and guidelines for schools. American National Standards Institute, New York
- ARMSTRONG S (1997) GennEm—single channel ACG emulator (version 1.0). Computer software, Burlington
- BILGER RC (1977) Electrical stimulation of the auditory nerve and auditory prostheses: a review of the literature. *Ann Otol Rhinol Laryngol* 86(Suppl 38):11–20
- BINGABR M, ESPINOZA-VARAS B, LOIZOU PC (2008) Simulating the effect of spread of excitation in cochlear implants. *Hear Res* 241:73–79
- BROWN GJ, PALOMAKI KJ (2006) Reverberation. In: Wang D, Brown GJ (eds) Computational auditory scene analysis: principles, algorithms, and applications, Chap.7. Wiley Interscience, Hoboken, pp 209–250
- DORMAN MF, DANKOWSKI K, McCANDLESS G, SMITH LM (1989) Consonant recognition as a function of the number of channels of stimulation by patients who use the Symbion cochlear implant. *Ear Hear* 10:288–291
- DORMAN MF, LOIZOU PC, 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
- DRENNAN WR, LONGNION JK, RUFFIN C, RUBINSTEIN JT (2008) Discrimination of Schroeder-phase harmonic complexes by normal-hearing and cochlear-implant listeners. *J Assoc Res Otolaryngol* 9:138–149
- DRENNAN WR, IVES DT, WON JH, NIE K, JAMEYSON EM, ANDERSON ES, RUBINSTEIN JT, LORENZI C (2013) Perception of frequency-modulation patterns based on recovered-envelope cues for cochlear implant users. Presented at the 36th Annual Midwinter Meeting of the Association of Research in Otolaryngology. Baltimore, MD, USA, February 16–20
- DUBBELBOER F, HOUTGAST T (2007) A detailed study on the effects of noise on speech intelligibility. *J Acoust Soc Am* 122:2865–2871
- FISHMAN K, SHANNON RV, SLATTERY WH (1997) Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *J Speech Hear Res* 40:1201–1215
- FOURCIN AJ, ROSEN SM, MOORE BCJ ET AL (1979) External electrical stimulation of the cochlea: clinical, psychophysical, speech perceptual and histological findings. *Br J Audiol* 13:85–107
- FRIESEN LM, SHANNON RV, BASKENT D, 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 QJ, NOGAKI G (2005) Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *J Assoc Res Otolaryngol* 6:19–27
- FU QJ, SHANNON RV, 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

- GELFAND SA (2009) Speech audiometry. In: *Essentials of audiology*, 3rd edn. Thieme, New York, p 256
- GHITZA O (2001) On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. *J Acoust Soc Am* 110:1628–1640
- GILBERT G, LORENZI C (2006) The ability of listeners to use recovered envelope cues from speech fine structure. *J Acoust Soc Am* 119:2438–2444
- GNANSIA D, PRESSNITZER D, PEAN V, MEYER B, LORENZI C (2010) Intelligibility of interrupted and interleaved speech in normal-hearing listeners and cochlear implantees. *Hear Res* 14:46–53
- HEINZ MG, SWAMINATHAN J (2009) Quantifying envelope and fine-structure coding in auditory nerve responses to Chimaeric speech. *J Assoc Res Otolaryngol* 10:407–423
- HOPKINS K, MOORE BCJ, STONE MA (2010) The effects of the addition of low-level, low-noise noise on the intelligibility of sentences processed to remove temporal envelope information. *J Acoust Soc Am* 128:2150–2161
- JENSTAD LM, SOUZA PE (2005) Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility. *J Speech Lang Hear Res* 48:651–667
- JENSTAD LM, SOUZA PE (2007) Temporal envelope changes of compression and speech rate: combined effects on recognition for older adults. *J Speech Lang Hear Res* 50:1123–1138
- KOKKINAKIS K, LOIZOU PC (2011) The impact of reverberant self masking and overlap- masking effects on speech intelligibility by cochlear implant listeners. *J Acoust Soc Am* 130:1099–1102
- KOKKINAKIS K, HAZRATI O, LOIZOU PC (2011) A channel-selection criterion for suppressing reverberation in cochlear implants. *J Acoust Soc Am* 129:3221–3232
- LOIZOU PC, DORMAN M, TU Z (1999) On the number of channels needed to understand speech. *J Acoust Soc Am* 106:2097–2103
- LORENZI C, GILBERT G, CARN H, GARNIER S, MOORE BCJ (2006) Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc Natl Acad Sci U S A* 103:18866–18869
- LORENZI C, WALLAERT N, GNANSIA D, LÉGER A, IVES DT, CHAYS A, GARNIER S, CAZALS Y (2012) Temporal-envelope reconstruction for hearing-impaired listeners. *J Assoc Res Otolaryngol* 13:853–865
- MCGOVERN SG (2003) A model for room acoustics. [computer program]. <http://www.sgm-audio.com/research/rir/rir.html>. Accessed 30 Jan 2014
- MOORE BCJ (2012) An introduction to the psychology of hearing, 6th edn. Emerald Group Publishing Limited, Bingley, p 145
- MOORE BCJ, SEK A (1996) Detection of frequency modulation at low modulation rates: evidence for a mechanism based on phase locking. *J Acoust Soc Am* 100:2320–2331
- PATTERSON RD, NIMMO-SMITH I, HOLDSWORTH J, RICE P (1987) An efficient auditory filterbank based on the gammatone function. In: *Paper Presented at a Meeting of the IOC Speech Group on Auditory Modeling at RSRE*, December 14–15
- PETERSON GE, LEHISTE I (1962) Revised CNC lists for auditory tests. *J Speech Hear Disord* 27:62–70
- PLOMP R (1988) The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *J Acoust Soc Am* 83:2322–2327
- POISSANT SF, WHITMAL NA, FREYMAN RL (2006) Effects of reverberation and masking on speech intelligibility in cochlear implant simulations. *J Acoust Soc Am* 119:1606–1615
- QIN MK, OXENHAM AJ (2003) Effects of simulated cochlear implant processing on speech reception in fluctuating maskers. *J Acoust Soc Am* 114:446–454
- SABERI K, HAFTER ER (1995) A common neural code for frequency and amplitude modulated sounds. *Nature* 374:537–539
- SHANNON RV, JENSVOLD A, PADILLA M, ROBERT ME, WANG X (1999) Consonant recordings for speech testing. *J Acoust Soc Am* 106:L71–L74
- SHEFT S, ARDOINT M, LORENZI C (2008) Speech identification based on temporal fine structure cues. *J Acoust Soc Am* 124:562–575
- SMITH ZM, DELGUTTE B, OXENHAM AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416:87–90
- SOUZA PE, JENSTAD LM, BOIKE KT (2006) Measuring the acoustic effects of compression amplification on speech in noise. *J Acoust Soc Am* 119:41–44
- SPAHR AJ, DORMAN MF, LOISELLE LH (2007) Performance of patients using different cochlear implant systems: effects of input dynamic range. *Ear Hear* 28:260–275
- STEENEKEN HJ, HOUTGAST T (1980) A physical method for measuring speech-transmission quality. *J Acoust Soc Am* 67:318–326
- STICKNEY GS, NIE K, ZENG F-G (2005) Contribution of frequency modulation to speech recognition in noise. *J Acoust Soc Am* 118:2412–2420
- STRYDOM T, HANEKOM JJ (2011) An analysis of the effects of electrical field interaction with an acoustic model of cochlear implants. *J Acoust Soc Am* 129:2213–2226
- STUDEBAKER GA (1985) A rationalized arcsine transform. *J Speech Hear Res* 28:455–462
- SWAMINATHAN J, HEINZ MG (2012) Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. *J Neurosci* 32:1747–1756
- TILLERY KH, BROWN CA, BACON SP (2012) Comparing the effects of reverberation and of noise on speech recognition in simulated electric-acoustic listening. *J Acoust Soc Am* 131:416–423
- TORICK EL, KELLER TB (1985) Improving the signal-to-noise ratio and coverage of FM stereophonic broadcasts. *J Audio Eng Soc* 33:938–943
- VERSCHUUR C (2009) Modeling the effect of channel number and interaction on consonant recognition in a cochlear implant peak-picking strategy. *J Acoust Soc Am* 125:1723–1736
- WHITE HE, WHITE DH (1980) *Physics and music: the science of musical sound*. Saunders College Publication, Philadelphia, p 375
- WHITMAL NA, POISSANT SF (2009) Effects of source-to-listener distance and masking on perception of cochlear implant processed speech in reverberant rooms. *J Acoust Soc Am* 126:2556–2569
- WILSON BS, DORMAN MF (2008) Cochlear implants: a remarkable past and a brilliant future. *Hear Res* 242:3–21
- WON JH, DRENNAN WR, RUBINSTEIN JT (2007) Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *J Assoc Res Otolaryngol* 8:384–392
- WON JH, LORENZI C, NIE K, LI X, JAMEYSON EM, DRENNAN WR, RUBINSTEIN JT (2012) The ability of cochlear implant users to use temporal envelope cues recovered from speech frequency modulation. *J Acoust Soc Am* 132:1113–1119
- ZENG FG (2004) Compression and cochlear implants. In: *Compression: from cochlear to cochlear implants*, vol 17. Springer Handbook of Auditory Research, New York, pp 184–220
- ZENG FG, GRANT G, NIPARKO J, GALVIN J, SHANNON R, OPIE J, SEGEL P (2002) Speech dynamic range and its effect on cochlear implant performance. *J Acoust Soc Am* 111:377–386
- ZENG FG, NIE K, STICKNEY GS, KONG YY, VONGPHOE M, BHARGAVE A, WEI C, CAO K (2005) Speech recognition with amplitude and frequency modulations. *Proc Natl Acad Sci U S A* 102:2293–2298
- ZHENG Y, KOEHNKE J, BESING J, SPITZER J (2011) Effects of noise and reverberation on virtual sound localization for listeners with bilateral cochlear implants. *Ear Hear* 32:569–572
- ZWICKER E (1962) Direct comparisons between the sensations produced by frequency modulation and amplitude modulation. *J Acoust Soc Am* 34:1425–1430