Maintaining intelligibility at high speech intensities: Evidence of lateral inhibition in the lower auditory pathway

James A. Bashford, Jr.,^{a)} Richard M. Warren, and Peter W. Lenz

*Psychology Department, University of Wisconsin-Milwaukee, P.O. Box 413,
Milwaukee, Wisconsin 53201*

bashford@uwm.edu, rmwarren@uwm.edu, plenz@uwm.edu

Abstract: Three experiments examined the intelligibility enhancement produced when noise bands flank high intensity narrowband speech. Enhancement was unaffected by noise gating (experiment 1), ruling out peripheral adaptation as a source, and was also unaffected by interaural decorrelation of noise bands flanking diotic speech (experiment 2), indicating that enhancement occurs prior to binaural processing. These results support previous suggestions that intelligibility loss at high intensities is reduced by lateral inhibition in the cochlear nuclei. Results from a final experiment suggest that this effect is only ipsilateral, implicating a specific population of inhibitory neurons.

© 2013 Acoustical Society of America

PACS numbers: 43.71.An, 43.71.Es [SGS]

Date Received: March 20, 2013 **Date Accepted:** May 14, 2013

1. Introduction

The auditory system operates very effectively over an extraordinarily broad range of intensities, with a difference limen (DL) of only about 1 dB for signal levels up to 100 dB sound pressure level (SPL) and a DL of only about 1.5 dB at a signal level of 120 dB SPL.¹ Yet, the large majority of auditory nerve (AN) fibers have far narrower ranges (20 to 40 dB) over which their firing rates can vary with signal level before they reach rate saturation. This interesting “dynamic range problem” is also manifest in the perception of speech, which may remain nearly perfectly intelligible at levels exceeding 90 dB SPL,² despite the fact that most AN fibers reach their firing-rate limits at conversational speech levels of about 65 dB SPL, and at higher intensities are unable to provide a rate-based encoding of either the fine spectral details³ or amplitude envelope fluctuations⁴ that encode the critical features of speech.

Viemeister⁵ has examined intensity discrimination under conditions that excluded as likely cues both the spread-of-excitation to unsaturated AN fibers and neural synchrony with temporal fine-structure. He obtained compelling behavioral evidence that the dynamic range of rate-based discrimination does extend to 100 dB SPL, with an acuity of 1 dB, despite the limited ranges of most individual AN fibers. He has also provided a theoretical account of discrimination at high signal intensities⁵ which relies on firing-rate information provided by the small population of AN fibers known to have high thresholds and wide dynamic ranges, and he has argued that these fibers are sufficient to account for rate-based processing at the upper end of the dynamic range, if input from the larger population of readily saturated, low-threshold fibers is excluded from analysis at high signal intensities. Physiological models have been proposed^{6,7} that attribute this exclusionary process to mechanisms of lateral suppression, which reduce input from AN fibers when sufficient stimulation occurs in spectral regions adjacent to their best frequencies. These mechanisms of mutual suppression

^{a)} Author to whom correspondence should be addressed.

may include both mechanical (two-tone) suppression within the cochlea and a substantially more effective neural inhibition of input to cells of the cochlear nucleus.⁸ It has been proposed specifically^{6,7} that, at high intensities, lateral inhibition within the cochlear nucleus selectively attenuates input from low-threshold, readily saturated AN fibers, producing a shift in the weighting of intensity analysis to favor input from high-threshold unsaturated fibers.

Behavioral predictions from this lateral inhibition hypothesis were initially tested by Bashford *et al.*,⁹ who used a steeply filtered 2/3-octave band of “everyday” sentences centered at 1500 Hz and found that the narrowband speech was much more vulnerable than broadband speech to a decline, or “rollover” of intelligibility as intensity was increased. A significant intelligibility loss was obtained when the narrowband speech intensity reached 65 dB SPL, a level at which, as discussed above, most AN neurons are incapable of providing rate-based speech cues. This intelligibility rollover at moderate signal intensities was considered likely due to the absence of lateral suppression that would normally be evoked by speech components spectrally adjacent to the 2/3-octave band. Bashford *et al.* then tested the conjoint prediction that adding flanking bands of white noise would restore intelligibility of the speech band, by restoring lateral inhibition, but only at a speech intensity sufficient to produce rollover. Listeners were presented with the narrowband sentences at either 45 or 75 dBA, without flanking noise, which yielded average intelligibility scores of about 80% vs 60%, respectively, indicating a substantial rollover effect. When flanking noise bands were added to the 75-dBA speech at relative spectrum levels ranging from −40 to −20 dB, intelligibility increased, with a maximum recovery of 13% with noise presented at a spectrum level of −30 dB. As predicted, however, flanking noise did not increase the intelligibility of the sentences presented at 45 dBA.

The present study extends these earlier findings with experiments designed to test the hypothesis that the reduction of intelligibility rollover produced by flanking noise is due to lateral inhibitory interactions solely within the cochlear nucleus (CN). Experiment 1 contrasted the effects of gated vs continuous flanking noise to test for the more peripheral effect of firing-rate adaptation in AN fibers, which might be produced by out-of-band noise through the spread of activation to fibers responsive to frequencies within the speech pass-band. Experiment 2 examined the possible influence of a process more central than the CN: A masking level difference (MLD) paradigm was employed to examine the effect of binaural processing upon rollover reduction. To foreshadow, negative findings from these experiments provide additional support for the CN lateral inhibition hypothesis proposed by Eriksson and Robert⁷ and Winslow *et al.*⁸ Finally, experiment 3 contrasted the effects of ipsilateral vs contralateral flanking noise upon the intelligibility of a monaurally presented high-intensity speech band. The absence of a contralateral effect in that experiment should prove useful in identifying the neural circuits involved in the reduction of intelligibility rollover.

2. Methods

2.1 Listeners

The 85 listeners in this study (two groups of 30 and one group of 25) were undergraduate students at the University of Wisconsin-Milwaukee who were paid for their participation. They ranged in age from 18 to 28 and were native monolingual English speakers who reported having no hearing problems and had normal bilateral hearing, as measured by pure tone thresholds of 20 dB hearing level (HL) or better at octave frequencies from 250 to 8000 Hz. Each listener participated in only one experiment.

2.2 Stimuli

The formal test stimuli were the 100 (10 lists of 10) CID “everyday” sentences (e.g., “I’d like some ice cream with my pie”), which contain 500 keywords (50 keywords per list) that are used for scoring.¹⁰ An additional 25 practice sentences were drawn from

the high-predictability sublist of the revised Speech Perception in Noise (SPIN) test.¹¹ The original broadband recordings (44.1 kHz sampling with 16-bit quantization) were produced by the same male speaker who has no evident regional accent and an average voicing frequency of 100 Hz. Prior to filtering, the sentences were transduced by a Sennheiser HD 250 Linear II headphone and their slow-rms peak levels were matched to within 0.2 dBA using a flat-plate coupler in conjunction with a Brüel and Kjaer model 2230 digital sound-level meter set at A-scale weighting (as were all level measurements reported). The sentences were then passed through two successive stages of 4000-order band-pass finite impulse response (FIR) filtering (producing slopes of approximately 3.2 dB/Hz), using the `fir1` function in MATLAB, to produce a 2/3-octave narrow band of speech centered at 1500 Hz (passband from 1191 Hz to 1890 Hz), with both low-pass and high-pass filter slopes exceeding 3000 dB/octave. White noise, low-pass filtered at 20 kHz, was used to produce the three added-noise conditions employed in the study. They were (1) flanking bands of low-pass and high-pass filtered noise, providing out-of-band stimulation; (2) narrowband noise matching the spectral limits of the speech band and providing within-band stimulation; and (3) broadband noise, providing both within-band and out-of-band stimulation. The flanking bands of low-pass and high-pass noise were each produced through two successive stages of 4000-order FIR filtering, as with the speech band. The low-frequency limit of the lower flanking band and the high-frequency limit of the higher flanking band matched those of the broadband noise. Separate pairs of flanking noise bands were prepared for each noise spectrum-level employed in the present study (−10, −20, and −30 dB relative to the speech spectrum-level). Noise cutoff-frequencies for each pair were adjusted so that the filter skirts of the flanking noise bands would intersect the skirts of the speech band at an average level 60-dB below that of the speech cutoff frequencies. A similar procedure was followed in preparing the narrowband noise that matched the spectral limits of the speech band: White noise was passed through two stages of 4000-order FIR bandpass filtering, with cutoffs adjusted so that the transitions bands of the noise covered those of the speech band.

3. General procedure

The design of each experiment incorporated repeated measures. Before receiving a given stimulus condition listeners were presented with five practice sentences, which were first presented broadband and then presented band-pass filtered, along with noise when appropriate, in the same manner as the test sentences that followed. The different experimental conditions presented to a given group of listeners were assigned to separate sets of the “everyday” test sentences. This assignment varied pseudorandomly, with the restriction that each condition was applied an equal number of times to each set of sentences and appeared an equal number of times in each serial position across listeners in each experiment. Testing was performed in a sound attenuating chamber, with the stimuli delivered through Sennheiser HD 250 Linear II Headphones. Listeners were instructed to call out what the voice was saying as best they could and were encouraged to guess when unsure. Their responses were recorded and scored online by the experimenter, who sat with them in the chamber during the experiment.

3.1 Experiment 1: Effects of noise gating upon within-band masking vs out-of-band enhancement of high-intensity speech-band intelligibility

Given the spread of excitation that could be produced by flanking noise bands, it is possible that some of the reduction of rollover observed by Bashford *et al.*⁹ may have been due to noise stimulation within the frequency range of the speech band. Since the noise was presented continuously in those experiments, both during and between stimulus sentence presentations, within-band stimulation by the noise may have produced firing-rate adaptation in AN fibers, and shifted their operating ranges to include higher signal levels.¹² To examine this possible peripheral effect upon rollover, the 30 listeners in experiment 1 received the effectively rectangular band of everyday sentences

(2/3-octave centered at 1500 Hz) along with white noise that was presented continuously on half of the experimental trials, and was gated on and off with the sentences on the remaining trials to reduce adaptation. Peak levels of the sentences were 75 dBA and the noise stimuli were presented at a spectrum level of -20 dB relative to the speech. For half of the trials, in both the gated and continuous noise conditions, speech was accompanied by narrowband noise matching its spectral limits (and producing within-band masking). During the remaining trials, the speech band was presented along with broadband noise (producing both within-band masking and potential enhancement of intelligibility by out-of-band noise components). Gating was expected to reduce the extent of firing-rate adaptation to the noise, and thus increase within-band masking and thereby decrease intelligibility relative to the continuous noise condition. This reduction of intelligibility was expected to occur in both the narrowband and broadband noise conditions, since within-band masking was operative in both. However, if the reduction of rollover produced by out-of-band noise is due to lateral inhibition in the cochlear nucleus, rather than peripheral adaptation of AN fibers, gating should not effect the enhancement of intelligibility by out-of-band noise. Hence, it was predicted that the effects of gating and noise bandwidth would not interact. The speech and noise stimuli were presented diotically to the 30 listeners in this experiment.

3.2 Results for experiment 1

The proportions of keywords correctly reported by listeners for the separate sets of 25 sentences presented in each of the four experimental conditions were arcsine transformed and subjected to a repeated measures factorial analysis of variance, which yielded significant main effects of noise bandwidth [$F(1,29)=24.5$, $p<0.0001$] and gating [$F(1,29)=9.8$, $p<0.01$], and, as predicted, a nonsignificant interaction [$F(1,29)=1.9$, $p>0.17$]. Hence, intelligibility in the broadband noise condition was about 11% higher than in the narrowband noise condition, regardless of gating, and, as would be expected of simultaneous masking, intelligibility in the continuous noise condition was about 5% higher than in the gated noise condition, regardless of noise bandwidth. The results plotted in Fig. 1 may be taken to suggest that some underlying interaction occurred between gating and bandwidth, but simply did not reach statistical significance in the present study. It should be noted that this apparent interaction, involving a greater enhancement of intelligibility by gated rather than continuous flanking noise, not only failed to reach statistical significance, but also is the converse of what would be predicted for AN-fiber firing-rate adaptation as a source of intelligibility enhancement under these conditions.

Fig. 1. Mean percent intelligibility scores and standard errors in experiment 1 for 2/3-octave narrowband sentences presented at a slow-rms peak level of 75 dBA along with either broadband or narrowband white noise at -20 dB spectrum level relative to the speech. Noise was continuous on half of the trials and gated on and off with the sentences on the remaining trials.

3.3 Experiment 2: Effects of noise interaural correlation upon within-band masking vs out-of-band enhancement of high-intensity speech-band intelligibility

For the 30 listeners in experiment 2, the narrowband and broadband noise were always presented continuously. However, the noise stimuli added to the diotic speech band were only diotic (i.e., interaurally identical) in half of the experimental trials, and in the remaining trials were interaurally uncorrelated. It was expected that interaural decorrelation of the noise added to the diotic speech would produce a binaural release from masking exerted by the within-band noise components in both the narrowband and broadband noise conditions. However, because binaural processing is believed to occur at levels above the CN, beginning in the superior olivary complex (SOC), it also was predicted that interaural noise decorrelation would not affect the intelligibility enhancement (i.e., recovery from rollover) evoked by the out-of-band noise components in the broadband noise condition. Hence, noise bandwidth and interaural correlation effects were predicted not to interact.

3.4 Results for experiment 2

The results plotted in Fig. 2 also conform to predictions. Analysis of variance for listeners' arcsine-transformed proportions of correct responses in the four noise conditions yielded significant main effects of bandwidth [$F(1,29)=34.8$, $p<0.0001$] and noise interaural correlation [$F(1,29)=5.3$, $p<0.0002$], and a predicted nonsignificant interaction of those effects [$F(1,29)=0.03$, $p>0.80$]. Intelligibility in the broadband noise condition was about 13% higher than in the narrowband noise condition, regardless of noise correlation, and intelligibility in the uncorrelated noise conditions was about 5% higher than in the correlated noise conditions, regardless of noise bandwidth. Hence, the results of this experiment clearly indicate that the out-of-band enhancement effect occurs prior to binaural processing in the SOC.

3.5 Experiment 3: Effects of ipsilateral vs contralateral flanking noise upon the intelligibility of monaural narrowband speech

Experiment 3 was designed to determine whether the out-of-band enhancement effect is evoked when flanking noise and the high intensity speech band are delivered to opposite ears of the listener. This was accomplished by presenting the 75-dBA 2/3-octave band of everyday sentences monaurally to 25 listeners along with contralateral flanking bands of white noise at three spectrum levels relative to the speech: -30 , -20 , and -10 dB. There was also a no-noise baseline condition, as well as a condition delivering the flanking noise bands ipsilaterally at a spectrum level of -30 dB.

Fig. 2. Mean percent intelligibility scores and standard errors in experiment 2 for 2/3-octave narrowband sentences presented at a slow-rms peak level of 75 dBA along with either broadband or narrowband white noise at -20 dB spectrum level relative to the speech. Speech and noise were diotic on half of the trials, and on the remaining trials the noise was interaurally uncorrelated.

Table 1. Mean percent intelligibility scores and standard errors in experiment 3 for 2/3-octave narrowband sentences presented at a slow-rms peak level of 75 dBA along with either: no flanking noise, ipsilateral noise at -30 dB, or contralateral noise at one of three spectrum levels relative to the speech: -30 , -20 , or -10 dB.

Flanking Noise	None	-30 dB Contralateral	-20 dB Contralateral	-10 dB Contralateral	-30 dB Ipsilateral
Intelligibility (%)	57.2 (2.2)	57.4 (2.3)	55.4 (2.1)	56.3 (2.2)	64.2 (2.6)

3.6 Results for experiment 3

As can be seen in Table 1, the results of experiment 3 indicate that contralateral stimulation by flanking noise did not enhance intelligibility at any spectrum level. This was confirmed by a one-way repeated measures analysis of variance (ANOVA) [$F(4,96) = 3.78$, $p < 0.01$] and subsequent Tukey HSD tests, which compared the arcsine-transformed proportions of keywords correctly reported for the separate sets of 20 sentences presented in the five experimental conditions. Tukey tests indicated that only the ipsilateral flanking noise condition improved intelligibility relative to the no-noise baseline condition ($p < 0.05$).

4. General discussion

The absence of a contralateral noise effect in experiment 3 rules out two binaurally activated sources of peripheral suppression as contributors to the enhancement effect: the aural reflexes and the medial olivocochlear (MOC) reflex. It should also be noted that, aside from the present finding, it is questionable whether these reflexes could be functionally activated by flanking noise bands presented at the relatively low intensities found effective for enhancement in the present study. Similarly, another peripheral, but ipsilateral mechanism, two-tone suppression within the cochlea, might provide limited reduction of AN fiber firing-rate saturation under some conditions,⁸ but available evidence¹³ suggests those effects would be negligible for speech signals as high in level as 75 dBA, as used in the present study, even with suppressors of equal intensity, let alone 30 dB lower in spectrum level as employed in experiment 3. Thus, by exclusion, it appears likely that lateral inhibition in the CN is the primary contributor to the enhancement of speech intelligibility at high intensities, and available evidence suggests a specific source of this inhibition. Studies of single units in the mammalian cochlear nucleus have identified inhibitory neurons (onset-choppers) that are activated by broadband stimulation, and are comprised of two separate populations: those activated only by ipsilateral input and those also activated by contralateral input.¹⁴ The results of experiment 3 suggest that the population of strictly ipsilateral onset-choppers may be the source of inhibition responsible for preserving speech intelligibility at high intensities.

Acknowledgments

The project described was supported by Award Number R01DC000208 from the National Institute on Deafness and Other Communication Disorders. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders or the National Institutes of Health.

References and links

- ¹N. F. Viemeister, "Intensity coding and the dynamic-range problem," *Hear. Res.* **34**, 267–274 (1988).
- ²G. A. Studebaker, R. L. Sherbecoe, D. M. McDaniel, and C. A. Gwaltney, "Monosyllabic word recognition at higher-than-normal speech and noise levels," *J. Acoust. Soc. Am.* **105**, 2431–2444 (1999).

- ³M. B. Sachs and E. D. Young, "Encoding of steady-state vowels," *J. Acoust. Soc. Am.* **66**, 470–479 (1979).
- ⁴A. R. Palmer and E. F. Evans, "On the peripheral coding of the level of individual frequency components of complex sounds at high levels," *Exp. Brain Res.* **2**, 19–26 (1979).
- ⁵N. F. Viemeister, "Auditory intensity discrimination at high frequencies in the presence of noise," *Science* **221**, 1206–1207 (1983).
- ⁶J. L. Eriksson and A. Robert, "The representation of pure tones and noise in a model of cochlear nucleus neurons," *J. Acoust. Soc. Am.* **106**, 1865–1879 (1999).
- ⁷R. L. Winslow, P. Barta, and M. B. Sachs, "Rate coding in the auditory nerve," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ, 1987), pp. 212–224.
- ⁸W. S. Rhode and S. Greenberg, "Lateral suppression and inhibition in the cochlear nucleus of the cat," *J. Neurophysiol.* **71**, 493–514. (1994).
- ⁹J. A. Bashford, Jr., R. M. Warren, and P. W. Lenz, "Enhancing intelligibility of narrowband speech with out-of-band noise: Evidence for lateral suppression at high-normal intensity," *J. Acoust. Soc. Am.* **117**, 365–369 (2005).
- ¹⁰S. R. Silverman and I. J. Hirsh, "Problems related to the use of speech in clinical audiometry," *Ann. Otol. Rhinol. Laryngol.* **64**, 1234–1244 (1995).
- ¹¹R. C. Bilger, J. M. Nuetzel, W. M. Rabinowitz, and C. Rzeczkowski, "Standardization of a test of speech perception in noise," *Speech Hear. Res.* **27**, 32–48 (1984).
- ¹²D. J. Gibson, E. D. Young, and J. A. Costalupes, "Similarity of dynamic range adjustment in auditory nerve and cochlear nuclei," *J. Neurophysiol.* **53**, 940–958 (1985).
- ¹³M. A. Ruggero, L. Robles, and N. C. Rich, "Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression," *J. Neurophysiol.* **68**, 1087–1099 (1992).
- ¹⁴N. J. Ingham, S. Bleack, and I. M. Winter, "Contralateral inhibitory and excitatory frequency response maps in the mammalian cochlear nucleus," *Eur. J. Neurosci.* **24**, 2515–2529 (2006).