The ability of cochlear implant users to use temporal envelope cues recovered from speech frequency modulation^{a)}

Jong Ho Won^{b)}

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

Christian Lorenzi

Equipe Audition, CNRS, Universite Paris Descartes, Ecole normale superieure, Paris, 75005, France

Kaibao Nie

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

Xing Li

Department of Electrical Engineering, University of Washington, Seattle, Washington 98195

Elyse M. Jameyson, Ward R. Drennan, and Jay T. Rubinstein

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

(Received 30 July 2011; revised 30 April 2012; accepted 6 May 2012)

Previous studies have demonstrated that normal-hearing listeners can understand speech using the recovered “temporal envelopes,” i.e., amplitude modulation (AM) cues from frequency modulation (FM). This study evaluated this mechanism in cochlear implant (CI) users for consonant identification. Stimuli containing only FM cues were created using 1, 2, 4, and 8-band FM-vocoders to determine if consonant identification performance would improve as the recovered AM cues become more available. A consistent improvement was observed as the band number decreased from 8 to 1, supporting the hypothesis that (1) the CI sound processor generates recovered AM cues from broadband FM, and (2) CI users can use the recovered AM cues to recognize speech. The correlation between the intact and the recovered AM components at the output of the sound processor was also generally higher when the band number was low, supporting the consonant identification results. Moreover, CI subjects who were better at using recovered AM cues from broadband FM cues showed better identification performance with intact (unprocessed) speech stimuli. This suggests that speech perception performance variability in CI users may be partly caused by differences in their ability to use AM cues recovered from FM speech cues.

© 2012 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4726013]

PACS number(s): 43.71.Gv, 43.71.Es, 43.71.Ky [PBN]

Pages: 1113–1119

I. INTRODUCTION

The role of amplitude modulation (AM) and frequency modulation (FM) cues in speech perception has been extensively studied for normal-hearing (NH) listeners (e.g., Drullman, 1995; Shannon *et al.*, 1995; Loizou *et al.*, 1999; Smith *et al.*, 2002; Zeng *et al.*, 2005; Gilbert and Lorenzi, 2006, 2010; Sheft *et al.*, 2008; Hopkins *et al.*, 2010; Kates, 2011; Eaves *et al.*, 2011). In these studies, speech signals were passed through the “analysis” filter bank of a vocoder to simulate the outputs of the cochlear filters. For each subband, the Hilbert transform was then used to decompose the subband signal into its AM and FM components. The Hilbert AM (also called the acoustic “temporal envelope”) was computed by taking the magnitude of the analytic signal of that

band. The FM signal was derived as the cosine of the angle of the analytic signal for each subband. These vocoder studies showed that, with adequate training, NH listeners could achieve high and quite comparable levels of speech intelligibility with speech stimuli having acoustic Hilbert AM speech cues only or acoustic Hilbert FM speech cues only (stimuli hereafter referred to as AM- and FM-speech, respectively).

These vocoder studies suggested that speech FM cues may convey as much phonetic information as speech AM cues do. This is not surprising because the AM and FM speech components are related when the Hilbert transform is used for demodulation processing (for analytical analysis, see Voelcker, 1966; Rice, 1973; Logan, 1977; Papoulis, 1983). However, it is still unclear how FM cues contribute to speech intelligibility. It has been proposed that the perception of FM can be mediated by two distinct early auditory mechanisms: (1) a first mechanism based on phase-locking in auditory-nerve fibers, and (2) a second mechanism based on an AM- (or envelope-) reconstruction process (e.g.,

^{a)}Portions of this work were presented at the Conference on Implantable Auditory Prostheses, Pacific Grove, CA, July 24–29, 2011.

^{b)}Author to whom correspondence should be addressed. Electronic mail: jhwon15@gmail.com

Moore and Sek, 1996; for a recent review, see Ernst and Moore, 2010).

Consistent with the first view, electrophysiological and modeling work demonstrated that the Hilbert FM information of FM-vocoded speech evokes “neural temporal fine-structure” cues (i.e., phase-locked temporal patterns of firing) in the auditory nerve fibers (Heinz and Swaminathan, 2009). However, psychoacoustical, electrophysiological, and modeling studies also demonstrated that the narrow-band speech AM components can be reconstructed from the broadband Hilbert FM components as a result of cochlear filtering, which is consistent with the second view (Ghitza, 2001; Zeng *et al.*, 2004; Gilbert and Lorenzi, 2006; Sheft *et al.*, 2008; Heinz and Swaminathan, 2009; Ibrahim and Bruce, 2010; see also Apoux *et al.*, 2011 for a recent study conducted with non-speech stimuli). Thus, when only Hilbert-FM speech cues (i.e., FM-speech) are presented to NH listeners, the recovered AM cues may serve as a strong cue for speech identification.

Gilbert and Lorenzi (2006) evaluated the contribution of the recovered AM cues to FM-speech identification as a function of the analysis filter bandwidth. They created Hilbert FM-speech signals with different numbers of bands for the analysis filter (1, 2, 4, 8, and 16), spanning between 80 and 8020 Hz. Then, the FM-speech signals were passed through a bank of 30 gammachirp auditory filters. This second set of gammachirp filters served as “reconstruction filters” to recover the AM cues from the Hilbert-FM signals at the outputs of the auditory filters. These recovered AM cues were then used to modulate the amplitude of sinusoidal carriers at the center frequencies of the auditory filters. Their hypothesis was that as more AM reconstruction from FM occurs, the better speech intelligibility would result when a wider analysis filter bandwidth was used. NH listeners could achieve about 60% consonant identification based on those recovered AM cues only when a single, wide analysis band was used, then identification performance dropped significantly as the number of analysis bands increased (i.e., as the bandwidth of each analysis band decreased). When the bandwidths of the analysis filters was less than or equal to four times the bandwidths of normal auditory filters, the contribution of the recovered AM cues to consonant identification almost vanished for NH listeners.

In agreement with this finding, the correlation between the original and the recovered AM components computed at the output of a simulated auditory filter bank was generally higher when the bandwidths of the analysis bands were wide [see also Sheft *et al.* (2008) for a similar demonstration]. Still, physiological experiments (Heinz and Swaminathan, 2009) and simulations (Ibrahim and Bruce, 2010) indicated that the recovered AM cues were not totally abolished in the response of auditory-nerve fibers when the bandwidths of the analysis filters approached the bandwidths of normal auditory filters. Consequently, a likely interpretation of the results of Gilbert and Lorenzi (2006) is that while the recovered AM cues were not entirely abolished for their NH listeners, these cues were not sufficient to contribute to the observed intelligibility. This is also supported by the preliminary model predictions of Ibrahim and Bruce (2011).

To the best of our knowledge, the role of AM cues recovered from FM cues in speech perception has been only investigated in NH and hearing-impaired listeners (e.g., Lorenzi *et al.*, 2006, 2009; Ardoint *et al.*, 2010). The present study extends this mechanism to cochlear implant (CI) users. CI sound processing is similar to the Hilbert vocoder processing which preserves only AM cues. For example, the continuous interleaved sampling (CIS) strategy extracts the AM component for each subband and uses it to modulate the amplitude of a fixed-rate pulse train. Therefore, FM information is discarded during this processing, and neural (phase-locking) temporal fine structure cues cannot be evoked by FM in the auditory nerve fibers of CI users. However, there is a possibility that FM could be transmitted through a CI sound processor in the form of recovered AM cues as a result of the implant filtering process. Specifically, when speech signals containing only Hilbert FM cues are presented through the CI sound processor, speech AM cues would be reconstructed at the output of the sound processor filters and modulated with pulse trains. In this regard, the sound processor filters act as reconstruction filters. Previous work demonstrated that CI users can make use of the recovered AM cues to detect and discriminate the simple and complex FM patterns produced by Schroeder-phase stimuli or noise modulators applied to sine-wave carriers (e.g., Drennan *et al.*, 2008; Sheft *et al.*, 2010). We hypothesized that if CI listeners can use the recovered AM cues to identify FM-speech signals, speech perception performance would increase as more recovered AM cues become available for them. To test this hypothesis, consonant identification was measured for stimuli having intact information (i.e., AM + FM), only AM cues, or only FM cues. For the FM condition, the number of frequency bands for the analysis filter was varied (1, 2, 4, and 8) to evaluate the effect of recovered AM cues on consonant identification (hereafter referred to as FM-1band, FM-2bands, FM-4bands, and FM-8bands, respectively). It was also reasoned that when AM and FM cues are presented simultaneously, AM and FM speech cues may potentially affect each other because they would share a common (and unique) coding mechanism for CI users. Consistent with this idea, previous vocoder studies with NH listeners (e.g., Zeng *et al.*, 2005; Nie *et al.*, 2005; Stickney *et al.*, 2005, 2007) suggested that FM speech cues may aid speech identification when presented simultaneously with AM speech cues. Consonant identification was therefore measured in CI users using stimuli having either intact (i.e., AM + FM) or only AM cues in 8 analysis bands to determine if the FM cues influence the auditory processing of the original AM cues. To compare CI and NH listeners' performance as the degree of envelope reconstruction varies, NH listeners were tested with vocoder simulations, where the stimuli were processed in a comparable way to the CI users.

II. METHOD

A. Subjects

Seven CI users and five NH subjects participated. One CI user used two implants and was tested on each ear individually, so a total of eight implanted ears were tested. Table I

TABLE I. Subject demographics.

Subject	Age (yrs)	Gender	Duration of CI use (yrs)	Implant device	Etiology
S04	66	M	7	Nucleus 24	Autoimmune
S48	71	F	4	HiRes90K	Genetic
S52	80	M	3	HiRes90K	Noise exposure
S72 (Left)	32	F	1.5	HiRes90K	Atypical Meniere's
S72 (Right)	32	F	4	HiRes90K	Atypical Meniere's
S84	47	M	1	HiRes90K	Meniere's
S88	32	M	6	Nucleus 24	Genetic
S91	22	F	7	Nucleus 24	Enlarged vestibular aqueduct

shows relevant information for the CI users. All CI users were fitted with an 8-channel CIS map to control the variability in the sound coding strategy. All NH subjects had audiometric thresholds of 20 dB HL or less at octave frequencies between 250 and 8000 Hz. This study was approved by the University of Washington Institutional Review Board.

B. Vocoder processing

The vocoder processing method used in this study was similar to the method reported by [Gilbert and Lorenzi \(2006\)](#). A total of three different vocoder processing conditions were used. These three vocoder processings were used for both CI and NH subjects.

1. Intact condition

Original speech signals sampled at 44 100 Hz were passed through a bank of 8 analysis bandpass filters (32nd order finite impulse response filters). The cutoff frequencies for the analysis filters were matched to those in the 8-channel CIS map (see Table II). The subband signals were summed over the bands without further processing. Thus both AM and FM information were preserved in this condition.

2. AM condition

The same analysis filters from the Intact condition were used to filter the original speech signals for the AM condition, where only AM cues were preserved while FM cues were discarded. The use of matched filters between the vocoder analysis filters and the CI processor ensured no spectral warping or compression. The subband AM signals were extracted using the Hilbert transform followed by low-pass

filtering (6th order, zero-phase Butterworth filter with a cut-off frequency of 64 Hz). These subband AM signals were then used to modulate the amplitude of pure tones at frequencies corresponding to the center of the analysis bands with random starting phase. The modulated waves were summed over all analysis bands to create tone-vocoded stimuli (i.e., AM speech).

3. FM condition

The FM condition preserved only FM cues while discarding AM cues. Four different sets of analysis bands were used including 1, 2, 4, and 8 bands. Table II shows the cutoff frequencies for the four different band conditions. The Intact and AM conditions used the cutoff frequencies for the 8-band condition listed in Table II. For the FM condition, the original stimuli were passed through a bank of 1, 2, 4, or 8 analysis bandpass filters. For each subband, the Hilbert transform was used to decompose the signals. Then the Hilbert FM (i.e., the cosine of the angle of the analytic signal) in each band was multiplied by the root-mean-square power of the subband signal. These subband FM signals were summed over all analysis bands.

C. Acoustic simulation of an 8-channel CIS strategy for NH subjects

CI subjects were presented with the speech signals that were processed with the three vocoder strategies described above. As indicated above, CI subjects have a second set of filters in their sound processor, which act as reconstruction filters for the FM signals. They were mapped with an 8-channel CIS map, thus they had 8 bands of reconstruction

TABLE II. Cutoff frequencies for Advanced Bionics (AB) and Cochlear devices as a function of the number of bands. All AB device users used HiRes90K implants and all Cochlear device users used Nucleus 24 implants as listed in Table I.

CI devices	# of bands		Cutoff frequencies in Hz							
AB	1	250								8700
	2	250				1387				8700
	4	250		697		1387		2762		8700
	8	250	494	697	983	1387	1958	2762	3898	8700
Cochlear	1	188								7938
	2	188				1563				7938
	4	188		688		1563		3438		7938
	8	188	438	688	1063	1563	2313	3438	5188	7938

filters. For NH subjects, this CI sound processing was simulated using an acoustic simulation of an 8-channel CIS strategy of Advanced Bionics (AB) devices. This way, NH subjects also had 8 bands of reconstruction filters as CI subjects. In collaboration with Advanced Bionics Corporation, a custom MATLAB program was created to implement a pre-emphasis filter, band-pass filters, and envelope detection processing that are employed by the actual CI sound processor.

A more specific description of the acoustic simulation of 8-channel CIS processing follows: The input speech signal was digitally sampled at 17 400 Hz. The signal was passed through a pre-emphasis filter and subsequently filtered into 8 channels using an array of infinite impulse response, 6th-order Butterworth filters. The frequency allocations for the 8 channels are equivalent to the frequencies for the AB 8-band condition listed in Table II. Threshold and most comfortable levels for each channel were set to 25 and 75 dB(A), respectively. The input dynamic range was set to 50 dB. Next, the envelope of each channel was computed on a frame-by-frame basis every 21 samples over 1.2 msec (21/17 400) by performing the half-wave rectification and averaging the outputs of each filter. The output of each channel was used to modulate a sine wave with frequency equal to the center frequency of the corresponding channel. Each modulated tone was then filtered with the initial bandpass filters. Finally, the modulated tones were summed together to create sine wave vocoded stimuli.

D. Pulse-train output analysis

The pulse-train outputs of CI processors were examined using an AB sound processing electrodiagram simulation implementing an 8-channel CIS map. A generic CIS map was used in the electrodiagram simulation, where the processing parameters were specified in such a way that the resulting electric signals were representative of the whole CI population. For example, threshold and most comfortable levels were set to 100 and 1000 μ A, respectively, for all 8 channels. Trains of biphasic pulses were computed for each speech token processed in all six conditions (Intact, AM, FM-8bands, FM-4bands, FM-2bands, and FM-1band). For each processing condition, the pulse amplitudes at eight electrodes were calculated. At each electrode, the pulse amplitudes were low-pass filtered (cutoff frequency = 300 Hz, 18 dB/octave rolloff) to extract the amplitude-modulation patterns. The modulation pattern at each electrode for the Intact condition was compared to the modulation patterns for the other five conditions. A Pearson correlation coefficient was computed for each token and an average across the 16 tokens was taken for each electrode.

E. Speech material and testing procedure

Sixteen bisyllables (vowel-consonant-vowels, VCV) read by a female speaker were used in a context of /aCa/ (/aba/, /ada/, /afa/, /aga/, /a3a/, /aka/, /ala/, /ama/, /ana/, /apa/, /ara/, /asa/, /a3a/, /ata/, /ava/, /aza/) (Shannon *et al.*, 1999). Before the actual testing, the subjects listened to each consonant more than three times for all processing conditions. All

testing sessions were conducted in a double-walled sound-proof booth. For CI subjects, the stimuli were presented in free field in random order at 75 dB(A) through a loudspeaker positioned at 0° azimuth and 0° elevation. CI subjects sat at 1-m from the loudspeaker, and were asked to face it during the course of the experiment. For NH subjects, the stimuli were presented monaurally to the right ear through an ER-3 insert earphone at 75 dB(A). Each consonant was presented three times for each trial. For the Intact and AM conditions, three testing trials were run. For the FM conditions, more than three trials (three to eight) were administered. Six testing conditions (Intact, AM, FM-8bands, FM-4bands, FM-2bands, and FM-1band) were run in random order. Percent correct scores 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. The RAU score for testing trials were averaged for each testing condition.

III. RESULTS

Figure 1 shows the mean RAUs (transformed percent correct, chance corrected) across 8 implanted ears and 5 NH subjects for each type of processing condition. For both CI and NH subjects, consonant identification scores were not significantly different for Intact and AM speech conditions (a paired *t*-test, $p = 0.1$ for CI subjects, $p = 0.15$ for NH subjects). For FM speech conditions, identification scores significantly decreased as the number of analysis filters increased from 1–8. This was true for both CI [$F(3,28) = 7.79$, $p = 0.001$] and NH subjects [$F(3,16) = 9.77$, $p = 0.001$], indicated by a one-way analysis of variance conducted on the four FM-speech data. On average, CI subjects reached an identification score of 42.3 RAU when FM-1band speech was used. For CI subjects, a *Post hoc* analysis using the Tukey test was performed on the FM-speech data with a Bonferroni correction value of $\alpha = 0.0125$ (0.05 divided by 4 demodulation band conditions) to determine if the FM-1band identification scores were significantly different from

FIG. 1. Mean identification performance (in chance-corrected RAU) across CI subjects (8 implanted ears, black bars) and 5 NH subjects (open bars) for six different processing conditions. Error bars show $\pm$ one standard error across subjects.

other conditions. This analysis revealed that FM-1band scores were significantly different from FM-4band ($p=0.005$) and FM-8band ($p=0.001$) scores. The FM-2band scores were not significantly different from any other conditions. The identification scores for FM-8band stimuli (13.2 RAU) were significantly different from 0 RAU (Wilcoxon t -test, $Z=-2.52$, $p=0.01$) in CI subjects.

An independent sample t -test was conducted to compare the consonant identification scores for CI and NH subjects. There was a significant difference in the consonant identification scores for the Intact condition [$t(11)=3.68$, $p=0.0036$] and for the AM condition [AM: $t(11)=2.86$, $p=0.016$]. The mean difference between CI and NH subjects for the FM-speech processing conditions did not reach statistical significance [FM-1band: $t(11)=1.72$; $p=0.11$; FM-2bands: $t(11)=1.52$, $p=0.16$; FM-4bands: $t(11)=0.33$, $p=0.75$; FM-8bands: $t(11)=0.68$, $p=0.51$].

For the FM-1band condition, consonant identification scores for CI subjects ranged from 26–74 RAU. Figure 2 shows the scatterplot of FM-1band scores and Intact condition scores for CI subjects. Despite the small number of subjects, a marginally significant correlation was found ($r=0.7$, $p=0.07$), suggesting that CI subjects who were better at using recovered AM cues from FM tended to also be better at identifying Intact speech. Previous work showed that differences in speech identification abilities of CI users are partly caused by differences in the ability to detect AM cues (e.g., Cazals *et al.*, 1994; Fu, 2002; Won *et al.*, 2011). The present work extends these initial studies by suggesting that differences in speech identification abilities among CI users may also be caused by differences in AM recovery from FM speech cues by the sound processor and/or differences in the ability of CI users to use recovered AM cues.

For each experimental condition and for each CI user, a confusion matrix was built from all VCV stimuli. A group matrix was then constructed by summing the individual confusion matrices. The specific reception of the individual pho-

FIG. 2. Relationship between FM-1band condition scores and Intact condition scores for CI users. Linear regression is represented by the solid line.

netic features of voicing, manner, and place of articulation was evaluated on the group matrices using an information-transmission analysis (Miller and Nicely, 1955). For this analysis, data collected from CI subjects wearing AB devices were used (total 5 implanted ears), so that the CI data can be more closely compared to 5 NH subjects, who were tested with an acoustic simulation of CIS strategy implemented by AB devices. The upper, middle, and bottom panels of Fig. 3 show the results for the reception of voicing, manner, and place of articulation, respectively. In each panel, the filled circles show the NH data and open triangles show the CI data as a function of the processing condition. NH subjects showed better reception of the three phonetic features than CI subjects, but the difference between NH and CI subjects varies with processing conditions. For example, when FM-4band and FM-8band signals were used, both NH and CI subjects were equally poor for all three phonetic features. NH and CI subjects showed similar reception of manner of articulation when FM-speech signals were presented, but the difference between NH and CI subjects was greater for Intact and AM speech conditions. Likewise, there was a big difference between NH and CI subjects for the reception of place when Intact and AM speech signals were used. Reception of voicing was generally greater than reception of manner and place for both NH and CI subjects.

FIG. 3. Phonetic feature reception for CI subjects (open triangles) and for 5 NH listeners (closed circles). The CI data are taken from five ears implanted with AB devices. The 5 NH subjects were tested with an acoustic simulation of an 8-channel CIS strategy of AB devices. The top, middle, and bottom panels show the results obtained for the reception of voicing, manner, and place of articulation, respectively.

FIG. 4. Mean correlation coefficients computed (across 16 consonants) between the pulse-trains' temporal envelopes of Intact stimuli and AM, FM-8 bands, FM-4 bands, FM-2 bands, and FM-1 band stimuli. Electrode 8 represents the highest frequency channel. Error bars show $\pm$ one standard error across sixteen consonants.

Figure 4 shows the correlation between the pulse train outputs of the CI processor for the Intact condition and for the other five conditions. A higher correlation indicates that more AM information was preserved at the processor outputs. As expected, the correlations were distinctly high for the AM-condition. Averaged across eight electrodes, the correlation was equal to 0.93, suggesting that Intact and AM conditions produced almost identical temporal envelopes at the pulse train outputs. For the FM-condition, the correlation generally increased as the number of demodulation bands decreased from 8 to 1, reaching relatively high values on some electrodes (e.g., correlation coefficient of 0.66 on electrode 4) in the FM-1band condition. Electrode 4 is tuned to mid-frequencies (around 1 kHz), where the frequency region encompasses the energy of the first and second formants of the vowel /a/ used in this study. This is consistent with the results of [Gilbert and Lorenzi \(2006\)](#) showing more AM recovery for channels where the amount of excitation is highest. This is also largely consistent with the consonant identification data in the present study; indeed, more AM recovery as well as better speech perception performance was obtained as a larger bandwidth was used for demodulation.

IV. DISCUSSION

The hypothesis that CI users can understand speech based on the recovered AM cues from FM processed by the sound processor is supported by these results. When speech sounds were demodulated into AM and FM components using a broadband analysis filter and the AM component was discarded (e.g., FM-1band), high levels of speech identification performance, voicing, and manner reception were observed in CI users. FM-1band speech scores were marginally correlated with Intact speech scores (Fig. 2), suggesting that the ability of CI subjects to use the recovered AM cues from speech FM signals relates to a variability in speech perception in CI users. When the mean difference between CI

and NH subjects was evaluated using an independent samples *t*-test, CI and NH subjects did not differ for the FM-speech conditions, suggesting that CI users could use the recovered envelope cues as well as NH subjects do. Consistent with the trend of consonant identification scores in CI subjects, the temporal envelope of the pulse-train outputs of CI processors showed a greater degree of similarity between the Intact and FM stimuli when a broader bandwidth was used for an analysis filter.

A similar pattern of results is observed in NH subjects. Consonant identification scores for NH subjects increased as the number of analysis filters is decreased (Fig. 1). The CI and NH data in the current study are largely consistent with NH data reported by [Gilbert and Lorenzi \(2006\)](#). Figure 3 compares the specific reception of voicing, manner, and place for NH and CI subjects. Again, a similar pattern is seen for both NH and CI listeners: Reception of all three phonetic features was best when the Intact and AM speech signals were used. When FM-speech signals were presented, reception of three phonetic features decreased as the number of analysis filters was increased. High voicing reception was observed for CI users with FM-speech signals. Interestingly, CI users nearly matched NH listeners for manner reception in the FM-1band condition. Together, these results demonstrate that CI users can use the AM cues recovered from the speech FM cues to achieve high levels of speech reception.

Finally, a comparable performance was obtained between the Intact- and AM-speech conditions (62 RAU vs 56 RAU) for CI subjects. This suggests that the recovered AM cues neither enhanced nor interfered with the auditory processing of the original AM cues for consonant identification when both AM and FM cues were presented to CI users. This was also true for NH subjects. Nevertheless, the present study demonstrated that the recovered AM cues contribute to robust speech perception when the speech AM cues are totally removed (i.e., FM-speech conditions). Further investigation is needed to determine if the recovered AM cues could contribute to robust speech perception when the speech AM cues are severely degraded by other acoustic distortions such as interfering background noise, peak clipping, amplitude compression, echoes, or reverberation. Recent psychophysical and modeling work by [Swinathan and Heinz \(2012\)](#) demonstrated that, for NH listeners, speech intelligibility in steady background noise can be predicted by the extent of AM recovery at the output of the peripheral auditory nerve. Altogether, these results suggest that the fidelity of AM recovery should be taken into account (and assessed) by CI manufacturers when evaluating current and future CI processors. Indeed, the fidelity of the AM recovery is likely to depend on specific aspects of the spectral decomposition employed by the CI processor. Future studies should investigate how to maximize the AM-recovery process by determining the optimal shapes and number of analysis filters. The AM-recovery process may also be an important factor in CI users when AM speech cues are degraded by the limited input dynamic range or automatic gain control applied by their sound processor before spectral decomposition.

V. CONCLUSIONS

The present study showed the following:

- (1) Temporal envelope recovery from broadband FM can occur through the CI sound processor.
- (2) When the recovered temporal envelopes were modulated with pulse trains, CI users could understand speech.
- (3) The present data also suggest that variability in speech identification abilities among CI users may be partly caused by differences in the degree of AM recovery from FM speech components provided by the sound processor, and/or differences in CI users' ability to use the recovered AM cues from FM speech cues for speech perception.

ACKNOWLEDGMENTS

We appreciate the dedicated efforts of our CI subjects. This study was supported by NIH Grant Nos. R01-DC007525, R01-DC010148, P30-DC04661, and T32-DC000033 and a fellowship from Advanced Bionics Corp. C.L. was supported by the Virginia Merrill Bloedel Traveling Scholar's Fund. We thank Leo Litvak of Advanced Bionics for providing the MATLAB code for the speech processing electrodogram simulation.

Apoux, F., Millman, R. E., Viemeister, N. F., Brown, C. A., and Bacon, S. P. (2011). "On the mechanisms involved in the recovery of envelope information from temporal fine structure," *J. Acoust. Soc. Am.* **130**, 273–282.

Ardoint, M., Sheft, S., Fleuriot, P., Garnier, S., and Lorenzi, C. (2010). "Perception of temporal fine structure cues in speech with minimal envelope cues for listeners with mild-to-moderate hearing loss," *Int. J. Audiol.* **49**, 823–831.

Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (1994). "Low-pass filtering in amplitude-modulation detection associated with vowel and consonant identification in subjects with cochlear implants," *J. Acoust. Soc. Am.* **96**, 2048–2054.

Drennan, W. R., Longnion, J. K., Ruffin, C., and Rubinstein, J. T. (2008). "Discrimination of Schroeder-phase harmonic complexes by normal-hearing and cochlear-implant listeners," *J. Assoc. Res. Otolaryngol.* **9**, 138–149.

Drullman, R. (1995). "Temporal envelope and fine structure cues for speech intelligibility," *J. Acoust. Soc. Am.* **97**, 585–592.

Eaves, J. M., Kitterick, P. T., and Summerfield, A. Q. (2011). "Benefit of temporal fine structure to speech perception in noise measured with controlled temporal envelopes," *J. Acoust. Soc. Am.* **130**, 501–507.

Ernst, S. M., and Moore, B. C. J. (2010). "Mechanisms underlying the detection of frequency modulation," *J. Acoust. Soc. Am.* **128**, 3642–3648.

Fu, Q. J. (2002). "Temporal processing and speech recognition in cochlear implant users," *NeuroReport* **13**, 1635–1639.

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., and Lorenzi, C. (2006). "The ability of listeners to use recovered envelope cues from speech fine structure," *J. Acoust. Soc. Am.* **119**, 2438–2444.

Gilbert, G., and Lorenzi, C. (2010). "Role of spectral and temporal cues in restoring missing speech information," *J. Acoust. Soc. Am.* **128**, EL294–EL299.

Heinz, M. G., and 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, B. C. J., and Stone, M. A. (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.

Ibrahim, R., and Bruce, I. (2011). "The interaction of envelope and temporal fine structure cues in speech perception," *Presented at the 34th ARO Meeting*, Abstract #742.

Ibrahim, R. A., and Bruce, I. C. (2010). "Effects of peripheral tuning on the auditory nerve's representation of speech envelope and temporal fine structure cues," in *The Neurophysiological Bases of Auditory Perception*, edited by E. A. Lopez-Poveda, A. R. Palmer, and R. Meddis (Springer, New York), Chap. 40, pp. 429–438.

Kates, J. M. (2011). "Spectro-temporal envelope changes caused by temporal fine structure modifications," *J. Acoust. Soc. Am.* **129**, 3981–3990.

Logan, B. F. (1977). "Information in the zero crossings of bandpass signals," *Bell Syst. Tech. J.* **56**, 487–510.

Loizou, P. C., Dorman, M., and Tu, Z. (1999). "On the number of channels needed to understand speech," *J. Acoust. Soc. Am.* **106**, 2097–2103.

Lorenzi, C., Debruille, L., Garnier, S., Fleuriot, P., and Moore, B. C. J. (2009). "Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal," *J. Acoust. Soc. Am.* **125**, 27–30.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (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.

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

Moore, B. C. J., and 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.

Nie, K., Stickney, G. S., and Zeng, F. G. (2005). "Encoding frequency modulation to improve cochlear implant performance in noise," *IEEE Trans. Biomed. Eng.* **52**, 64–73.

Papoulis, A. (1983). "Random modulation: a review," *IEEE Trans. Acoust., Speech, Signal Process.* **31**, 96–105.

Rice, S. O. (1973). "Distortion produced by band limitation of an FM wave," *Bell Syst. Tech. J.* **52**, 605–626.

Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., and Wang, X. (1999). "Consonant recordings for speech testing," *J. Acoust. Soc. Am.* **106**, L71–L74.

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

Sheft, S., Ardoint, M., and Lorenzi, C. (2008). "Speech identification based on temporal fine structure cues," *J. Acoust. Soc. Am.* **124**, 562–575.

Sheft, S., Cheng, M. Y., McMullen, R., Farrell, C., Shafiro, V., and Lorenzi, C. (2010). "Stochastic FM discrimination by elderly listeners and cochlear-implant users," presented at the annual meeting of the American Auditory Society, Scottsdale, AZ.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). "Chimaeric sounds reveal dichotomies in auditory perception," *Nature (London)* **416**, 87–90.

Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F. G. (2007). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," *J. Acoust. Soc. Am.* **122**, 1069–1078.

Stickney, G. S., Nie, K., and Zeng, F. G. (2005). "Frequency modulation added to implant simulations improves speech recognition in noise," *J. Acoust. Soc. Am.* **118**, 2412–2420.

Studebaker, G. A. (1985). "A rationalized arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.

Swaminathan, J., Heinz, M. G. (2012). "Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise," *J. Neurosci.* **32**, 1747–1756.

Voelcker, H. B. (1966). "Toward a unified theory of modulation. Part I: Phase-envelope relationships," *Proc. IEEE* **54**, 340–353.

Won, J. H., Drennan, W. R., Nie, K., Jameyson, E. M., and Rubinstein, J. T. (2011). "Acoustic temporal modulation detection and speech perception in cochlear implant listeners," *J. Acoust. Soc. Am.* **130**, 376–388.

Zeng, F. G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y. Y., and Chen, H. (2004). "On the dichotomy in auditory perception between temporal envelope and fine structure cues," *J. Acoust. Soc. Am.* **116**, 1351–1354.

Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (2005). "Speech recognition with amplitude and frequency modulations," *Proc. Natl. Acad. Sci. USA* **102**, 2293–2298.