Systems/Circuits

Adaptation to Noise in Human Speech Recognition Unrelated to the Medial Olivocochlear Reflex

Miriam I. Marrufo-Pérez,[1,2] **Almudena Eustaquio-Martín,**[1,2] **and Enrique A. Lopez-Poveda**[1,2,3]

[1]Instituto de Neurociencias de Castilla y León, [2]Instituto de Investigación Biomédica de Salamanca, and [3]Departamento de Cirugía, Facultad de Medicina, Universidad de Salamanca, 37007 Salamanca, Spain

Sensory systems constantly adapt their responses to the current environment. In hearing, adaptation may facilitate communication in noisy settings, a benefit frequently (but controversially) attributed to the medial olivocochlear reflex (MOCR) enhancing the neural representation of speech. Here, we show that human listeners ($N = 14$; five male) recognize more words presented monaurally in ipsilateral, contralateral, and bilateral noise when they are given some time to adapt to the noise. This finding challenges models and theories that claim that speech intelligibility in noise is invariant over time. In addition, we show that this adaptation to the noise occurs also for words processed to maintain the slow-amplitude modulations in speech (the envelope) disregarding the faster fluctuations (the temporal fine structure). This demonstrates that noise adaptation reflects an enhancement of amplitude modulation speech cues and is unaffected by temporal fine structure cues. Last, we show that cochlear implant users ($N = 7$; four male) show normal monaural adaptation to ipsilateral noise. Because the electrical stimulation delivered by cochlear implants is independent from the MOCR, this demonstrates that noise adaptation does not require the MOCR. We argue that noise adaptation probably reflects adaptation of the dynamic range of auditory neurons to the noise level statistics.

Key words: **adaptation; cochlear implant; envelope; medial olivocochlear reflex; olivocochlear efferents; temporal fine structure**

Significance Statement

People find it easier to understand speech in noisy environments when they are given some time to adapt to the noise. This benefit is frequently but controversially attributed to the medial olivocochlear efferent reflex enhancing the representation of speech cues in the auditory nerve. Here, we show that the adaptation to noise reflects an enhancement of the slow fluctuations in amplitude over time that are present in speech. In addition, we show that adaptation to noise for cochlear implant users is not statistically different from that for listeners with normal hearing. Because the electrical stimulation delivered by cochlear implants is independent from the medial olivocochlear efferent reflex, this demonstrates that adaptation to noise does not require this reflex.

Introduction

Verbal communication is harder in noisy than in quiet places. Sensory systems, however, constantly adapt their responses to the current environment (Webster, 2012), and hearing is not an exception. Listeners with normal hearing (NH) recognize more syllables or words in noisy backgrounds when the speech tokens are delayed from the noise onset than when they start at the same time as the noise (Cervera and Ainsworth, 2005; Cervera and Gonzalez-Alvarez, 2007; Ben-David et al., 2012, 2016). This adaptation to the noise probably facilitates everyday communication in noisy settings. The present study aimed at shedding light on the cues and mechanisms underlying this adaptation.

The mammalian cochlea effectively operates as a bank of bandpass filters, separating the multiple frequency components in speech for further processing by the auditory system (Robles and Ruggero, 2001). The signal at the output of each filter carries information in the slow fluctuations in amplitude over time (the envelope) as well as in the rapid fluctuations with a rate close to the center frequency of the filter [the temporal fine structure (TFS); Rosen, 1992]. Our sensitivity to amplitude modulations (AMs) improves when unmodulated noise precedes the AM carrier sound (Viemeister, 1979; Sheft and Yost, 1990; Almishaal et al., 2017; Marrufo-Pérez et al., 2018). This suggests that noise adaptation in AM sensitivity may underlay the adaptation to noise in speech rec-

Received Jan. 5, 2018; revised Feb. 26, 2018; accepted March 24, 2018.

Author contributions: E.A.L.-P. designed research; M.I.M.-P. performed research; A.E.-M. contributed unpublished reagents/analytic tools; M.I.M.-P. analyzed data; M.I.M.-P. and E.A.L.-P. wrote the paper.

This work was supported by a doctoral contract of the University of Salamanca and Banco Santander to M.I.M.-P., and by the European Regional Development Fund and the Spanish Ministry of Economy and Competitiveness (Grant BFU2015-65376-P to E.A.L.-P.). We thank Milagros J. Fumero for help with data collection, and MED-EL GmbH for technical support with the cochlear implant experiments.

The authors declare no competing financial interests.

Correspondence should be addressed to Enrique A. Lopez-Poveda, Instituto de Neurociencias de Castilla y León, Universidad de Salamanca, Calle Pintor Fernando Gallego 1, 37007 Salamanca, Spain. E-mail: ealopezpoveda@usal.es.

DOI:10.1523/JNEUROSCI.0024-18.2018

 0270-6474/18/384138-08$15.00/0

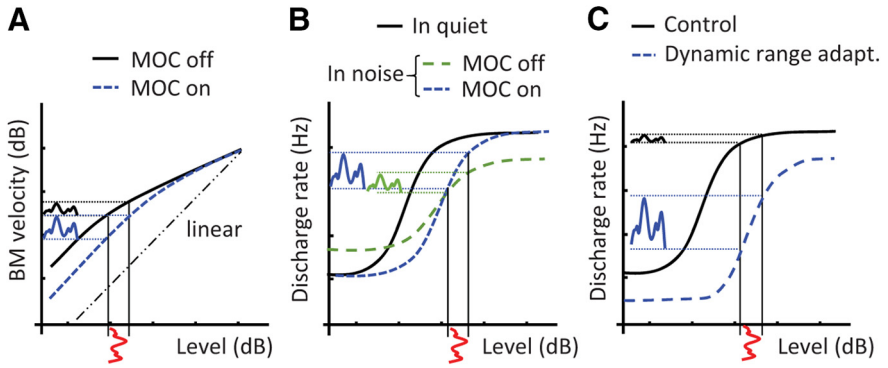

Figure 1. Schematic representation of three mechanisms that could enhance the amplitude modulation cues in speech. ***A***, Basilar membrane velocity versus sound pressure level with and without MOC efferent activation (adapted from Cooper and Guinan, 2006). ***B***, Auditory nerve rate-level functions in quiet and in noise with and without MOC efferent activation (adapted from Guinan, 2006). ***C***, Dynamic range adaptation of auditory neurons to the most frequently occurring level (adapted from Wen et al., 2009). Each panel illustrates the internal representation (ordinate) of the acoustic envelope depth (abscissa) associated with each mechanism. See main text for further details.

ognition. TFS cues, however, are also important for robust speech recognition (Qin and Oxenham, 2003; Zeng et al., 2005; Lorenzi et al., 2006; Hopkins and Moore, 2009; Johannesen et al., 2016; Lopez-Poveda et al., 2017), and it is still uncertain whether they may also contribute to the adaptation to noise.

Also uncertain are the mechanisms underlying adaptation. Some authors have reasoned that the adaptation to noise in AM sensitivity is due to the leading noise activating the medial olivocochlear reflex (MOCR; Almishaal et al., 2017; Marrufo-Pérez et al., 2018). This would linearize basilar membrane (BM) responses (Murugasu and Russell, 1996; Cooper and Guinan, 2003, 2006) and enhance the representation of speech AM cues in the BM response, as illustrated in Figure 1*A*. The MOCR may also enhance the representation of AM in the auditory nerve by restoring the dynamic range of auditory nerve fibers in noise to values observed in quiet, as shown in Figure 1*B* (Winslow and Sachs, 1988; Guinan, 2006). Because speech recognition in noise is easier with enhanced AM cues (Lorenzi et al., 1999; Apoux et al., 2001), the MOCR may underlay the adaptation to noise in speech recognition. This possibility is supported by the facts that (1) speech-in-noise recognition worsens when olivocochlear efferents are cut (Giraud et al., 1997), and (2) the time course of MOCR activation (280 ms; Backus and Guinan, 2006) is similar to the time course of adaptation to noise (~300 ms; Ben-David et al., 2012, 2016). However, the adaptation of the dynamic range of auditory neurons to the noise level statistics (Dean et al., 2005; Watkins and Barbour, 2008; Wen et al., 2009) can also enhance the AM cues in speech, as shown in Figure 1*C* (Almishaal et al., 2017; Marrufo-Pérez et al., 2018). Moreover, the time course of dynamic range adaptation (100–400 ms; Wen et al., 2012) is similar to the time course of adaptation to noise in speech recognition. It is still uncertain which of these two mechanisms (MOCR-related effects or dynamic range adaptation) plays a more prominent role in the adaptation to noise in speech recognition.

The present study aimed at investigating whether adaptation to noise in speech recognition reflects an enhancement of AM and/or TFS cues in speech, and whether the MOCR is necessary for adaptation to noise to occur.

Materials and Methods

Experimental design. We compared the signal-to-noise ratio (SNR) for 50% word recognition [termed the speech reception threshold (SRT)] for words presented monaurally at the onset ("early" condition) or delayed 300 ms from the onset ("late" condition) of a steady-state noise. The SRT improvement in the late versus the early condition was regarded as the amount of adaptation to the noise, or the temporal effect. Adaptation was measured for cochlear implant (CI) users and for NH listeners presented with natural and vocoded words. For NH listeners, adaptation was measured to ipsilateral, contralateral, and binaural (diotic) noise; for CI users, adaptation was measured to ipsilateral noise. Both the electrical stimulation delivered by the CIs and the vocoded words maintained the AM cues in speech but disregarded TFS cues (Shannon et al., 1995; Schatzer et al., 2010). Therefore, the presence of a normal temporal effect for CI users and for NH listeners tested with vocoded words would indicate that TFS cues contribute little to the adaptation to noise in normal hearing. On the other hand, the MOCR exerts its inhibitory effect via the outer hair cells (OHCs; Guinan, 1996). The electrical stimulation delivered to CI users bypasses OHCs and is independent from the MOCR (Lopez-Poveda et al., 2016). Therefore, the presence of a temporal effect for CI users would indicate that mechanisms other than MOCR-related adaptation can contribute to adaptation to noise.

Participants. Fourteen NH listeners (five men) participated in the study. Their mean ± SD age was 28.8 ± 8.2 years. All but two of them had audiometric thresholds of ≤20 dB hearing level (HL) in both ears at octave frequencies between 125 Hz and 8 kHz (American National Standards Institute, 1996). The exceptions were one participant whose threshold was 25 dB HL at 4 kHz in the left ear, and another participant whose threshold was 25 dB HL at 8 kHz in the left ear. Thirteen of the 14 NH listeners were tested with both natural and vocoded words; the other participant was tested with vocoded words only.

The CI group included seven users (four men) of CIs manufactured by MED-EL. Five users wore bilateral CIs, and two users wore a hearing aid in the ear opposite to the CI. Table 1 gives further information about the CI participants.

SRTs were measured monaurally in the left ear (NH listeners), the implanted ear (unilateral CI users), or the self-reported better ear (bilateral CI users; Table 1). All participants were native speakers of Castilian Spanish. They were volunteers and not paid for their services. The study was approved by the Ethics Committee of the University of Salamanca (Spain).

Stimuli. Thirty-five disyllabic words were used to measure each SRT. The first 10 words were always the same but were presented in random order, and they were included to give participants the opportunity to become familiar with the test condition. The last 25 words corresponded to one of the nine phonetically balanced lists from Cárdenas and Marrero (1994). Lists for children were used to test the younger CI user (Table 1, participant S5). Words were presented in random order across test conditions to minimize the possibility of the participant remembering the words.

A steady-state noise with a speech-shaped long-term spectrum [speech-shaped noise (SSN)] was used to measure the SRTs in noise background. For NH listeners, the noise level was fixed at 60 dB sound pressure level (SPL). This level was chosen because broadband noise at this level is capable of activating the MOCR without activating the middle-ear muscle reflex (Lilaonitkul and Guinan, 2009; Aguilar et al., 2013; Mishra and Lutman, 2014), a factor that could have confounded the results. For CI users, the noise level was set at −25 dB full scale (FS), where 0 dB FS corresponds to a signal with peak amplitude at unity. For reference, this level corresponds to ~70 dB SPL in the clinical CI audio processors of MED-EL. (Note that middle-ear muscle reflex could not have affected the SRTs for CI users because the electrical stimulation delivered by the CIs was independent from this reflex.) For the two participant groups, the noise finished 200 ms after the word offset. The

Table 1. CI user information

Participant ID	Bilateral/ unilateral	Better (test) ear	Gender	Etiology	Age (years)	CI use (months)	No. of channels active/used for testing	Pulse rate (pps)
S1	Unilateral	R	M	Unknown	49	128	11	1617
S2	Unilateral	R	F	Unknown	48	132	11	1653
S3	Bilateral	R	M	Unknown	16	172	10	1099
S4	Bilateral	L	F	Unknown	44	22	10	1754
S5	Bilateral	R	M	Genetic	8	83	12	1515
S6	Bilateral	L	M	Meningitis	48	175	9	1846
S7	Bilateral	L	F	Meningitis	35	147	11	1405

For bilateral CI users, data refer to the test ear. R, Right; L, left; M, male; F, female; pps, pulses per second; ID, identification.

noise was gated with raised cosine onset and offset ramps of a duration of 10 ms (NH listeners) or 50 ms (CI users).

Stimuli processing. CI users were not tested with their clinical audio processors but with an experimental processing strategy that lacked automatic gain control or any other form of dynamic (time-varying) processing that could have caused a temporal effect (Lopez-Poveda et al., 2016). The strategy included a high-pass pre-emphasis filter (a first-order Butterworth filter with a 3 dB cutoff frequency of 1.2 kHz); a bank of sixth-order Butterworth bandpass filters whose 3 dB cutoff frequencies followed a modified logarithmic distribution between 100 and 8500 Hz; envelope extraction via full-wave rectification and low-pass filtering (fourth-order Butterworth low-pass filter with a 3 dB cutoff frequency of 400 Hz); a fixed logarithmic compression function to map the wide dynamic range of sounds in the environment into the relatively narrow dynamic range of electrically evoked hearing (Boyd, 2006); and continuous interleaved sampling of compressed envelopes with biphasic electrical pulses (Wilson et al., 1991). The number of filters in the bank was identical to the number of active electrodes in the implant (Table 1).

The vocoder used to test NH listeners was virtually identical to the processing strategy used to test CI users except that (1) the number of bandpass filters in the bank was fixed to 12; (2) the carrier for each channel was a sinusoid at the central frequency of the channel, rather than a train of electrical pulses; and (3) no back-end compression function was used.

Procedure. To measure an SRT, the speech level varied adaptively using a one-down, one-up adaptive rule: it decreased after a correct response and increased after an incorrect response. The SRT was thus defined as the SNR giving 50% correct word recognition in the psychometric function (Levitt, 1971). The initial SNR was typically 0 or 5 dB, and the speech level changed in 4 dB steps (for NH listeners) or 3 dB steps (for CI users) between words 1 and 14, and in 2 dB steps between words 14 and 35. The SRT was calculated as the mean of the SNRs for the final 20 words for NH listeners, or final 17 words for CI users. Feedback was not given to the participants on the correctness of their responses.

For CI users, SRTs were measured for words embedded in ipsilateral noise. SRTs for the early and late conditions were measured in random order. For NH listeners, SRTs were measured for seven conditions [1 quiet + 3 noise lateralities (bilateral, ipsilateral, and contralateral) × 2 temporal positions (early and late)] for natural and vocoded words. For any given type of word (natural or vocoded), the seven conditions were administered in random order. SRTs for natural and vocoded words were measured in alternate order across participants. Three SRTs were always obtained for each test condition, and their mean was regarded as the SRT. Most participants were experienced in speech-in-noise listening tests; those who were not were trained in the SRT task until their performance became stable. The two or more SRTs measured during training were discarded from further analyses.

During the measurements, NH listeners were seated in a double-walled sound-attenuating booth, and the presentation of each word was controlled by the experimenter, who was sitting outside the booth without visual interaction with the listener. A sound cue (a 1 kHz pure tone with 500 ms duration) was presented 500 ms before the noise onset (or before the word onset in the quiet condition) to warn the listener about the stimulus presentation and to focus his/her attention on the speech recognition task. Without the cue, the listener may have been more distracted in the early condition than in the late condition (because the noise served as a cue in the late condition), which may have produced a "fake" temporal effect. The cue was used with NH listeners and with two CI users (participants S2 and S7). However, its use for testing CI users was deemed unnecessary because during the measurements CI users were seated in front of the experimenter, who implicitly warned them before presenting each word.

Before testing CI users, electrical current levels at maximum comfortable loudness (MCL) were measured using the method of adjustment. Minimum stimulation levels (i.e., thresholds) were set to 0%, 5%, or 10% of MCL values, according to each participant's clinical fitting (Boyd, 2006).

Apparatus. Stimuli were digitally stored and presented through custom-made Matlab software (version R2014a, MathWorks; RRID:SCR_001622). For NH listeners, stimuli were played via an RME FIREFACE 400 soundcard at a sampling rate of 44.1 kHz, and with 24 bit resolution. Stimuli were presented to the listeners using circumaural Sennheiser HD580 Headphones. Sound pressure levels were calibrated by placing the headphones on a KEMAR head (Knowles Electronics) equipped with a Zwislocki (model DB-100, Knowles Electronics) artificial ear connected to a sound level meter (model 2238, Brüel & Kjaer). Calibration was performed at 1 kHz, and the obtained sensitivity was used at all other frequencies.

For CI users, stimuli were stored digitally (at a 20 kHz sampling rate, 16 bit quantization), processed through the coding strategy, and the resulting electrical stimulation patterns delivered using the Research Interface Box 2 (RIB2; Department of Ion Physics and Applied Physics, the University of Innsbruck, Innsbruck, Austria) and each patient's implanted receiver/stimulator.

Statistical analyses. Statistical analyses were performed with SPSS Statistics, version 23 (IBM; RRID:SCR_002865). Repeated-measures ANOVA (RMANOVA) or paired Student's *t* test was used, as appropriate, to test for the statistical significance of temporal position (early vs late), noise laterality (bilateral, ipsilateral, and contralateral), and/or word type (natural vs vocoded) on group mean SRTs. We hypothesized that SRTs would be better (lower) in the late condition than in the early conditions. Accordingly, we applied one-tailed tests when testing for differences in temporal position, and two-tailed tests for all other comparisons. Greenhouse–Geisser corrections were applied when the sphericity assumption was violated. For tests involving multiple groups or variables, *post hoc* pairwise comparisons were conducted using Bonferroni corrections for multiple comparisons. An effect was regarded as statistically significant when the null hypotheses could be rejected with 95% confidence ($p < 0.05$).

Results

Listeners with normal hearing

The mean SRT values in quiet were 21.2 ± 4.3 and 23.6 ± 4.1 dB SPL, respectively, for natural and vocoded words, and the difference was statistically significant ($t_{(12)} = -6.40$, $p < 0.001$, paired *t* test). Worse speech recognition in quiet has been reported previously when the TFS was removed and when the speech envelope was obtained using a reduced number of spectral channels in the vocoder (Apoux and Healy, 2010). Figure 2 shows the mean SRTs (SNR in decibels) for NH listeners, for natural (diamonds) and vocoded (circles) words, and for the early and late conditions. A three-way RMANOVA, using word type, noise laterality, and temporal position as factors, revealed worse (higher) SRTs for vocoded words than for natural words ($F_{(1,12)} = 218.4$, $p \leq 0.001$). This is consistent with the study by Qin and Oxenham (2003), who reported worse (higher) SRTs in SSN for vocoded speech than for natural speech when 4, 8, and 24 channels were used in the vocoder. The RMANOVA also revealed that SRTs were significantly different for bilateral, ipsilateral, and contralateral noises ($F_{(1.1,12.8)} = 593.0$, $p \leq 0.001$). Multiple pairwise com-

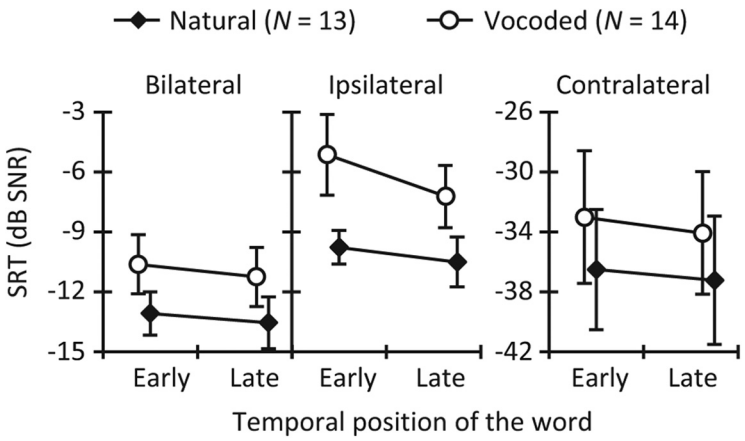

Figure 2. Speech reception thresholds for NH listeners. Mean SRTs in the early (0 ms word–noise-onset delay) and late (300 ms word-noise-onset delay) conditions for natural (diamonds) and vocoded (circles) disyllabic words presented monaurally in bilateral (left), ipsilateral (middle), and contralateral (right) noise. Note the different ordinate scale for contralateral noise. Error bars illustrate 1 SD.

parisons indicated significantly different SRTs across the three noise lateralities ($p \leq 0.001$).

For natural words, mean SRTs were better (lower) in the late condition than in the early condition for the three noise lateralities (Fig. 2). This occurred also for most individuals (Fig. 3*A*). A two-way RMANOVA with the factors temporal position and noise laterality revealed that SRTs were better (lower) in the late than in the early condition ($F_{(1,12)} = 12.38$, $p = 0.002$). *Post hoc* comparisons showed that the SRT improvement in the late versus the early condition was significant for each of the three noise lateralities (bilateral, $p = 0.023$; ipsilateral, $p = 0.005$; and contralateral, $p = 0.040$; Fig. 3*A*). The interaction between temporal position and noise laterality was not significant ($F_{(1.4,16.8)} = 0.3$, $p = 0.665$), indicating that the magnitude of the improvement in SRT was comparable across the three noise lateralities.

The pattern of mean SRTs for vocoded words was similar to that for natural words (Fig. 2), and most individual participants also showed a temporal effect for vocoded words (Fig. 3*B*). A two-way RMANOVA with factors temporal position and noise laterality revealed better (lower) SRTs in the late condition than in the early condition ($F_{(1,13)} = 25.1$, $p \leq 0.001$). *Post hoc* comparisons indicated that the improvement in SRT was significant for bilateral ($p = 0.049$), ipsilateral ($p \leq 0.001$), and contralateral ($p = 0.008$) noises (Fig. 3*B*). The interaction between temporal position and noise laterality was significant ($F_{(2,26)} = 3.7$, $p = 0.039$), indicating that the magnitude of the SRT improvement was different across the noise lateralities. However, although pairwise comparisons without corrections revealed a statistically greater improvement for the ipsilateral noises than for the bilateral or contralateral noises, those differences became not significant after applying the Bonferroni corrections.

Figure 3*C* allows a direct comparison of the SRT improvement for natural and vocoded words. For natural words, the mean improvement was 0.5 ± 0.8, 0.7 ± 0.9, and 0.7 ± 1.3 dB, respectively, for bilateral, ipsilateral, and contralateral noises. For vocoded words, the mean improvement was 0.6 ± 1.4, 2.1 ± 1.9, and 1.2 ± 1.4 dB, respectively, for bilateral, ipsilateral, and contralateral noises. A two-way RMANOVA with noise laterality and word type as factors indicated that the SRT improvement was not statistically different for natural and vocoded words ($F_{(1,12)} = 4.2$, $p = 0.064$). The interaction between word type and noise laterality was not significant ($F_{(2,24)} = 1.9$, $p = 0.168$).

Cochlear implant users

Six of the seven CI users showed better (lower) SRTs in the late condition than in the early condition (Fig. 4*A*). The group mean improvement in SRT was 2.8 ± 1.9 dB (Fig. 4*B*) and was statistically significant ($t_{(6)} = 3.95$, $p = 0.004$, paired *t* test). The improvement was not statistically different from that for NH listeners tested with vocoded words in ipsilateral noise ($t_{(19)} = 0.82$, $p = 0.420$, independent-samples *t* test; Fig. 4*B*).

"Good" performers benefit less from adaptation to noise

The effects of MOCR activation, as measured using otoacoustic emissions, are stronger for bilateral than for unilateral MOCR elicitors (Berlin et al., 1995; Guinan et al., 2003; Lilaonitkul and Guinan, 2009). Similarly, dynamic range adaptation is almost certainly stronger with diotic than with monaural stimuli (Wen et al., 2009). Based on this, one would expect greater adaptation to bilateral than to ipsilateral or contralateral noise, regardless of whether noise adaptation is mediated by the MOCR and/or dynamic range adaptation. In contrast with this, we found adaptation to be largest for vocoded words in ipsilateral noise (Fig. 3*C*). This, however, need not mean that adaptation is largest in ipsilateral noise. Instead, it may indicate that adaptation is greater when baseline performance is worse or, conversely, that good performers show less adaptation to noise.

This is illustrated in Figure 5, which depicts the magnitude of adaptation (or SRT improvement in the late condition relative to the early condition) as a function of the SRT in the early condition. Results for the contralateral noise are shown in the left half of Figure 5, whereas results for other conditions are shown in the right half of the figure. Second-order polynomials were fitted (using a least-squares procedure) to the two datasets. Though more evident in the right half of Figure 5, the two polynomials suggest that the magnitude of adaptation increased with increasing SRT in the early condition; that is, the worse the listener performance in the early condition, the greater the SRT improvement in the late condition. This analysis suggests that adaptation was largest for NH listeners tested with vocoded words in ipsilateral noise because their baseline SRTs (the SRTs in the early condition) were the worst in this condition. The same reasoning might explain the slightly greater adaptation for CI users than for NH listeners (Fig. 4*B*). This "ceiling" effect makes it difficult to interpret the magnitude of adaptation for NH listeners in terms of noise laterality.

Discussion

Listeners with normal hearing showed adaptation (i.e., better SRTs in the late condition than in the early condition) to bilateral, ipsilateral, and contralateral noises, for both natural and vocoded words (Fig. 3*A*,*B*). The magnitude of adaptation was not statistically different across the three noise lateralities or word types (Fig. 3*C*). Cochlear implant users also showed adaptation to ipsilateral noise (Fig. 4), and the magnitude was not statistically different from that for NH listeners tested with vocoded words in ipsilateral noise. Good performers tended to show less adaptation to noise than bad performers (Fig. 5).

Comparison with previous studies

Our NH listeners showed better SRTs for natural words presented late rather than early in the noise. This is consistent with previous findings (Cervera and Ainsworth, 2005; Cervera and Gonzalez-Alvarez, 2007). Ben-David et al. (2016) reported SRTs to improve by ~4.5 dB over the first 300 ms, an improvement greater than found here for natural words and ipsilateral noise, the most similar condition (0.7 dB, $N = 13$; Fig. 3*A*). The reason for the difference in magnitude is uncertain, but we found the temporal improvement in SRT to be smaller when baseline performance was better (Fig. 5). The SRT values reported by Ben-David et al. (2016) ranged between 0 and 3 dB SNR for 0 ms speech–noise-onset delay; thus, they were much higher (worse) than the SRTs found here (mean, −9.8 dB SNR; Fig. 2). Those differences in SRTs might explain the differences in the temporal effect magnitudes observed between the two studies.

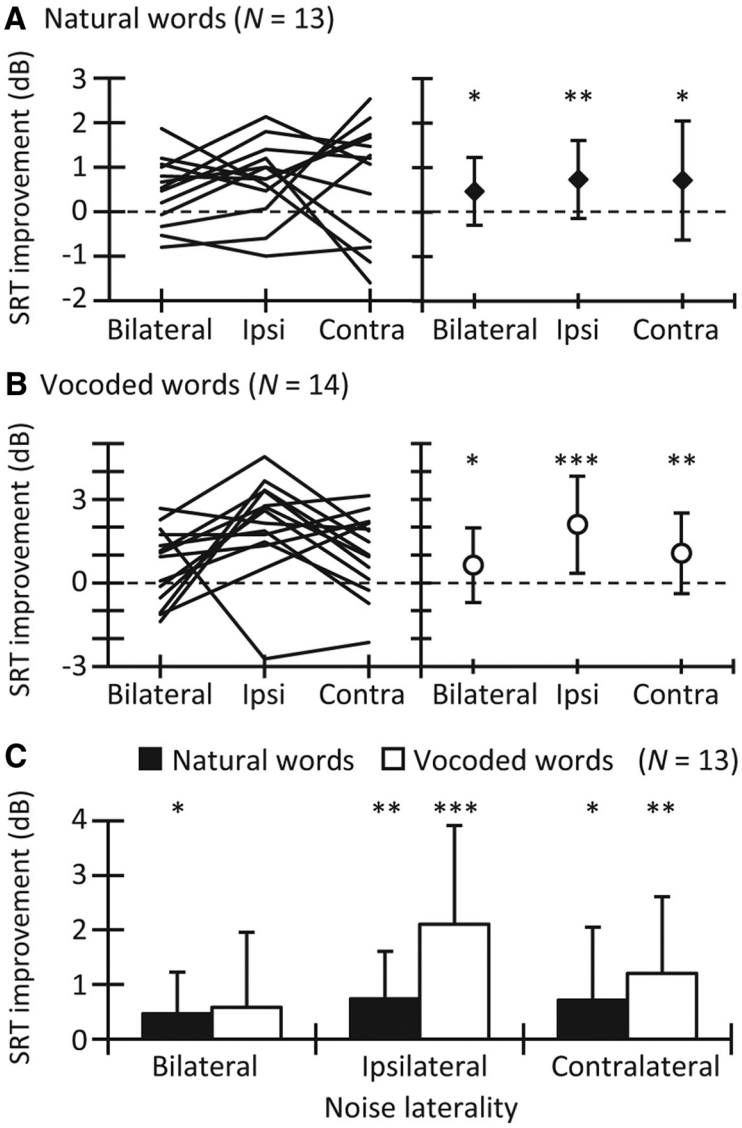

Figure 3. Improvement in SRT for words delayed 300 ms from the noise onset. Data are for NH listeners and for the three noise lateralities, as indicated in the abscissa of each panel. ***A***, For natural words. Each line in the left panel represents data for one participant ($N = 13$), and symbols in the right panel illustrate the mean data (±1 SD). ***B***, For vocoded words ($N = 14$). The layout is the same as in ***A***. ***C***, Mean data for natural and vocoded words ($N = 13$) replotted from ***A*** and ***B***. Error bars represent 1 SD. Asterisks indicate statistically significant differences: $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Contra, Contralateral; Ipsi, ipsilateral.

Mechanisms involved in adaptation to noise

Ben-David et al. (2012, 2016) provided evidence that the temporal improvement in word recognition is unlikely due to linguistic or cognitive processes and reasoned that it is likely due to physiological mechanisms. As explained in the Introduction, the linearization of BM responses caused by the MOCR could facilitate the recognition of speech delayed in noise by enhancing the AM cues in speech (Fig. 1*A*). Dubno et al. (2012) reported that the recognition of vocoded speech at −6 dB SNR improved when the speech level increased from 45 to 60 dB SPL, peaked at 60 dB SPL, and decreased progressively when the speech level increased from 60 to 85 dB SPL. In addition, they reported speech recognition to be better over the range of levels where the speech envelope is flattened as a result of BM compression. According to that, the linearization of BM responses caused by the MOCR (Fig. 1*A*) should result in a degradation rather than an improvement of speech recognition, which undermines the idea that the MOCR is involved in the adaptation to noise.

Most importantly, we have found that CI users show adaptation to noise (Fig. 4). The electrical stimulation provided by CIs is independent from the MOCR, and the sound-processing strategy used here lacked any form of dynamic processing. Therefore, although we cannot rule out the involvement of the MOCR in the temporal improvement found in SRT for NH listeners, the results for CI users demonstrate that mechanisms different from the MOCR can produce adaptation to noise in word recognition.

The mechanisms in question remain uncertain. However, studies in the AN (Wen et al., 2009, 2012), the inferior colliculus (Dean et al., 2005, 2008), and the auditory cortex (Watkins and Barbour, 2008) have demonstrated that, for a continuous sound of varying levels, neurons adapt their rate-level functions toward the most frequently occurring level in the sound. In other words, neurons show a displacement of the rate-level function (together with a decrement in firing rate) to adapt their dynamic range of operation to the sound-level distribution (Fig. 1*C*). This "neural dynamic range adaptation" might contribute to the reported temporal effect on word recognition in noise. Here, auditory neurons might have adapted their rate-level functions toward the leading noise and thus facilitate envelope coding when the speech level falls within the adapted rate-level curve (Fig. 1*C*), improving the SRT.

Auditory nerve fibers stimulated by electric pulse trains show firing rate adaptation (Zhang et al., 2007; Hu et al., 2010), and short-term firing rate adaptation and dynamic range adaptation are likely mediated by a common neural mechanism (Wen et al., 2012). Therefore, dynamic range adaptation is likely to occur also for electrical stimulation, which probably explains why CI users show adaptation to noise.

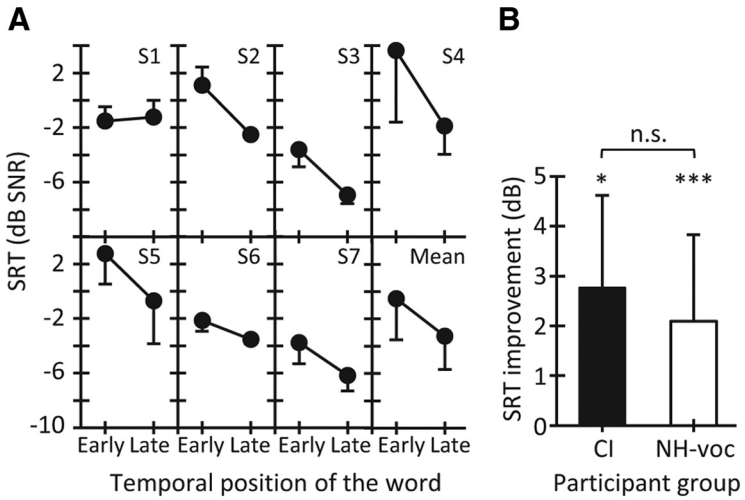

Figure 4. Results for cochlear implant users. ***A***, Individual and mean SRTs for words presented early and late in ipsilateral noise. ***B***, Mean SRT improvement in the late relative to the early condition for CI users ($N = 7$) and NH listeners ($N = 14$) tested with vocoded words in ipsilateral noise. Error bars represent 1 SD. NH-voc, NH listeners with vocoded words. Asterisks indicate statistically significant differences: $*p < 0.05$; $***p < 0.001$.

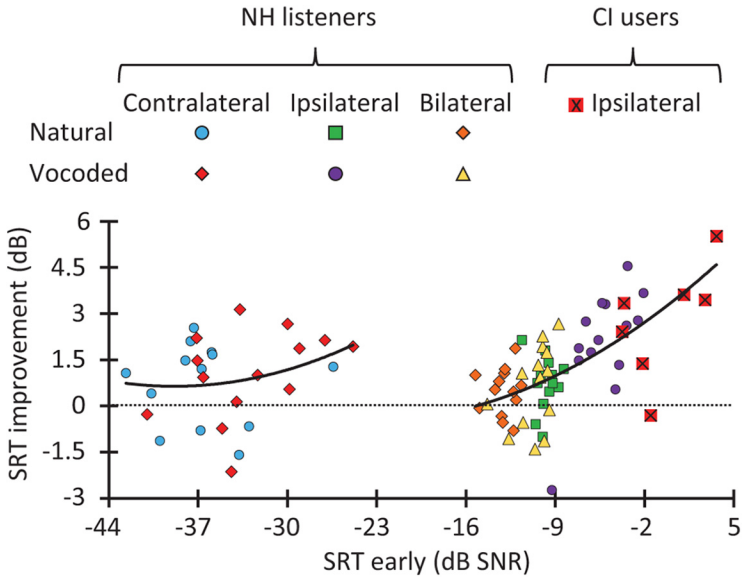

Figure 5. SRT improvement against the SRT in the early condition. Data are shown for all conditions and participants. The two continuous lines illustrate second-order polynomial fits to the data for contralateral noise conditions (left) and for the other conditions (right) separately. Notice that the worse (higher) the SRT in the early condition, the greater the SRT improvement (see main text for details).

The TFS does not influence adaptation to noise

Though not statistically significant, the amount of adaptation to ipsilateral noise tended to be smaller for NH listeners when tested with natural rather than with vocoded words (Fig. 3*C*), and to be smaller for NH listeners tested with vocoded words than for CI users (Fig. 4*B*). This, however, need not imply that adaptation is less when TFS cues are available. Instead, it may reflect differences in baseline performance across test conditions and groups. In other words, if TFS cues do not contribute to adaptation, temporal adaptation should be similar for vocoded and natural words across conditions of equal baseline performance (i.e., for conditions producing equal SRTs in the early condition). We tested this possibility by comparing the magnitude of the temporal effect for natural words in ipsilateral noise with that for vocoded words in bilateral noise, two conditions where baseline performance was not statistically different ($t_{(10)} = 1.63$, $p = 0.133$, paired t test; Fig. 5). In the comparison, we omitted data for two NH listeners whose SRTs for natural and vocoded words differed by >2 dB. These two conditions produced significant temporal effects (0.6 dB for natural words and 0.8 dB for vocoded words) that were not statistically different from each other ($t_{(10)} = -0.37$, $p = 0.718$, paired t test). That TFS cues were available with natural words but not with vocoded words suggests that the TFS cues in speech hardly affect adaptation to noise in natural hearing.

Limitations

The words used here lasted ~650 ms. The MOCR activation time constant is ~280 ms (Backus and Guinan, 2006), and dynamic range adaptation occurs over 100–400 ms (Wen et al., 2012). Therefore, for NH listeners, MOCR-related adaptation and dynamic range adaptation could have already occurred toward the end of each word in the early condition, thus reducing the magnitude of the temporal effect as measured here. The use of shorter speech tokens, such as syllables or phonemes, might reduce the amount of adaptation in the early condition and might reveal greater temporal effects than those reported here.

For a minority of NH listeners, broadband noise can activate the middle-ear muscle reflex at levels <60 dB SPL, the level of the SSN used here (Feeney et al., 2017). Here, adaptation to noise occurred for most NH listeners and in most conditions (Fig. 3). Therefore, it is unlikely that the middle-ear muscle reflex alone is responsible for this adaptation. Certainly, the middle-ear muscle reflex did not contribute to the adaptation found for CI users.

Implications

We have shown that human speech recognition adapts to the background noise, and that the magnitude of this adaptation varies depending on the difficulty of the task and/or the baseline (unadapted) performance (Fig. 5). Most speech intelligibility models assume that intelligibility in noise is constant over time [articulation index (Kryter, 1962); the speech intelligibility index (American National Standards Institute, 1997); the speech transmission index (Steeneken and Houtgast, 1980); the short-term objective intelligibility (Taal et al., 2011); or the speech-based envelope power spectrum model (Jørgensen et al., 2013)]. The incorporation of adaptation to these models would make them more accurate.

Since the discovery of the "antimasking" effect of the MOCR for detecting signals in low-level noise (Nieder and Nieder, 1970a,b), researchers have speculated that the MOCR might facilitate the recognition of speech in noise (for review, see Guinan, 1996, 2006). The reasoning was that the activation of the MOCR re-

stores the dynamic range of auditory nerve fibers in noise to values observed in quiet (Fig. 1*B*; Winslow and Sachs, 1988), which would enhance the neural coding of transient speech features, thus facilitating the perception of speech in noise. Though deeply investigated, the evidence in support of this idea is inconclusive (Lopez-Poveda, 2018). While some studies have reported worse speech-in-noise recognition after vestibular neurectomy (a surgery designed to cut MOC efferents; Giraud et al., 1997), others have demonstrated that vestibular neurectomy is ineffective in cutting the MOC efferents (Chays et al., 2003). Similarly, while some studies have found better speech-in-noise recognition with greater suppression of otoacoustic emissions levels by MOCR elicitor sounds (Mishra and Lutman, 2014; Bidelman and Bhagat, 2015), others have found the opposite trend (de Boer et al., 2012). Some studies have argued that factors such as selective attention, which has been shown to modulate MOC activity (de Boer and Thornton, 2007), are not usually controlled for and may be responsible for the lack of correlation in some studies. The present findings demonstrate that the MOCR is not necessary for adaptation to noise in speech recognition to occur, which might also be contributing to the controversial findings in previous studies.

References

Aguilar E, Eustaquio-Martin A, Lopez-Poveda EA (2013) Contralateral efferent reflex effects on threshold and suprathreshold psychoacoustical tuning curves at low and high frequencies. J Assoc Res Otolaryngol 14: 341–357. CrossRef Medline

Almishaal A, Bidelman GM, Jennings SG (2017) Notched-noise precursors improve detection of low-frequency amplitude modulation. J Acoust Soc Am 141:324–333. CrossRef Medline

American National Standards Institute (1996) S3.6 Specification for audiometers. New York: American National Standards Institute.

American National Standards Institute (1997) S3.5 Methods for calculating the speech intelligibility index. New York, NY: American National Standards Institute.

Apoux F, Healy EW (2010) Relative contribution of off- and on-frequency spectral components of background noise to the masking of unprocessed and vocoded speech. J Acoust Soc Am 128:2075–2084. CrossRef Medline

Apoux F, Crouzet O, Lorenzi C (2001) Temporal envelope expansion of speech in noise for normal-hearing and hearing-impaired listeners: effects on identification performance and response times. Hear Res 153: 123–131. CrossRef Medline

Backus BC, Guinan JJ Jr (2006) Time-course of the human medial olivocochlear reflex. J Acoust Soc Am 119:2889–2904. CrossRef Medline

Ben-David BM, Tse VY, Schneider BA (2012) Does it take older adults longer than younger adults to perceptually segregate a speech target from a background masker? Hear Res 290:55–63. CrossRef Medline

Ben-David BM, Avivi-Reich M, Schneider BA (2016) Does the degree of linguistic experience (native versus nonnative) modulate the degree to which listeners can benefit from a delay between the onset of the maskers and the onset of the target speech? Hear Res 341:9–18. CrossRef Medline

Berlin CI, Hood LJ, Hurley AE, Wen H, Kemp DT (1995) Binaural noise suppresses linear click-evoked otoacoustic emissions more than ipsilateral or contralateral noise. Hear Res 87:96–103. CrossRef Medline

Bidelman GM, Bhagat SP (2015) Right-ear advantage drives the link between olivocochlear efferent "antimasking" and speech-in-noise listening benefits. Neuroreport 26:483–487. CrossRef Medline

Boyd PJ (2006) Effects of programming threshold and maplaw settings on acoustic thresholds and speech discrimination with the MED-EL COMBI 40+ cochlear implant. Ear Hear 27:608–618. CrossRef Medline

Cárdenas MR, Marrero V (1994) Cuaderno de logoaudiometría. Madrid, Spain: Universidad Nacional de Educación a Distancia.

Cervera T, Ainsworth WA (2005) Effects of preceding noise on the perception of voiced plosives. Acta Acust United Acust 91:132–144.

Cervera T, Gonzalez-Alvarez J (2007) Temporal effects of preceding bandpass and band-stop noise on the recognition of voiced stops. Acta Acust United Acust 93:1036–1045.

Chays A, Maison S, Robaglia-Schlupp A, Cau P, Broder L, Magnan J (2003) Are we sectioning the cochlear efferent system during vestibular neurectomy? Rev Laryngol Otol Rhinol (Bord) 124:53–58. Medline

Cooper NP, Guinan JJ Jr (2003) Separate mechanical processes underlie fast and slow effects of medial olivocochlear efferent activity. J Physiol 548: 307–312. CrossRef Medline

Cooper NP, Guinan JJ Jr (2006) Efferent-mediated control of basilar membrane motion. J Physiol 576:49–54. CrossRef Medline

Dean I, Harper NS, McAlpine D (2005) Neural population coding of sound level adapts to stimulus statistics. Nat Neurosci 8:1684–1689. CrossRef Medline

Dean I, Robinson BL, Harper NS, McAlpine D (2008) Rapid neural adaptation to sound level statistics. J Neurosci 28:6430–6438. CrossRef Medline

de Boer J, Thornton AR (2007) Effect of subject task on contralateral suppression of click evoked otoacoustic emissions. Hear Res 233:117–123. CrossRef Medline

de Boer J, Thornton AR, Krumbholz K (2012) What is the role of the medial olivocochlear system in speech-in-noise processing? J Neurophysiol 107: 1301–1312. CrossRef Medline

Dubno JR, Ahlstrom JB, Wang X, Horwitz AR (2012) Level-dependent changes in perception of speech envelope cues. J Assoc Res Otolaryngol 13:835–852. CrossRef Medline

Feeney MP, Keefe DH, Hunter LL, Fitzpatrick DF, Garinis AC, Putterman DB, McMillan GP (2017) Normative wideband reflectance, equivalent admittance at the tympanic membrane, and acoustic stapedius reflex threshold in adults. Ear Hear 38:e142–e160. CrossRef Medline

Giraud AL, Garnier S, Micheyl C, Lina G, Chays A, Chéry-Croze S (1997) Auditory efferents involved in speech-in-noise intelligibility. Neuroreport 8:1779–1783. CrossRef Medline

Guinan JJ (1996) Efferent physiology. In: The cochlea (Dallos P, Popper AN, Fay RR, eds), pp 435–502. New York, NY: Springer.

Guinan JJ Jr (2006) Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. Ear Hear 27:589–607. CrossRef Medline

Guinan JJ Jr, Backus BC, Lilaonitkul W, Aharonson V (2003) Medial olivocochlear efferent reflex in humans: otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs. J Assoc Res Otolaryngol 4:521–540. CrossRef Medline

Hopkins K, Moore BC (2009) The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. J Acoust Soc Am 125:442–446. CrossRef Medline

Hu N, Miller CA, Abbas PJ, Robinson BK, Woo J (2010) Changes in auditory nerve responses across the duration of sinusoidally amplitude-modulated electric pulse-train stimuli. J Assoc Res Otolaryngol 11:641–656. CrossRef Medline

Johannesen PT, Pérez-González P, Kalluri S, Blanco JL, Lopez-Poveda EA (2016) The influence of cochlear mechanical dysfunction, temporal processing deficits, and age on the intelligibility of audible speech in noise by hearing-impaired listeners. Trends Hear 20:2331216516641055. CrossRef Medline

Jørgensen S, Ewert SD, Dau T (2013) A multi-resolution envelope-power based model for speech intelligibility. J Acoust Soc Am 134:436–446. CrossRef Medline

Kryter KD (1962) Methods for the calculation and use of the articulation index. J Acoust Soc Am 34:1698–1702. CrossRef

Levitt H (1971) Transformed up-down methods in psychoacoustics. J Acoust Soc Am 49:467–677. CrossRef

Lilaonitkul W, Guinan JJ Jr (2009) Human medial olivocochlear reflex: effects as functions of contralateral, ipsilateral, and bilateral elicitor bandwidths. J Assoc Res Otolaryngol 10:459–470. CrossRef Medline

Lopez-Poveda EA (2018) Olivocochlear Efferents in Animals and Humans: From Anatomy to Clinical Relevance. Front. Neurol 9:197. CrossRef

Lopez-Poveda EA, Eustaquio-Martín A, Stohl JS, Wolford RD, Schatzer R, Wilson BS (2016) A binaural clochlear implant sound coding strategy inspired by the contralateral medial olivocochlear reflex. Ear Hear 37: e138–e148. CrossRef Medline

Lopez-Poveda EA, Johannesen PT, Pérez-González P, Blanco JL, Kalluri S, Edwards B (2017) Predictors of hearing aid outcomes. Trends Hear 21: 2331216517730526. CrossRef Medline

Lorenzi C, Berthommier F, Apoux F, Bacri N (1999) Effects of envelope expansion on speech recognition. Hear Res 136:131–138. CrossRef Medline

Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BC (2006) Speech percep-

tion problems of the hearing impaired reflect inability to use temporal fine structure. Proc Natl Acad Sci U S A 103:18866–18869. CrossRef Medline

Marrufo-Pérez MI, Eustaquio-Martín A, López-Bascuas LE, Lopez-Poveda EA (2018) Temporal effects on monaural amplitude-modulation sensitivity in ipsilateral, contralateral and bilateral noise. J Assoc Res Otolaryngol 19:147–161. CrossRef Medline

Mishra SK, Lutman ME (2014) Top-down influences of the medial olivocochlear efferent system in speech perception in noise. PLoS One 9:e85756. CrossRef Medline

Murugasu E, Russell IJ (1996) The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea. J Neurosci 16:325–332. Medline

Nieder P, Nieder I (1970a) Stimulation of efferent olivocochlear bundle causes release from low level masking. Nature 227:184–185. CrossRef Medline

Nieder P, Nieder I (1970b) Antimasking effect of crossed olivococochlear bundle stimulation with loud clicks in guinea pig. Exp Neurol 28:179–188. CrossRef Medline

Qin MK, Oxenham AJ (2003) Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. J Acoust Soc Am 114:446–454. CrossRef Medline

Robles L, Ruggero MA (2001) Mechanics of the mammalian cochlea. Phsyiol Rev 81:1305–1352. CrossRef

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 336:367–373. CrossRef Medline

Schatzer R, Krenmayr A, Au DK, Kals M, Zierhofer C (2010) Temporal fine structure in cochlear implants: preliminary speech perception results in Cantonese-speaking implant users. Acta Otolaryngol 130:1031–1039. CrossRef Medline

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304. CrossRef Medline

Sheft S, Yost WA (1990) Temporal integration in amplitude modulation detection. J Acoust Soc Am 88:796–805. CrossRef Medline

Steeneken HJ, Houtgast T (1980) A physical method for measuring speech-transmission quality. J Acoust Soc Am 67:318–326. CrossRef Medline

Taal CH, Hendriks RC, Heusdens R, Jensen J (2011) An algorithm for intelligibility prediction of time–frequency weighted noisy speech. IEEE Trans Audio Speech Lang Process 19:2125–2136. CrossRef

Viemeister NF (1979) Temporal modulation transfer functions based upon modulation thresholds. J Acoust Soc Am 66:1364–1380. CrossRef Medline

Watkins PV, Barbour DL (2008) Specialized neuronal adaptation for preserving input sensitivity. Nat Neurosci 11:1259–1261. CrossRef Medline

Webster MA (2012) Evolving concepts of sensory adaptation. F1000 Biol Rep 4:21. CrossRef Medline

Wen B, Wang GI, Dean I, Delgutte B (2009) Dynamic range adaptation to sound level statistics in the auditory nerve. J Neurosci 29:13797–13808. CrossRef Medline

Wen B, Wang GI, Dean I, Delgutte B (2012) Time course of dynamic range adaptation in the auditory nerve. J Neurophysiol 108:69–82. CrossRef Medline

Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM (1991) Better speech recognition with cochlear implants. Nature 352:236–238. CrossRef Medline

Winslow RL, Sachs MB (1988) Single-tone intensity discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle. Hear Res 35:165–189. CrossRef Medline

Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargave A, Wei C, Cao K (2005) Speech recognition with amplitude and frequency modulations. Proc Natl Acad Sci U S A 102:2293–2298. CrossRef Medline

Zhang F, Miller CA, Robinson BK, Abbas PJ, Hu N (2007) Changes across time in spike rate and spike amplitude of auditory nerve fibers stimulated by electric pulse trains. J Assoc Res Otolaryngol 8:356–372. CrossRef Medline