Hearing Research xxx (2017) 1–10

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

Spectro-temporal cues enhance modulation sensitivity in cochlear implant users

Yi Zheng [a, 1], Monty Escabí [b], Ruth Y. Litovsky [a, *]

[a] Waisman Center, University of Wisconsin Madison, 1500 Highland Avenue, Madison, WI, 53705, USA
[b] Biomedical Engineering, Electrical and Computer Engineering, University of Connecticut, 371 Fairfield Rd., U1157, Storrs, CT, 06269, USA

ARTICLE INFO

Article history:
Received 11 March 2017
Received in revised form
12 May 2017
Accepted 23 May 2017
Available online xxx

Keywords:
Cochlear implant
Spectro-temporal
Psychoacoustics
Auditory modulation detection

ABSTRACT

Although speech understanding is highly variable amongst cochlear implants (CIs) subjects, the remarkably high speech recognition performance of many CI users is unexpected and not well understood. Numerous factors, including neural health and degradation of the spectral information in the speech signal of CIs, likely contribute to speech understanding. We studied the ability to use spectro-temporal modulations, which may be critical for speech understanding and discrimination, and hypothesize that CI users adopt a different perceptual strategy than normal-hearing (NH) individuals, whereby they rely more heavily on joint spectro-temporal cues to enhance detection of auditory cues. Modulation detection sensitivity was studied in CI users and NH subjects using broadband "ripple" stimuli that were modulated spectrally, temporally, or jointly, i.e., spectro-temporally. The spectro-temporal modulation transfer functions of CI users and NH subjects was decomposed into spectral and temporal dimensions and compared to those subjects' spectral-only and temporal-only modulation transfer functions. In CI users, the joint spectro-temporal sensitivity was better than that predicted by spectral-only and temporal-only sensitivity, indicating a heightened spectro-temporal sensitivity. Such an enhancement through the combined integration of spectral and temporal cues was not observed in NH subjects. The unique use of spectro-temporal cues by CI patients can yield benefits for use of cues that are important for speech understanding. This finding has implications for developing sound processing strategies that may rely on joint spectro-temporal modulations to improve speech comprehension of CI users, and the findings of this study may be valuable for developing clinical assessment tools to optimize CI processor performance.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Humans communicate using sounds that are modulated in amplitude, across both frequency and time. These modulations are not mutually exclusive. Human speech is characterized by a complex and continuous pattern of jointly varying spectral and temporal modulations. Temporal modulation patterns carry important syllable boundary information, and spectral modulation patterns carry formant and pitch information; joint spectro-temporal patterns carry information about formant transitions (Liberman, 1996). The rates of temporal amplitude modulations in speech range from those found in syllables (2–5 Hz) and phonemes (15–30 Hz) to those of vocal fold oscillations (>100 Hz) (Rosen, 1992). Spectral modulation rates critical for vowel identification and speech comprehension range from1-4 cycles/kHz (for a center frequency of 500 Hz, 4 cycles/kHz = 2 cycles/octave) (Liu and Eddins, 2008), and rates for vocal gender identification occur at 3–7 cycles/kHz (Elliott and Theunissen, 2009). In normal hearing (NH) subjects, accurate perception of spectral, temporal and spectro-temporal modulations facilitates success on basic auditory tasks such as sound recognition and speech intelligibility (Bregman, 1990; Shannon et al., 1995; Elhilali et al., 2003; Woolley et al., 2005). Further, accurate speech perception depends on the integrity of neural mechanisms in the auditory periphery and central system. Sounds are filtered into frequency-specific "channels" by the auditory periphery, i.e., the cochlea and auditory nerve. In the central auditory system, areas including the auditory midbrain and cortex selectively respond to spectro-temporal auditory cues (Kowalski et al., 1996; Theunissen

* Corresponding author.
E-mail address: Litovsky@waisman.wisc.edu (R.Y. Litovsky).
[1] Currently at: Beijing Advanced Innovation Center for Future Education, Beijing Normal University, Beijing, 100875, China.

http://dx.doi.org/10.1016/j.heares.2017.05.009
0378-5955/© 2017 Elsevier B.V. All rights reserved.

et al., 2000; Escabí et al., 2003). Maintaining an intact peripheral representation is essential for forward-transfer of information to central auditory centers, which are ultimately responsible for speech perception.

An increasing number of patients with sensorineural deafness receive cochlear implants (CIs) with the purpose of improving everyday communication, which relies on accurate speech intelligibility. The CI is aimed at replacing the function of the cochlea, bypassing damaged sensory hair cells and electrically stimulating the auditory nerve (Wouters et al., 2015). CI users do show some sensitivity to spectral and temporal cues, thus gaining access to at least some aspects of the speech signal. However, variability amongst CI users in speech understanding is very high, such that some patients perform significantly worse than NH subjects while others exhibit near-normal speech comprehension (Peterson et al., 2010; Holden et al., 2013).

The highly variable performance may partly arise from adoption of different strategies for extracting temporal and spectral cue information. The way frequency and timing cues are perceptually integrated can be evaluated by measuring sensitivity to changes in combined spectro-temporal cues. Although spectro-temporal sensitivity, as it pertains to speech intelligibility, has been studied in NH subjects (Chi et al., 1999), spectro-temporal sensitivity in CI users remains unexplored. We hypothesize and demonstrate that, unlike NH subjects, CI users adopt a novel perceptual strategy in which they have heightened detection sensitivity to spectro-temporal modulations. Such an enhancement in the spectro-temporal modulation sensitivity could facilitate extraction and detection of auditory cues in spoken language through the degraded electrical stimulation that CIs provide.

2. Materials and methods

2.1. Subjects

Nine bilateral CI users (9 females) with Nucleus devices participated in the study (Table 1). Nine normal hearing subjects (4 males and 5 females), ranging in age from 20 to 34 years were tested in this study. All subjects had pure tone thresholds at or below 20 dB HL octave interval frequencies between 250 and 8000 Hz. Also, for each subject, the thresholds between the two ears differed by less than 15 dB at any tested frequency. They consented to participation in the study and were paid an hourly wage. This study was approved by the Health Sciences Institutional Review Board of the University of Wisconsin-Madison.

2.2. Stimulus

Fig. 1A illustrates a schematic of the auditory stimuli used in this study. The schematic represents spectrograms of spectral, temporal, and spectro-temporal modulated "moving ripple" sounds. Each moving ripple sound has specific modulation values for the spectral and temporal domains. For spectral-modulation-only sounds, the spectral ripples are static over time (Fig. 1A right column; vertical spacing of bars) while for temporal-modulation-only sounds, the ripples fluctuate dynamically overtime, but the modulations are synchronous across frequency channels (Fig. 1A, top row; horizontal spacing of bars). For joint spectro-temporal modulations, the spectral ripples fluctuate over time asynchronously across frequency channels, generating "moving ripples" where the spectral peaks shift in frequency over time (Fig. 1A). Thus, spectral-modulation-only and temporal-modulation-only sounds comprise one-dimensional modulated functions, whereas spectro-temporal

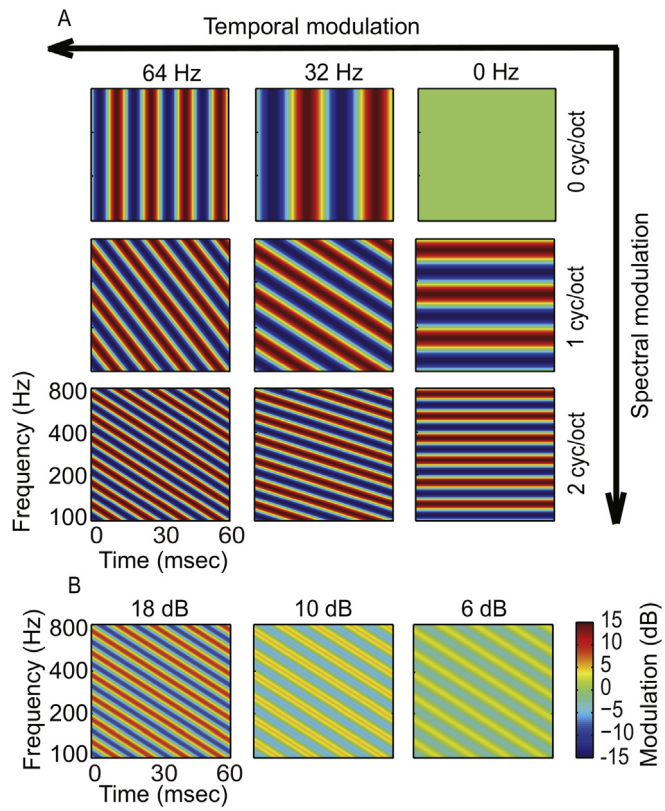

Fig. 1. Experimental stimuli. (A) Schematic diagram of spectro-temporal modulation envelope profiles. Each plot depicts the spectro-temporal modulation to which a particular spectral modulation frequency and temporal modulation frequency is tuned. The stimulus envelopes are modulated to spectral-only (vertical spacing of bars, right column), temporal-only (horizontal spacing of bars, top row), or spectro-temporal modulation (tilted spacing bars). The top-right plot represents the unmoduated signal. (B) Spectro-temporal modulation envelope with different peak-to-valley modulation contrasts. From left-to-right, the decreasing modulation contrast is demonstrated by the color contrast which transitions from sharp (red-to-blue) to shallow (yellow-to-cyan).

Table 1
Clinical etiology.

Subject	Age	Years CI Experience	Ear Tested	Internal Device	Processor
IAJ	57	15	R	CI24R	ESPrit 3G
IBO	48	5	R	Freedom Contour Advance	Freedom
IBP	62	8	R	CI24M	Freedom
IBQ	80	10	R	CI24R	Freedom
IBY	49	5	R	CI512	CP810
IBZ	45	6	R	Freedom Contour Advance	Freedom
ICA	53	11	R	CI24R	CP810
ICB	51	11	R	CI24R	CP810
ICF	71	3	L	CI512	CP810

Please cite this article in press as: Zheng, Y., et al., Spectro-temporal cues enhance modulation sensitivity in cochlear implant users, Hearing Research (2017), http://dx.doi.org/10.1016/j.heares.2017.05.009

modulated sounds comprise two-dimensional function. Fig. 1B demonstrates the effect of changing the modulation contrast (e.g., the modulation depth, defined as the peak-to-valley modulation amplitude in decibels) of a modulated sound on a given moving ripple sound. As modulation contrast decreases, the peak-to-valley amplitude of the sound envelope transitions from sharp (Fig. 1B left, red-to-blue of color contrast) to shallow (Fig. 1B right, yellow-to-cyan of color contrast). As in prior modulation detection studies (Litvak et al., 2007; Anderson et al., 2012), the modulation contrast is used to quantify the detection sensitivity of modulated signals.

The spectro-temporal envelope ($S_{dB}(t, X)$, in dB) of each moving ripple is defined by

$$S_{dB}(t,X) = \frac{M}{2}\cdot\sin(2\pi\omega t + 2\pi\Omega X + \Phi) - \frac{M}{2} \quad (1)$$

where M is the modulation contrast in dB, Ω is the spectral modulation frequency (i.e., the number of spectral peaks per octave, in units of cycles/octave), ω is the temporal modulation frequency (in Hz) and Φ is the spectral phase (randomly chosen). The variable $X = log_2(f/f_1)$ represents the frequency in octaves above the lowest frequency in the sound (f_1=100 Hz). Note that the spectral peaks in the moving ripple sound can move in an upward (low to high frequency) or downward (high to low frequency) direction over time. In this study, only the downward moving ripple conditions were tested.

For each modulated sound, the modulation contrast (M, units of dB) or, equivalently, the linear modulation depth ($1 - 10^{-M/20}$, units of linear amplitude; shown at 30 dB or equivalently 97% linear modulation depth ($1 - 10^{-30/20} = 0.97$) in Fig. 1A) (Escabí et al., 2003) is defined as the peak-to-valley modulation amplitude in decibels (Fig. 1B) or linear amplitude units, respectively. A logarithmic amplitude ripple spectrum (i.e., in dB) was used in this study for several reasons. First, perceptual rules for loudness and intensity discrimination follow logarithmic relationships such that decibels units provide a more faithful representation of the perceptual variables (Weber, 1834; Miller, 1947). Second, in terms of neural coding, natural sounds including speech have modulation amplitudes with approximately log-normal distribution (in dB units) that span several orders of magnitude and central auditory neurons can more faithfully encode spectro-temporal modulations with such characteristics (Escabí et al., 2003). Finally, the logarithmic amplitude scale for modulation detection and discrimination has also been widely used in cochlear implant users (Drennan et al., 2010; Won et al., 2011a). Each sound contained a broadband spectrum consisting of 948 tones in logarithmically spaced frequencies spanning 100–8000 Hz (~150 tones per octave), and which were modulated by the envelopes which were shown in Fig. 1B (for clarity the frequency axis in the plots span 100–800 Hz).

NH subjects were presented with stimuli through monaural earphones (ER-2) in the preferred ear as indicated by the listener (all nine NH subjects used the right ear). Similarly, CI users listened to the stimuli using their own CI processor through monaural presentation (directly to the audio port of the CI processor) using the ear with the longest auditory experience (first ear to be activated with a CI; 8/9 had the right ear). As such, the effect of CI processing on the spectral and temporal cues was evaluated in this experiment. All stimuli were presented at 60 dB SPL with ±4 dB of level rove for both the NH and CI groups; the levels were confirmed to be in the range of comfortable sound level for both the NH and CI subjects. All sounds were 500 ms duration, including 50-ms rise/fall ramps and the starting modulation phase (Φ) was randomized for each sound to minimize the possibility that users detected transient cues during the task. Thus, the possibility that intensity cues were available for detection was minimized using level roving and the randomized starting modulation phase.

2.3. Procedure

A three-interval, two-alternative forced choice procedure was used to measure modulation detection thresholds for each sound configuration. Subjects were required to distinguish a modulated stimulus (signal) from an unmodulated noise stimulus (standard). The first of the three intervals were the standard unmodulated stimulus presented as the "cue" interval. The second and third intervals contained the signal and another instance of the standard. The signal could occur in the second or third interval with equal a-priori probability (Litvak et al., 2007). On each trial, the three intervals were separated by 500 ms silent intervals. The timing of the three intervals was visually indicated by highlights of the interval buttons displayed on a computer screen. No correct-answer feedback was provided. A 2-down, 1-up adaptive procedure was used to measure the detection threshold of modulation contrast (Jones et al., 2013), converging on 70.7% correct (Levitt, 1971). The initial modulation contrast in each track was set to a level that was determined in extensive pilot testing to be above threshold (≥8 dB). The step-size of the adaptive track was 4 dB until the first reversal, which was reduced to 2 dB until the second reversal and then 0.5 dB thereafter. Each run terminated after a maximum of 9 reversals or 30 trials, whichever occurred first. For reliable and accurate threshold estimates of adaptive tracks, 30 trials are necessary and sufficient (Amitay et al., 2006). In the adaptive tracks the minimum of 4 reversals had occurred by trial 30 for all tests. Since other studies have shown that there is little loss in accuracy in averaging over two rather than four or more reversals (Baker and Rosen, 2001), data analysis was limited to last two reversals. For each adaptive track, the last two reversals, which excluded the first two reversals at large step sizes, were averaged to estimate the detection threshold.

The spectro-temporal modulated signals were tested at 5 spectral modulation rates: 0.5, 1, 2, 4, 8 cycles/octave, at each of 6 temporal modulation rates: 8, 16, 32, 64, 128, 256 Hz for NH subjects, and at 4 spectral modulation rates (0.5, 1, 2, 4 cycles/octave) at each of 5 temporal modulation rates (8, 16, 32, 64, 128 Hz) for CI users. The spectral-modulated-only signals were also tested at 6 spectral modulation rates (0.5, 1, 2, 4, 8, 16 cycles/octave) for NH subjects but 5 spectral modulation rates (0.5, 1, 2, 4, 8 cycles/octave) for CI users. The temporal-modulation-only signals were tested at 7 temporal modulation rates (8, 16, 32, 64, 128, 256, 512 Hz) for NH subjects, but 6 temporal modulation rates (8, 16, 32, 64, 128, 256 Hz) for CI users. The CI users were stimulated by a reduced modulation space because most CI users lost detection ability at high spectral and temporal modulation frequencies. The thresholds were measured for all combinations of spectro-temporal modulations and spectral-modulation-only and temporal-modulation-only, with the order of conditions and three adaptive tracks for each condition randomized. Nine CI users and nine NH subjects completed three adaptive tracks for each combination and all conditions on different days. Thresholds for each condition were computed by averaging three individual thresholds.

2.4. Analysis

2.4.1. Spectro-temporal modulation detection sensitivity of CI users and NH subjects

The modulation detection threshold (MDT) was defined as the modulation depth yielding 70.7% correct detection. Subjects identified the single interval containing a modulated signal (either spectral, temporal or spectro-temporal) from amongst three intervals, two of which contained an un-modulated sound (See

Section 2.2, 'Stimulus'). Detection thresholds at each condition (temporal and spectral modulation frequency) are evaluated with the modulation transfer function (MTF, detection thresholds as a function of temporal or spectral modulation frequency). Performance was tested in conditions with spectral-modulation-only and temporal-modulation-only, and also in a distinct configuration of combined spectro-temporal modulations. Thus, we could derive and compare the temporal (tMTF), spectral (sMTF) and spectro-temporal (stMTF) modulation transfer function for each subject.

2.4.2. Decomposing spectro-temporal sensitivity into spectral and temporal sensitivity

We determined whether sensitivity to spectro-temporal cues can be accounted for by the sensitivity to spectral-only and temporal-only cues. We approached this question from two separate points of view. First, we asked if the spectro-temporal detection sensitivity of CI and NH subjects can be approximated by separable functions of spectral and temporal sensitivity. That is, can the stMTF be decomposed and approximated as a product of a temporal and a spectral function? This first approach is analogous to the separability analysis performed by Chi et al. (1999) and answers the question of whether temporal and spectral cues are processed independently when they are delivered concurrently. We also asked whether the measured stMTF of subjects could be predicted by the measured spectral-only and temporal-only MTF (sMTF and tMTF). This second test is more stringent as it requires that the perceptual sensitivity to spectro-temporal cues is fully accounted by the sensitivities for each cue delivered in isolation (temporal and spectral only). Since the separability and prediction analysis performed below are based on linear operations, all the forthcoming analysis of the sMTF, tMTF and stMTF is carried out in the linear amplitude domain (i.e., modulation depth in units of %) even though, graphically, all the figures are shown in units of dB.

First, we tested whether the stMTF could be decomposed into independent spectral and temporal components analogous to the procedure performed by Chi et al. (1999). Separability of the spectro-temporal sensitivity was assessed by decomposing the stMTF into temporal and spectral functions using singular value decomposition (SVD)

$$stMTF(\omega,\Omega) = U\cdot S\cdot V^{*} = \sum_{i=1}^{L}\lambda_i\cdot stMTF_i(\omega,\Omega)$$

$$\mathbf{S} = diag(\lambda_1,\lambda_2,\ldots\ldots,\lambda_L),\quad \lambda_1\geq\lambda_2\geq\ldots\geq\lambda_L>0 \tag{2}$$

S is a diagonal matrix with real, non-negative elements, λ_i, in descending rank order according to energy; U and V are unitary orthogonal matrixes containing the two decomposed uMTF and vMTF functions that are used to approximate the *stMTF* component ($stMTF_i$); * denotes the Hermitian transpose. The i-th transfer function component, $stMTF_i(\omega,\Omega)$, is obtained by the product

$$stMTF_i(\omega,\Omega) = uMTF_i(\omega)\cdot\lambda_i\cdot vMTF_i(\Omega) \tag{3}$$

where $uMTF_i(\Omega)'$ and $vMTF_i(\omega)$ are the i-th unitary orthogonal vectors of U and V, respectively. According to the SVD procedure, every component $stMTF_i$ is separable in the spectral-temporal domain, although the entire measured stMTF may be non-separable.

The separability index (α_{sep}) is defined as (Singh and Theunissen, 2003)

$$\alpha_{sep} = \frac{\lambda_1}{\sum_{=1}^{L}\lambda_l} \tag{4}$$

where a value near one indicates that the stMTF is roughly separable and can be approximated as the product of a temporal and spectral function. Based on our calculations, only the first singular value is significantly different from zero ($p < 0.01$), thus the stMTF is predominantly separable, and we thus use the first component $stMTF_1$ to compare against spectral-only and temporal-only sensitivity.

Since the unitary $uMTF_1$ and $vMTF_1$ components derived from SVD are normalized such that $stMTF_1(\omega,\Omega) = uMTF_1(\omega)\times vMTF_1(\Omega)$ has unit power, each component was renormalized so that the composite MTF preserved the original total power. The decomposed temporal and spectral MTFs ($uMTF_d$ and $vMTF_d$) are expressed as

$$vMTF_d(\Omega) = \sqrt{\frac{E_{u1}}{E_{v1}}\cdot\lambda_1}\cdot vMTF_1(\Omega) \tag{5}$$

$$uMTF_d(\omega) = \sqrt{\frac{E_{v1}}{E_{u1}}\cdot\lambda_1}\cdot uMTF_1(\omega) \tag{6}$$

where $E_{u1} = rms(uMTF_1)$, $E_{v1} = rms(vMTF_1)$.

To identify potential differences in processing between temporal-only and spectral–only vs. spectro-temporal processing by CI and NH subjects, we compared the temporal-only and spectral-only MTF (tMTF and sMTF) against the first and dominant SVD component of the stMTF ($uMTF_d$ and $vMTF_d$).

2.4.3. Predicting spectro-temporal sensitivity from spectral- and temporal-only sensitivity

One hypothesis is that the spectro-temporal sensitivity can be fully accounted by the sensitivities for spectral and temporal cues delivered in isolation. Thus, using the spectral-only and temporal-only MTF (sMTF and tMTF), we first derived a predicted stMTF ($stMTF_p$) as the product of the temporal-only and spectral-only MTF in the linear amplitude domain: $stMTF_p = sMTF\times tMTF$. Any potential improvement of spectro-temporal sensitivity over spectral-only and temporal-only sensitivity was estimated as the difference of the predicted MTF and the measured spectro-temporal sensitivity (stMTF). The spectro-temporal gain is defined as $Gain = 20\cdot\log_{10}(stMTF) - 20\cdot\log_{10}(stMTFp)$ in units of dB. The overall estimated gain was then averaged across conditions and across subjects. Values near 0 dB indicate that spectro-temporal sensitivity is comparable to spectral-only and temporal-only sensitivity. Values > 0 dB indicate an enhancement while <0 dB indicate an impoverishment of spectro-temporal sensitivity over spectral-only and temporal-only sensitivity.

3. Results

3.1. CI electrode outputs to spectro-temporal modulated stimuli

Fig. 2 shows the electrode output recordings of the 22 electrodes through the Advanced Combinational Encoder (ACE) strategy for the STM stimuli shown in Fig. 1. Higher electrode numbers indicate low frequency channels. The data were collected by Ni-6343 DAQ (National instruments) and the sampling frequency was 22,727 Hz. First, there were fewer responses at the higher frequency channels compared to lower frequency channels. For the temporal-modulation-only sound, electrodes represented here contained periodic outputs that were synchronized across all electrodes

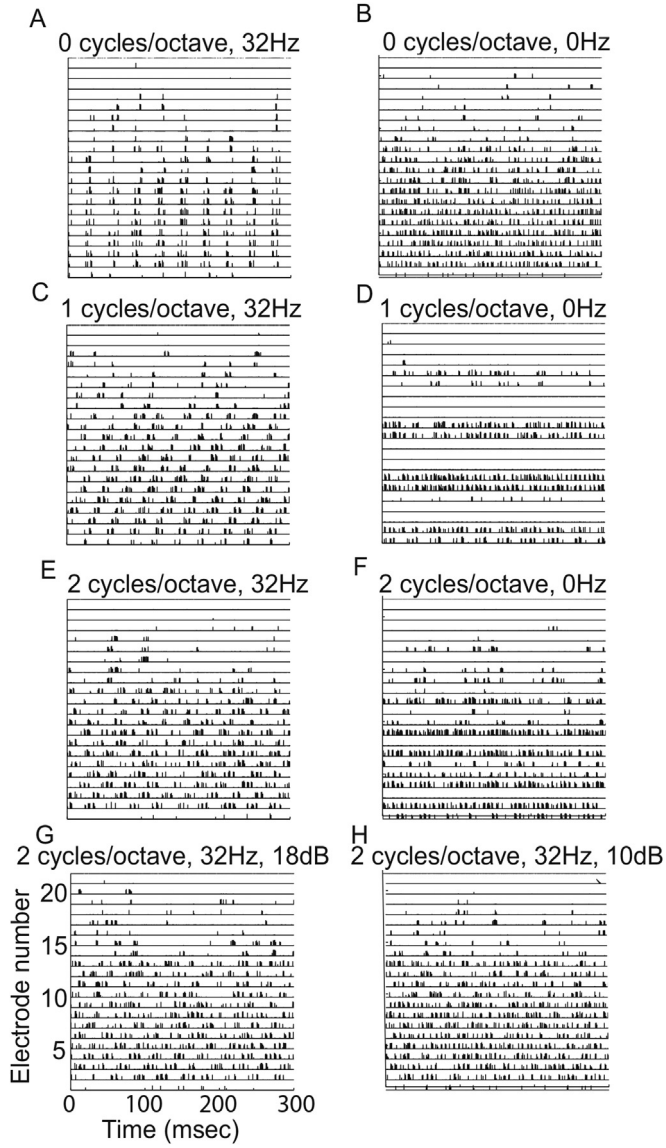

Fig. 2. The electrode outputs for various spectro-temporal modulated stimuli. (A–F) various spectral and temporal modulated stimuli with 30 dB modulation contrast. (G, H) 2 cycles/octave, 32 Hz spectro-temporal modulated stimuli with different modulation contrast.

(Fig. 2A). For the spectral modulation only sound, the electrode outputs displayed interleaved responses from low to high electrodes, but the responses lacked temporal modulation (Fig. 2D and E). Spectral modulation was also evident in the electrode outputs. For 1 cycles/octave stimuli, the outputs displayed interleaved responses every two or three electrodes from low to high electrodes (Fig. 2D). By comparison, a spectral modulation of 2 cycles/octave produced a denser electrode output with less distinct response peaks that were interleaved by one or two electrodes (Fig. 2F). For the spectro-temporal modulation stimulus, the peak of the envelope modulation occurred periodically in each channel, but the peaks shifted along time systematically across electrodes (Fig. 2C, E). Though the dynamic range of CI processors is different with that of normal hearing, these electrode output patterns reflect the spectrograms of spectral, temporal, and spectro-temporal modulated "moving ripple" sounds (Fig. 1). In addition, although responses were temporally modulated and synchronized at 32 Hz temporal modulation, response became progressively less synchronized with increase spectral modulation rate between 0 and 2 cycles/octave (Fig. 2A, C, E). Finally, the modulation contrast also affected the amount of output synchrony, such that decreasing the modulation contrast (Fig. 1B) decreased the overall output synchrony, thus producing a noticeably weaker envelope pattern (Fig. 2G and H).

3.2. Spectral, temporal and spectro-temporal modulation transfer functions for CI users and NH subjects

We measured listener's just-noticeable-difference for modulation contrast between a spectro-temporal modulation stimulus and an unmodulated stimulus, as the detection modulation threshold. Fig. 3 shows the averaged sMTFs, tMTFs, and stMTFs for CI users and NH subjects. The stMTFs generally exhibited a low-pass trend in both the spectral and temporal dimensions. Thresholds in both CI and NH subjects increased with increasing spectral modulation frequency or increasing temporal modulation frequency. NH subjects exhibited modulation threshold <10 dB to joint spectro-temporal modulations up to 128 Hz and 4 cycles/octave (Fig. 3B, D, F). CI subjects maintained reasonably low spectro-temporal

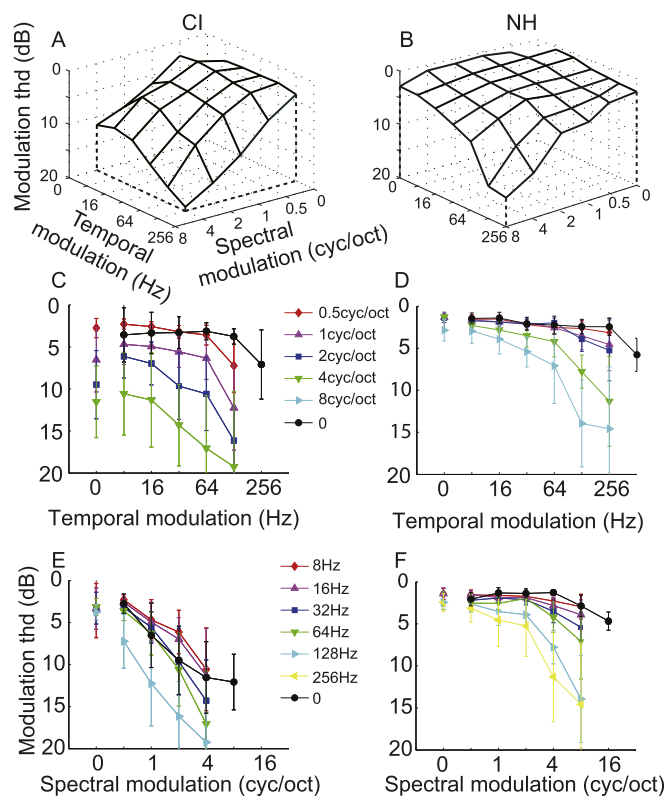

Fig. 3. Spectro-temporal modulation transfer functions. (A, B) 3D-plots of the averaged detection thresholds to spectro-temporal modulated sounds from 9 CI users and 9 NH subjects as a function of spectral modulation and temporal modulation frequencies. (C) Same spectro-temporal detection threshold data as A (color curves, different colors indicate different spectral modulation frequencies of the spectro-temporal modulations) from CI subjects plotted as a function of temporal modulation frequency, and compared to the averaged detection thresholds obtained from temporal modulation only stimuli (black curve). The error bars indicate the standard deviation error. (E) Same spectro-temporal detection threshold data as A (color curves, different colors indicate different temporal modulation frequencies of the spectro-temporal modulations) from CI users plotted as a function of spectral modulation frequency, and compared to the averaged detection thresholds to spectral modulation only stimuli (black curve). (D, F) Similar to (C, D) for NH subjects. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Zheng, Y., et al., Spectro-temporal cues enhance modulation sensitivity in cochlear implant users, Hearing Research (2017), http://dx.doi.org/10.1016/j.heares.2017.05.009

modulation detection thresholds<10 dB to temporal modulations up to 64 Hz and spectral modulations up to 2 cycles/octave (Fig. 3A, C, E). A two-way repeated-measures analysis of variance (ANOVA) on stMTF values indicated a significant effect of spectral modulation for both the CI and NH groups (NH: $F_{4,240} = 46.27$, $p = 10^{-29}$; CI:$F_{3,160} = 52.04$, $p = 10^{-23}$), and a significant effect of temporal modulation for both groups (NH:$F_{5,240} = 35.33$, $p = 10^{-27}$; CI: $F_{4,160} = 19.19$, $p = 10^{-13}$). A significant interaction of spectral and temporal modulation (i.e., effect of joint spectro-temporal modulation sensitivity) was seen for NH subjects ($F_{20,240} = 4.87$, $p = 10^{-10}$). For CI subjects, however, that interaction was not significant ($F_{12.160} = 0.72$, $p = 0.73$, see Fig. 3C, E). The post-hoc multiple group comparison tests were applied in two-way ANOVA. For CI users, all four spectral modulation groups (0.5, 1, 2, 4 cycles/octave) displayed significantly different each other at the level of 0.05, but for the five temporal modulation groups, the three low temporal modulation groups (8, 16, 32 Hz) are not significantly different each other, but significantly different with high temporal modulation groups (64, 128 Hz) and the two high temporal modulation groups are significantly different each other at the level of 0.05. For NH subjects spectral modulation groups, the three low spectral modulation groups (0.5, 1, 2 cycles/octave) are not significantly different each other but the two high modulation groups (4, 8 cycles/octave) are significantly different with all other groups at the level of 0.05; for temporal modulation groups, the three low temporal modulation groups (8, 16, 32 Hz) are not significantly different each other and 32 Hz is not significantly different with 64 Hz, but all the other pairs are significantly different at the level of 0.05.

The one-dimensional sMTF of the CI group exhibited a gradual decrease in sensitivity with increasing spectral modulation rate (Fig. 3E, black curve), as confirmed with results of a one-way ANOVA ($F_{5,48} = 10.39$, $p = 10^{-7}$). The sMTF of the NH group (Fig. 3F, black curve) showed a weak although significant band-pass characteristic with the maximum sensitivity between 1 and 3 cycles/octave, as confirmed with a one-way ANOVA ($F_{4,40} = 6.11$, $p = 0.0006$), and which is consistent with prior studies (Bernstein and Green, 1987; Summers and Leek, 1994; Chi et al., 1999; Saoji and Eddins, 2007). The tMTF showed low-pass characteristics in both the NH and CI groups (Fig. 3D and 3C, black curves). These finding were confirmed with one-way ANOVAs (NH: $F_{6,56} = 5.18$, $p = 0.0003$; CI: $F_{5,48} = 3.37$, $p = 0.01$).

Comparing the results from CI users to NH subjects, for spectro-temporal modulations, the CI users showed significantly higher thresholds than did NH subjects for high spectral (>1 octaves/cycle) and high temporal (>32 Hz) modulations (*t*-test, *all* $p < 0.005$). For spectral-only-modulations, the CI users displayed markedly higher thresholds than did NH subjects (*t*-test, $p < 0.005$ for spectral modulation frequencies> 0.5 octaves/cycle, $p = 0.24$ for spectral modulation frequency = 0.5 octaves/cycle, Fig. 3E and F, black curves). By comparison, although temporal-only-modulation thresholds for CI subjects were slightly lower than NH subjects, the differences were not significant (*t*-test, $p > 0.005$ for all temporal modulation frequencies, Fig. 3C and D, *black* curves). While these results indicate that degraded spectral representation is responsible for higher spectro-temporal thresholds, the findings do not address the exact mechanism by which spectral and temporal cues are combined and integrated.

3.3. Enhanced spectro-temporal sensitivity for CI users but not for NH subjects

We further asked whether the spectral-only and temporal-only sensitivity of NH and CI subjects could account for their combined spectro-temporal sensitivity. One possibility is that spectral-only and temporal-only sensitivities are independent of one another and they can therefore account for the combined spectro-temporal sensitivity (stMTF). Prior studies have shown that spectro-temporal modulation sensitivity (stMTF) can be approximated as a separable function in NH subjects (Chi et al., 1999). That is, the stMTF can be approximated as a product of a temporal-only and spectral-only function, indicating that the cues can be treated as independent when tested together. However, this study did not address whether the spectral-only and temporal-only sensitivity, with each of the cues delivered in isolation, accounts for the combined sensitivity to spectro-temporal modulations. Rather, we applied a more stringent test to determine whether temporal-only or spectral-only sensitivity account for the joint spectro-temporal sensitivity in NH and CI subjects. We first performed a singular value decomposition (SVD) of the stMTF. The spectro-temporal sensitivity from both the NH and CI subjects is approximately separable, since only the first singular value (λ_1) was significantly different from zero ($p < 0.01$, *t*-test using bootstrap) and dominant (separability index for CI, $\lambda_1/\sum\lambda_i = 0.91$; for NH, $\lambda_1/\sum\lambda_i = 0.86$; see Methods). This finding implies that the stMTF is approximately the product of a one-dimensional function of the spectral modulation frequency and a one-dimensional function of the temporal modulation frequency, as reported for NH subjects (Chi et al., 1999).

More importantly, the temporal and spectral components of the stMTF (obtained via SVD, $tMTF_d$ and $sMTF_d$, see Methods; Fig. 4, *solid* curves) were compared with the measured temporal-only and spectral-only sensitivities (tMTF and sMTF; Fig. 3, *dashed* curves). For the NH group (Fig. 3C and D), the $vMTF_d$ and $uMTF_d$ closely mirrored the measured tMTF (*t*-test, $p > 0.01$ for temporal modulation frequencies<256 Hz, $p = 0.003$ for temporal modulation frequency = 256 Hz; root mean square: 0.79 dB; Pearson's correlation: 0.83) and the measured sMTF (*t*-test, $p > 0.01$ for spectral modulation frequencies>1 cycles/octave, $p = 0.003$ for spectral modulation frequency = 0.5 cycles/octave; root mean square: 0.93 dB; Pearson's correlation: 0.62). This result confirms that the tMTF and sMTF measured with temporal-modulation-only and

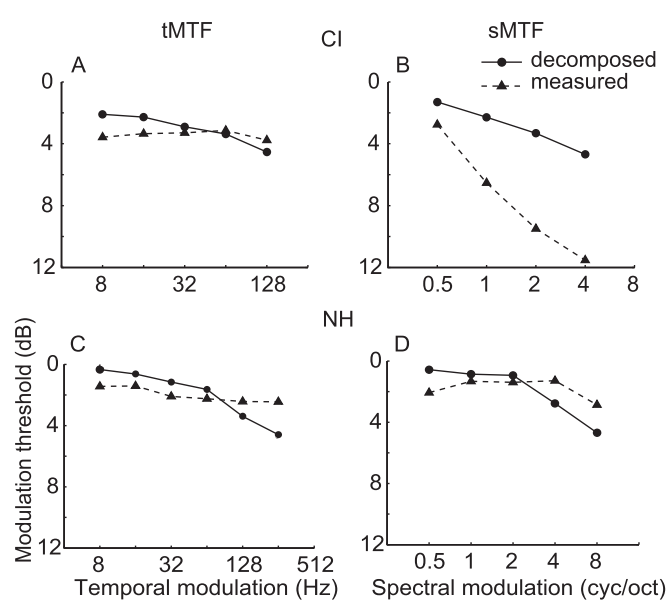

Fig. 4. Comparison of the temporal and spectral components of the stMTF with the measured sMTF and tMTF. Population averaged $uMTF_d$ and $vMTF_d$ (solid curves) and measured averaged sMTF and tMTF (dashed curves) for the CI subjects (A and B) and the NH subjects (C and D). A singular value decomposition procedure (SVD; see Methods) is used to decompose the measured averaged stMTF from the data in Fig. 2A and B.

Please cite this article in press as: Zheng, Y., et al., Spectro-temporal cues enhance modulation sensitivity in cochlear implant users, Hearing Research (2017), http://dx.doi.org/10.1016/j.heares.2017.05.009

spectral-modulation-only stimuli account for the joint spectro-temporal modulation sensitivity (stMTF) of NH subjects, because of the balanced and complete use of spectral and temporal cues. Surprisingly, this was not true for CI subjects. The spectral components derived from the stMTF exhibited lower thresholds than the isolated spectral sensitivity obtained with spectral-only sounds (Fig. 3B) and the enhancement was larger at the higher spectral frequencies (6.84 dB at 4cycles/octave) than at the lower spectral frequencies (1.46 dB at 0.5 cycles/octave). However, a similar effect was not observed for the temporal component of the stMTF (Fig. 3A, tMTF: *t*-test, $p > 0.01$ for all temporal modulation frequencies; root mean square difference: 0.91 dB; Pearson's correlation: 0.98; sMTF: *t*-test, $p < 0.01$ for all spectral modulation frequency; root mean square difference: 5.12 dB; Pearson's correlation: 0.37). This result indicates that CI subjects have a heightened sensitivity when listening to joint spectro-temporal modulations that is not accounted by their spectral-only and temporal-only sensitivity. This enhancement of the detection sensitivity through integration of spectro-temporal information was not observed in NH subjects. The significant larger enhancement over spectral-only sensitivity than over temporal-only sensitivity can be accounted for by the brain's compensatory mechanism in which the spectro-temporal sensitivity relies on relatively intact temporal and low spectral frequency cues to overcome reduced sensitivity to higher spectral frequency regions. The unbalanced use of spectral and temporal cues on spectro-temporal sensitivity might be caused by the limited spectral resolution of CI processing.

Results of SVD analysis for the individual CI users and NH subjects are provided in Fig. 5. CI users displayed large across-subject variability. Seven of the nine CI subjects exhibited different lower thresholds for spectro-temporal modulations than for the spectral-only and temporal-only modulation. Two (IAJ and IBO) subjects did not show the heightened spectro-temporal sensitivity. None of the NH subjects showed the heightened spectro-temporal sensitivity.

It is important to note that the separability demonstrated by the SVD analysis does not conflict with the result of the ANOVA yielding a two-way interaction. A high separability index (>0.8) from the SVD decomposition indicates that the separable spectro-temporal sensitivity accounts for most of the stMTF power (see Method); hence the joint spectro-temporal sensitivity can be roughly separated into two independent measures of sensitivity: temporal and spectral. In theory, a significant effect of spectro-temporal interaction remains possible even if it is weak. Therefore, the ANOVA examines the significance of interaction whereas SVD measured the size of the effect (larger separability index, smaller effect of interaction). A significant but small interaction could occur, which is the case of NH subjects in this study. CI users showed a small although non-significant interaction of the two dimensions.

To further confirm and quantify the enhanced spectro-temporal sensitivity we measured the detection gain afforded by spectro-temporal cues. Assuming that spectro-temporal sensitivity can be fully accounted by the sensitivities for spectral and temporal cues delivered in isolation, we expect a gain enhancement of 0 dB. The modulation gain enhancement is obtained as the difference in thresholds between the stMTF and the predicted stMTF ($stMTF_p$) from the isolated temporal-only and spectral-only MTFs ($stMTF_p = sMTF \times tMTF$, see Method). Fig. 6B demonstrates that the measured stMTF for the NH group did not have a significant improvement over the predicted $stMTF_p$ (Average Gain = 0.42 dB, see Method, *t*-test, $t_{29} = 0.75$, $p = 0.46$). By comparison, the CI group (Fig. 6A) exhibited a positive and significant improvement (Average Gain = 2.27 dB, *t*-test, $t_{19} = 3.24$, $p = 0.004$), most prominent for spectro-temporal modulations with low spectral (≤2cycles/octave) and low temporal (≤64 Hz) modulation frequencies. Unlike NH counterparts, CI users gained enhanced spectro-temporal sensitivity over spectral-only and temporal-only sensitivity.

4. Discussion

4.1. Spectro-temporal sensitivity of CI users

The main finding of this study is that the spectro-temporal

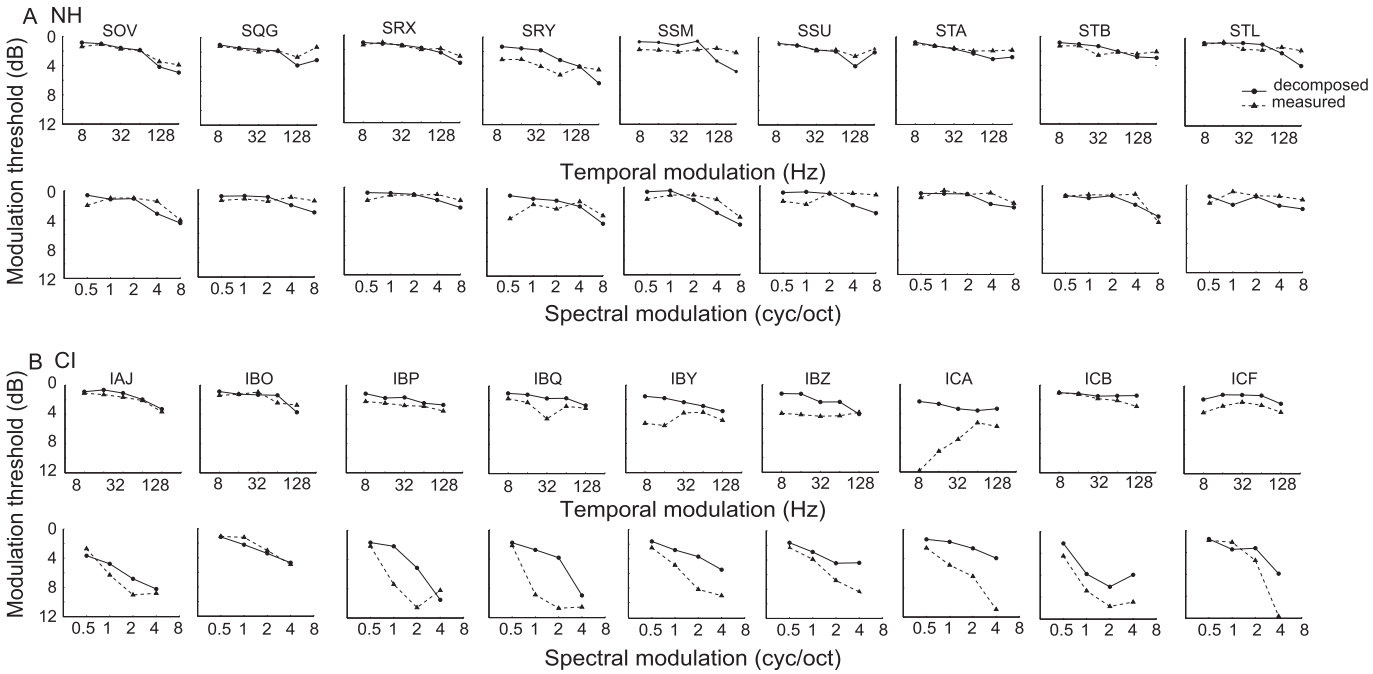

Fig. 5. Comparison of the temporal and spectral components of the stMTF with the measured sMTF and tMTF from individual subjects. $uMTF_d$ and $vMTF_d$(solid curves) and measured sMTF and tMTF (dashed curves) for nine NH subjects (A) and nine CI subjects (B).

Please cite this article in press as: Zheng, Y., et al., Spectro-temporal cues enhance modulation sensitivity in cochlear implant users, Hearing Research (2017), http://dx.doi.org/10.1016/j.heares.2017.05.009

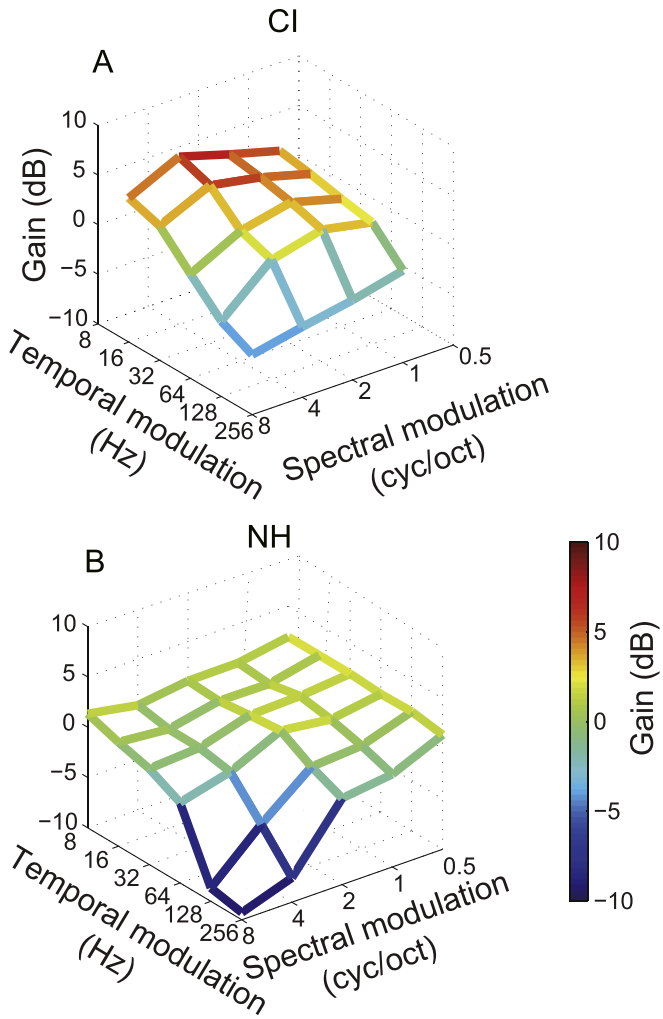

Fig. 6. Comparison of the predicted $stMTF_p$ to the measured stMTF. A $stMTF_p$ is constructed by a product of measured sMTF and tMTF. The subtraction of the predicted $stMTF_p$ from the measured stMTF is defined as the improvement of stMTF. Averaged improvement from 9 CI subjects (A) and 9 NH subjects (B) were plotted.

detection sensitivity of CI users can be separated into roughly independent components (Fig. 4) but cannot be predicted by the measured spectral- and temporal-modulation-only detection sensitivity (Fig. 6). This disparity is due to notable enhancement in spectro-temporal detection sensitivity over spectral-only or temporal-only detection sensitivity. In contrast, enhancement was not observed in NH subjects, because of the balanced and complete use of spectral and temporal cues on spectro-temporal detection. As CI users receive severely degraded spectral cues, and thus have more room for improvement in sensitivity, it is possible that mutual interaction between temporal and spectral dimensions may enhance spectro-temporal detection when the two cues are combined. The relative high sensitivity to temporal cue can compensate for the relative low sensitivity to spectral cue for spectro-temporal detection. In addition, the CI processor might contribute to better sensitivity in the spectro-temporal condition more so than to the spectral-only or temporal-only conditions. For instance, temporally-modulated spectral components that fall within the same analysis filter have the potential to produce more useable modulation cues than the non-modulated spectral components.

Although CI users exhibited poorer spectral modulation sensitivity than the NH counterparts, both groups had similar temporal modulation sensitivity. Poor sensitivity to high-rates of spectral modulation is likely affected by the known poor specificity in cochlear stimulation regarding location along the array of electrodes. The neurobiological factors mediating these effects are likely a combination of limited number of electrodes, amount of intra-cochlear current spread, and regions of neural death or atrophy (Horst, 1987; Henry and Turner, 2003; Won et al., 2011b; Azadpour and McKay, 2012; Jones et al., 2013).

Numerous studies on spectral-modulation-only detection in CI users suggest that the reduced upper cutoff frequency of the sMTF is caused by broadened spectral tuning (Horst, 1987) due to the diminished spectral resolution (Summers and Leek, 1994; Eddins and Bero, 2007; Saoji and Eddins, 2007). The frequency distribution tables in CI speech processors take the entire speech spectrum and allocate broad regions of the speech spectrum to 12–22 channels, each channel corresponding to one electrode in the cochlear array. Unlike the normal acoustic system where processing of speech through cochlear filters results in narrowly-tuned filters and fine-grained spectral resolution, the bandwidths of CI analysis filters are typically broader than one-fourth of octave on average, which explained our results from CI users that spectral modulation frequencies greater than 4 cycles/octave are not easily detected (Fig. 3E, *black* curve). Our results showed that NH subjects could detect spectral modulation out to the highest spectral modulation frequency tested (16 cycles/octave; Fig. 3F, *black* curve, threshold less than 5 dB). Anderson et al. (2012) observed the spectral detection sensitivity up to 60 cycles/octave, which is beyond the typical auditory filter bandwidths of one-sixth to one-tenth of an octave depending on the center frequency and the method of measurement (Glasberg and Moore, 1990; Oxenham and Shera, 2003). It is possible that interaction between closely spaced spectral peaks at high spectral modulation in the spectral-modulation-only signal produces temporal envelope fluctuations, or beats, which provide a viable temporal cue when compared with the unmodulated flat-spectrum noise signal in the detection task (Anderson et al., 2012).

As for temporal-modulation-only detection sensitivity, we found that CI users are able to follow temporal envelope fluctuation up to 300 Hz, near the limit observed in NH subjects (Shannon, 1992; Won et al., 2011a). A likely explanation for the fine temporal modulation sensitivity is that high temporal synchrony is maintained with electrical stimulation to the auditory nerve (van den Honert and Stypulkowski, 1987).

4.2. Spectro-temporal enhancement mechanisms for CI

In this study, all CI subjects used the CI device from Cochlear Corporation (Table 1). Speech processors are typically programmed with the Advanced Combinational Encoder (ACE) or Continuous Interleaved Sampling (CIS) strategies (Vandali et al., 2000), and today the majority of Nucleus users are fitted with the ACE strategy (Parkinson et al., 2002), which is based on a channel selection scheme. Stimulus pulses are delivered only to 8–10 electrodes out of 22 implanted electrodes corresponding to channels with highest amplitudes in each time frame. For a given spectral-modulation-only sound, the same electrodes would be stimulated in each time frame. However, for a dynamic spectro-temporal modulation stimulus, there is potential for different electrodes to be stimulated during each time frame, creating ongoing variance, or "jitter" in the spectral information conveyed. Furthermore, in CI users it is likely that some auditory neural substrates are degraded, due to auditory deprivation or other factors (Kral et al., 2006; Kral and Sharma, 2012). Thus, response patterns to spectral-modulation-only stimuli, which limit stimulation to particular electrodes (Fig. 2D, F), might not be as informative as when ACE-like strategies are used

with stimuli that carry spectro-temporal modulations, in which different electrodes are stimulated at different time frames (Fig. 2C, E). With spectro-temporal modulations, there might be release from spike rate adaptation patterns in the periods during which specific electrodes are not stimulated. Thus, the overall information transmitted for spectro-temporal modulation might be less prone to neural adaptation and thus could be transmitted to central auditory levels with higher-fidelity.

Another factor that may affect spectro-temporal sensitivity is CI multi-channel loudness summation. Previous studies suggested that MDTs are level-dependent (Donaldson and Viemeister, 2000; Galvin and Fu, 2005). Galvin (Galvin et al., 2014) measured single- and multi-channel amplitude modulation detection in CI users. They found that multi-channel MDTs were significantly better than MDTs with the best single channel when there was no level compensation for loudness summation and suggested that CI users combine envelope information across channels. This is in line with our finding that spectro-temporal enhancement could arise from cues resulting from multi-channel summation and spatio-temporal channel interactions of CI processors. In addition, there is the possibility that low-pass filter cutoff of the envelope filter used in the CI processors might influence the multi-channel interaction and modulation sensitivity (Fig. 2).

Physiological studies have demonstrated that spectro-temporal modulation tuning of midbrain inferior colliculus neurons (Slee and David, 2015) and cortical neurons (Fritz et al., 2003) are modified when an animal engages in an auditory detection task, producing changes that enhance discriminability of task-relevant sounds. Our findings, demonstrating the unique use of spectro-temporal cues in CI users, provide evidence that auditory plasticity and adaptation to CI processing strategies in adults who became deaf has the potential to enhance detection of cues that are involved in speech understanding. Future work on NH subjects' ability to adapt to CI simulations may be informative but would require extensive training and experience with the simulations to detect effects of auditory adaptation to the degraded cues. In addition, the healthy auditory system of NH subjects may not require the same adaptive strategies as the system of CI users in whom neural pathology may dominate aspects of the adoption of perceptual strategies.

Age-related changes in the central auditory system is another potential factor that may influence the enhanced spectro-temporal sensitivity. Evidence of age-related difficulties in the temporal processing of acoustic cues, independent of peripheral listening, has been previously observed (Phillips et al., 1994; Fitzgibbons and Gordon-Salant, 1995; Souza and Kitch, 2001). In this study, the age of CI users (aged 45–80 years) were significantly older than the NH subjects (aged 20–34 years), such that age-related changes in the peripheral and central auditory pathways could be a factor that impact perceptual strategy.

4.3. Implications for speech perception

Several studies have shown that spectral modulation tests (e.g., spectral modulation detection or and spectral ripple discrimination) can predict speech recognition outcomes in CI users (Throckmorton and Collins, 1999; Donaldson and Nelson, 2000; Henry et al., 2000, 2005; Litvak et al., 2007; Won et al., 2007; Saoji et al., 2009; Drennan et al., 2010; Anderson et al., 2011, 2012; Jones et al., 2013). One remaining question is whether changes in the intensity over time provide a cue that is extracted and perceptually utilized by CI users in the spectral-ripple discrimination test (Goupell et al., 2008). A second question is whether global spectral and temporal cues other than fine spectral detail contribute to speech understanding by CI users (Azadpour and McKay, 2012). Speech in quiet remains intelligible even with poor spectral resolution (Shannon et al., 1995), but higher spectral resolution is required for speech understanding in noisy and complex environment (Friesen et al., 2001). On the other hand, temporal sensitivity is critical for recognition of vowel and consonants (Cazals et al., 1994; Fu, 2002) as well as phonemes (Fu and Shannon, 2000; Xu et al., 2005). Previous studies showed that good temporal modulation detection performance for low modulation frequencies in CI users was significantly correlated with speech perception in CI users (Shannon, 1992; Cazals et al., 1994; Fu, 2002; Won et al., 2011a; Gnansia et al., 2014). Yet, neither of these measurements provides information about joint sensitivity to spectral and temporal information. In NH subjects spectro-temporal modulations are considered important for sentence comprehension and vocal gender identification (Elliott and Theunissen, 2009). In hearing impaired subjects spectro-temporal modulation sensitivity appears to be a predictor of speech intelligibility (Bernstein et al., 2013; Choi et al., 2016).

Future work should investigate the relationship of the non-speech spectro-temporal modulation sounds and speech intelligibility would be interesting, provided that the speech materials are selected appropriately. Given the broad range of speech testing materials that exist, a comprehensive study on this topic would be informative.

The current findings also indicate that future CI signal processing or (re)habilitation strategies can theoretically take advantage of the enhanced spectro-temporal modulation sensitivity in CI patients to potentially improve speech recognition and intelligibility. Sabin et al. (2012) found that in NH subjects, improved sensitivity to a specific spectro-temporal modulation stimulus cannot be influenced by training the corresponding stimuli that have either spectral-modulation-only or temporal-modulation-only. The results of our study thus suggest that there may be a benefit to using training strategies that focus on spectro-temporal sensitivity, rather than training on the cues in isolation. Training effects may thus have more robust benefits for speech understanding if combined spectro-temporal sounds are used, rather than single-cue stimuli. Finally, the findings indicate that dual spectro-temporal processing strategies for CI speech processors could theoretically exploit the enhancements offered by spectral-temporal cues to improve speech intelligibility.

Acknowledgements

This work was supported by NIH-NIDCD (grant R01DC003083, Litovsky) and also in part by a core grant to the Waisman Center from the NIH-NICHD (P30 HD03352). The authors would like to thank Jacob Bergal for assistance with generating the electrode output recordings, and Dr. Leslie Bernstein for comments on an earlier version of this manuscript. The authors declare no competing financial interests. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Amitay, S., Irwin, A., Hawkey, D.J., Cowan, J.A., Moore, D.R., 2006. A comparison of adaptive procedures for rapid and reliable threshold assessment and training in naive subjects. J. Acoust. Soc. Am. 119, 1616–1625.

Anderson, E.S., Nelson, D.A., Kreft, H., Nelson, P.B., Oxenham, A.J., 2011. Comparing spatial tuning curves, spectral ripple resolution, and speech perception in cochlear implant users. J. Acoust. Soc. Am. 130, 364–375.

Anderson, E.S., Oxenham, A.J., Nelson, P.B., Nelson, D.A., 2012. Assessing the role of spectral and intensity cues in spectral ripple detection and discrimination in cochlear-implant users. J. Acoust. Soc. Am. 132, 3925–3934.

Azadpour, M., McKay, C.M., 2012. A psychophysical method for measuring spatial resolution in cochlear implants. J. Assoc. Res. Otolaryngol. 13, 145–157.

Baker, R.J., Rosen, S., 2001. Evaluation of maximum-likelihood threshold estimation with tone-in-noise masking. Br. J. Audiol. 35, 43–52.

Bernstein, L.R., Green, D.M., 1987. Detection of simple and complex changes of spectral shape. J. Acoust. Soc. Am. 82, 1587–1592.
Bernstein, J.G., Mehraei, G., Shamma, S., Gallun, F.J., Theodoroff, S.M., Leek, M.R., 2013. Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired subjects. J. Am. Acad. Audiol. 24, 293–306.
Bregman, A., 1990. Auditory Scene Analysis: the Perceptual Organization of Sound. The MIT Press, Cambridge, MA.
Cazals, Y., Pelizzone, M., Saudan, O., 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.
Chi, T., Gao, Y., Guyton, M.C., Ru, P., Shamma, S., 1999. Spectro-temporal modulation transfer functions and speech intelligibility. J. Acoust. Soc. Am. 106, 2719–2732.
Choi, J.E., Hong, S.H., Won, J.H., Park, H.S., Cho, Y.S., Chung, W.H., Cho, Y.S., Moon, I.J., 2016. Evaluation of cochlear implant Candidates using a non-linguistic spectrotemporal modulation detection test. Sci Rep-UK 6.
Donaldson, G.S., Nelson, D.A., 2000. Place-pitch sensitivity and its relation to consonant recognition by cochlear implant subjects using the MPEAK and SPEAK speech processing strategies. J. Acoust. Soc. Am. 107, 1645–1658.
Donaldson, G.S., Viemeister, N.F., 2000. Intensity discrimination and detection of amplitude modulation in electric hearing. J. Acoust. Soc. Am. 108, 760–763.
Drennan, W.R., Won, J.H., Nie, K., Jameyson, E., Rubinstein, J.T., 2010. Sensitivity of psychophysical measures to signal processor modifications in cochlear implant users. Hear Res. 262, 1–8.
Eddins, D.A., Bero, E.M., 2007. Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region. J. Acoust. Soc. Am. 121, 363–372.
Elhilali, M., Chi, T., Shamma, S., 2003. Intelligibility and the spectrotemporal representation of speech in the auditory cortex. Speech Commun. 41, 331–348.
Elliott, T.M., Theunissen, F.E., 2009. The modulation transfer function for speech intelligibility. PLoS Comput. Biol. 5, e1000302.
Escabí, M.A., Miller, L.M., Read, H.L., Schreiner, C.E., 2003. Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus. J. Neurosci. 23, 11489–11504.
Fitzgibbons, P.J., Gordon-Salant, S., 1995. Age effects on duration discrimination with simple and complex stimuli. J. Acoust. Soc. Am. 98, 3140–3145.
Friesen, L.M., Shannon, R.V., 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.
Fritz, J., Shamma, S., Elhilali, M., Klein, D., Niwa, M., Johnson, J.S., O'Connor, K.N., Sutter, M.L., 2003. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex Active engagement improves primary auditory cortical neurons' ability to discriminate temporal modulation. Nat. Neurosci. 6, 1216–1223.
Fu, Q.J., 2002. Temporal processing and speech recognition in cochlear implant users. Neuroreport 13, 1635–1639.
Fu, Q.J., Shannon, R.V., 2000. Effect of stimulation rate on phoneme recognition by nucleus-22 cochlear implant subjects. J. Acoust. Soc. Am. 107, 589–597.
Galvin 3rd, J.J., Fu, Q.J., 2005. Effects of stimulation rate, mode and level on modulation detection by cochlear implant users. J. Assoc. Res. Otolaryngol. 6, 269–279.
Galvin 3rd, J.J., Oba, S., Fu, Q.J., Baskent, D., 2014. Single- and multi-channel modulation detection in cochlear implant users. PLoS One 9, e99338.
Glasberg, B.R., Moore, B.C., 1990. Derivation of auditory filter shapes from notched-noise data. Hear Res. 47, 103–138.
Gnansia, D., Lazard, D.S., Leger, A.C., Fugain, C., Lancelin, D., Meyer, B., Lorenzi, C., 2014. Role of slow temporal modulations in speech identification for cochlear implant users. Int. J. Audiol. 53, 48–54.
Goupell, M.J., Laback, B., Majdak, P., Baumgartner, W.D., 2008. Current-level discrimination and spectral profile analysis in multi-channel electrical stimulation. J. Acoust. Soc. Am. 124, 3142–3157.
Henry, B.A., Turner, C.W., 2003. The resolution of complex spectral patterns by cochlear implant and normal-hearing subjects. J. Acoust. Soc. Am. 113, 2861–2873.
Henry, B.A., McKay, C.M., McDermott, H.J., Clark, G.M., 2000. The relationship between speech perception and electrode discrimination in cochlear implantees. J. Acoust. Soc. Am. 108, 1269–1280.
Henry, B.A., Turner, C.W., Behrens, A., 2005. Spectral peak resolution and speech recognition in quiet: normal hearing, hearing impaired, and cochlear implant subjects. J. Acoust. Soc. Am. 118, 1111–1121.
Holden, L.K., Finley, C.C., Firszt, J.B., Holden, T.A., Brenner, C., Potts, L.G., Gotter, B.D., Vanderhoof, S.S., Mispagel, K., Heydebrand, G., Skinner, M.W., 2013. Factors affecting open-set word recognition in adults with cochlear implants. Ear Hear. 34, 342–360.
Horst, J.W., 1987. Frequency discrimination of complex signals, frequency selectivity, and speech perception in hearing-impaired subjects. J. Acoust. Soc. Am. 82, 874–885.
Jones, G.L., Won, J.H., Drennan, W.R., Rubinstein, J.T., 2013. Relationship between channel interaction and spectral-ripple discrimination in cochlear implant users. J. Acoust. Soc. Am. 133, 425–433.
Kowalski, N., Depireux, D.A., Shamma, S.A., 1996. Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. J. Neurophysiol. 76, 3503–3523.
Kral, A., Sharma, A., 2012. Developmental neuroplasticity after cochlear implantation. Trends Neurosci. 35, 111–122.
Kral, A., Tillein, J., Heid, S., Klinke, R., Hartmann, R., 2006. Cochlear implants: cortical plasticity in congenital deprivation. Prog. Brain Res. 157, 283–313.
Levitt, H., 1971. Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49 (Suppl 2), 467.
Liberman, A.M., 1996. Speech: a special Code. MIT Press, Cambridge. Massachusetts.
Litvak, L.M., Spahr, A.J., Saoji, A.A., Fridman, G.Y., 2007. Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder subjects. J. Acoust. Soc. Am. 122, 982–991.
Liu, C., Eddins, D.A., 2008. Effects of spectral modulation filtering on vowel identification. J. Acoust. Soc. Am. 124, 1704–1715.
Miller, G.A., 1947. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. J. Acoust. Soc. Am. 191, 609–619.
Oxenham, A.J., Shera, C.A., 2003. Estimates of human cochlear tuning at low levels using forward and simultaneous masking. J. Assoc. Res. Otolaryngol. 4, 541–554.
Parkinson, A.J., Arcaroli, J., Staller, S.J., Arndt, P.L., Cosgriff, A., Ebinger, K., 2002. The nucleus 24 contour cochlear implant system: adult clinical trial results. Ear Hear 23, 41S–48S.
Peterson, N.R., Pisoni, D.B., Miyamoto, R.T., 2010. Cochlear implants and spoken language processing abilities: review and assessment of the literature. Restor. Neurol. Neurosci. 28, 237–250.
Phillips, S.L., Gordon-Salant, S., Fitzgibbons, P.J., Yeni-Komshian, G.H., 1994. Auditory duration discrimination in young and elderly subjects with normal hearing. J. Am. Acad. Audiol. 5, 210–215.
Rosen, S., 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. Philos. Trans. R. Soc. Lond B Biol. Sci. 336, 367–373.
Sabin, A.T., Eddins, D.A., Wright, B.A., 2012. Perceptual learning evidence for tuning to spectrotemporal modulation in the human auditory system. J. Neurosci. 32, 6542–6549.
Saoji, A.A., Eddins, D.A., 2007. Spectral modulation masking patterns reveal tuning to spectral envelope frequency. J. Acoust. Soc. Am. 122, 1004–1013.
Saoji, A.A., Litvak, L., Spahr, A.J., Eddins, D.A., 2009. Spectral modulation detection and vowel and consonant identifications in cochlear implant subjects. J. Acoust. Soc. Am. 126, 955–958.
Shannon, R.V., 1992. Temporal modulation transfer functions in patients with cochlear implants. J. Acoust. Soc. Am. 91, 2156–2164.
Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.
Singh, N.C., Theunissen, F.E., 2003. Modulation spectra of natural sounds and ethological theories of auditory processing. J. Acoust. Soc. Am. 114, 3394–3411.
Slee, S.J., David, S.V., 2015. Rapid task-related plasticity of spectrotemporal receptive fields in the auditory midbrain. J. Neurosci. 35, 13090–13102.
Souza, P.E., Kitch, V., 2001. The contribution of amplitude envelope cues to sentence identification in young and aged subjects. Ear Hear 22, 112–119.
Summers, V., Leek, M.R., 1994. The internal representation of spectral contrast in hearing-impaired subjects. J. Acoust. Soc. Am. 95, 3518–3528.
Theunissen, F.E., Sen, K., Doupe, A.J., 2000. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J. Neurosci. 20, 2315–2331.
Throckmorton, C.S., Collins, L.M., 1999. Investigation of the effects of temporal and spatial interactions on speech-recognition skills in cochlear-implant subjects. J. Acoust. Soc. Am. 105, 861–873.
van den Honert, C., Stypulkowski, P.H., 1987. Temporal response patterns of single auditory nerve fibers elicited by periodic electrical stimuli. Hear Res. 29, 207–222.
Vandali, A.E., Whitford, L.A., Plant, K.L., Clark, G.M., 2000. Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system. Ear Hear 21, 608–624.
Weber, E.H., 1834. De pulse, resorptione, auditu et tactu. Koehler, Leipzig.
Won, J.H., Drennan, W.R., Rubinstein, J.T., 2007. Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. J. Assoc. Res. Otolaryngol. 8, 384–392.
Won, J.H., Drennan, W.R., Nie, K., Jameyson, E.M., Rubinstein, J.T., 2011a. Acoustic temporal modulation detection and speech perception in cochlear implant subjects. J. Acoust. Soc. Am. 130, 376–388.
Won, J.H., Jones, G.L., Drennan, W.R., Jameyson, E.M., Rubinstein, J.T., 2011b. Evidence of across-channel processing for spectral-ripple discrimination in cochlear implant subjects. J. Acoust. Soc. Am. 130, 2088–2097.
Woolley, S.M., Fremouw, T.E., Hsu, A., Theunissen, F.E., 2005. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. Nat. Neurosci. 8, 1371–1379.
Wouters, J., McDermott, H.J., Francart, T., 2015. Sound coding in cochlear implants: from electric pulses to hearing. IEEE Signal Process. Mag. 32, 67–80.
Xu, L., Thompson, C.S., Pfingst, B.E., 2005. Relative contributions of spectral and temporal cues for phoneme recognition. J. Acoust. Soc. Am. 117, 3255–3267.

Please cite this article in press as: Zheng, Y., et al., Spectro-temporal cues enhance modulation sensitivity in cochlear implant users, Hearing Research (2017), http://dx.doi.org/10.1016/j.heares.2017.05.009