

Speech Masking in Normal and Impaired Hearing: Interactions Between Frequency Selectivity and Inherent Temporal Fluctuations in Noise

Andrew J. Oxenham and Heather A. Kreft

Abstract Recent studies in normal-hearing listeners have used envelope-vocoded stimuli to show that the masking of speech by noise is dominated by the temporal-envelope fluctuations inherent in noise, rather than just overall power. Because these studies were based on vocoding, it was expected that cochlear-implant (CI) users would demonstrate a similar sensitivity to inherent fluctuations. In contrast, it was found that CI users showed no difference in speech intelligibility between maskers with and without inherent envelope fluctuations. Here, these initial findings in CI users were extended to listeners with cochlear hearing loss and the results were compared with those from normal-hearing listeners at either equal sensation level or equal sound pressure level. The results from hearing-impaired listeners (and in normal-hearing listeners at high sound levels) are consistent with a relative reduction in low-frequency inherent noise fluctuations due to broader cochlear filtering. The reduced effect of inherent temporal fluctuations in noise, due to either current spread (in CI users) or broader cochlear filters (in hearing-impaired listeners), provides a new way to explain the loss of masking release experienced in CI users and hearing-impaired listeners when additional amplitude fluctuations are introduced in noise maskers.

Keywords Cochlear hearing loss · Hearing in noise · Speech perception

A. J. Oxenham (✉)

Departments of Psychology and Otolaryngology, University of Minnesota—Twin Cities,
Minneapolis, MN 55455, USA
e-mail: oxenham@umn.edu

H. A. Kreft

Department of Otolaryngology, University of Minnesota—Twin Cities,
Minneapolis, MN 55455, USA

© The Author(s) 2016

P. van Dijk et al. (eds.), *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*, Advances in Experimental Medicine and Biology 894,
DOI 10.1007/978-3-319-25474-6_14

1 Introduction

Speech perception is a major communication challenge for people with hearing loss and with cochlear implants (CIs), particularly when the speech is embedded in background noise (e.g., Humes et al. 2002; Zeng 2004). Recent work has suggested that it is not so much the overall noise energy that limits speech perception in noise, as suggested by earlier work (French and Steinberg 1947; Kryter 1962; George et al. 2008), but rather the energy in the inherent temporal-envelope modulations in noise (Dubbelboer and Houtgast 2008; Jorgensen and Dau 2011; Stone et al. 2011, 2012; Jorgensen et al. 2013; Stone and Moore 2014). In a recent study (Oxenham and Krefl 2014) we examined the effects of inherent noise fluctuations in CI users. In contrast to the results from normal-hearing (NH) listeners, we found that CI users exhibited *no* benefit of maskers without inherent fluctuations. Further experiments suggested that the effective inherent noise envelope fluctuations were reduced in the CI users, due to the effects of current spread, or interactions between adjacent electrodes, leading to smoother temporal envelopes.

One remaining question is whether HI listeners exhibit the same loss of sensitivity to inherent noise fluctuations as CI users. If so, the finding may go some way to explaining why both HI and CI populations exhibit less masking release than NH listeners when additional slow fluctuations are imposed on noise maskers (e.g., Festen and Plomp 1990; Nelson and Jin 2004; Stickney et al. 2004; Gregan et al. 2013). On one hand, cochlear hearing loss is generally accompanied by a loss of frequency selectivity, due to loss of function of the outer hair cells; on the other hand this “smearing” of the spectrum occurs before extraction of the temporal envelope, rather than afterwards, as is the case with CI processing. Because the smearing is due to wider filters, rather than envelope summation, the resultant envelopes from broader filters still have Rayleigh-distributed envelopes with the same overall relative modulation power as the envelopes from narrower filters (i.e., the same area under the modulation power spectrum, normalized to DC; see Fig. 1). In contrast, with CIs, the envelopes derived from summing the envelope currents from adjacent electrodes are no longer Rayleigh distributed and have lower modulation power, relative to the overall (DC) power in the envelope (Hu and Beaulieu 2005).

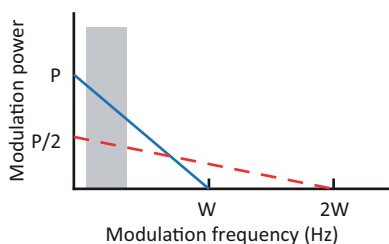


Fig. 1 Schematic diagram of the effect of broadening the filter from a bandwidth of W to a bandwidth of $2W$ on the modulation spectrum of filtered Gaussian noise. The relative modulation power (area under the line) remains constant, but the area under the lines within the speech-relevant range (*shaded rectangle*) is reduced

One factor suggesting that HI listeners may also experience less influence of inherent noise fluctuations is that the modulation spectrum is altered by broadening the filters: for an ideal rectangular filter, the modulation power of Gaussian noise after filtering has a triangular distribution, reaching a minimum of no power at a frequency equal to the bandwidth of the filter (Lawson and Uhlenbeck 1950). Although widening the filter does not alter the area under the modulation spectrum, it results in relatively less power at lower modulation frequencies (see Fig. 1). Given that low modulation frequencies are most important for speech, the relative reduction in modulation power at low modulation frequencies may reduce the influence of the inherent fluctuations for listeners with broader filters, due to hearing loss.

The aim of this experiment was to test the resulting prediction that hearing loss leads to less effect of inherent noise fluctuations on speech masking. We compared the results of listeners with cochlear hearing loss with the performance of young NH listeners and age-matched NH listeners. Performance was compared for roughly equal sensation levels (SL), and for equal sound pressure levels (SPL) to test for the effect of overall level on performance in NH listeners.

2 Methods

2.1 Listeners

Nine listeners with mild-to-moderate sensorineural hearing loss (4 male and 5 female; mean age 61.2 years) took part in this experiment. Their four-frequency pure-tone average thresholds (4F-PTA from 500, 1000, 2000, and 4000 Hz) ranged from about 25 to 65 dB HL (mean \sim 40 dB HL). Nine listeners with clinically normal hearing (defined as 20 dB HL or less at octave frequencies between 250 and 4000 Hz; mean 4F-PTA 7.6 dB HL), who were matched for age (mean age 62.2 years) and gender with the HI listeners, were run as the primary comparison group. In addition, a group of four young (mean age 20.5 years; mean 4F-PTA 2.8 dB HL) NH listeners were tested. All experimental protocols were approved by the Institutional Review Board of the University of Minnesota, and all listeners provided informed written consent prior to participation.

2.2 Stimuli

Listeners were presented with sentences taken from the AZBio speech corpus (Spahr et al. 2012). The sentences were presented to the HI listeners at an overall rms level of 85 dB SPL. The sentences were presented to the NH listeners at two different levels: in the equal-SL condition, the sentences were presented at

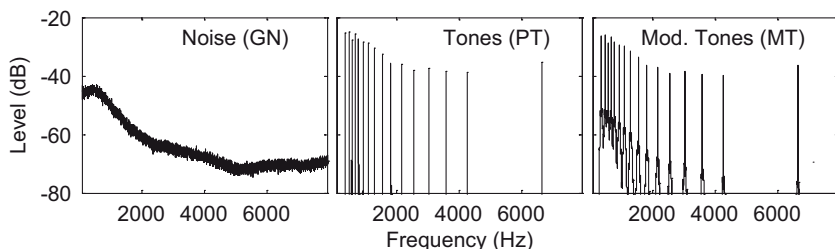


Fig. 2 Representation of the three masker types used in the experiment. The three panels provide spectral representations of the noise (*top*), tone (*middle*), and modulated tone (*bottom*) maskers

40 dB above the detection threshold of the speech for the NH listeners, similar to the sensation level (SL) of the 85-dB SPL speech for the HI listeners; in the equal-SPL condition, the speech was presented at 85 dB SPL. The sentences were presented in three different types of masker, as in Oxenham and Krefl (2014); see Fig. 2. The first was a Gaussian noise, spectrally shaped to match the long-term spectrum of the target speech. The second was comprised of 16 pure tones, approximately evenly spaced on a logarithmic frequency scale from 333 to 6665 Hz, corresponding to the center frequencies on a standard CI map for Advanced Bionics. The amplitudes of the tones were selected to produce the same long-term output of a 16-channel vocoder as the target speech (or Gaussian noise masker), assuming the same center frequencies. The third was comprised of the same 16 tones, but each tone was modulated independently with the temporal envelope of a noise masker, bandpass filtered using a vocoder filter with the same center frequency as the tone carrier. These three maskers produced equal amounts of speech masking in the CI users tested by Oxenham and Krefl (2014). The masker was gated on 1 s before the beginning of each sentence, and was gated off 1 s after the end of each sentence. The masker in each trial was a sample of a longer 25-s sound file, cut randomly from within that longer waveform. The speech and masker were mixed before presentation and the signal-to-masker ratios were selected in advance, based on pilot data, to span a range of performance between 0 and 100% word recognition.

The speech and the masker were mixed and low-pass filtered at 4000-Hz, and were either presented unprocessed or were passed through a tone-excited envelope vocoder that simulates certain aspects of CI processing (Dorman et al. 1998; Whitmal et al. 2007). The stimulus was divided into 16 frequency subbands, with the same center frequencies as the 16 tone maskers. The temporal envelope from each subband was extracted using a Hilbert transform, and then the resulting envelope was lowpass filtered with a 4th-order Butterworth filter and a cutoff frequency of 50 Hz. This cutoff frequency was chosen to reduce possible voicing periodicity cues, and to reduce the possibility that the vocoding produced spectrally resolved components via the amplitude modulation. Each temporal envelope was then used to modulate a pure tone at the center frequency of the respective subband.

2.3 Procedure

The stimuli were generated digitally, converted via a 24-bit digital-to-analog converter, and presented via headphones. The stimuli were presented to one ear (the better ear in the HI listeners), and the speech-shaped noise was presented in the opposite ear at a level 30 dB below the level of the speech. The listeners were seated individually in a double-walled sound-attenuating booth, and responded to sentences by typing what they heard via a computer keyboard. Sentences were scored for words correct as a proportion of the total number of keywords presented. One sentence list (of 20 sentences) was completed for each masker type and masker level. Presentation was blocked by condition (natural and vocoded, and speech level), and the order of presentation was counterbalanced across listeners. The test order of signal-to-masker ratios was random within each block.

3 Results

The results from the conditions where the speech was presented at roughly equal sensation level to all participants (85 dB SPL for the HI group, and 40 dB SL for the NH groups) are shown in Fig. 3. The upper row shows results without vocoding; the lower row shows results with vocoding. The results from the young NH, age-matched NH, and HI listeners are shown in the left, middle and right columns, respectively.

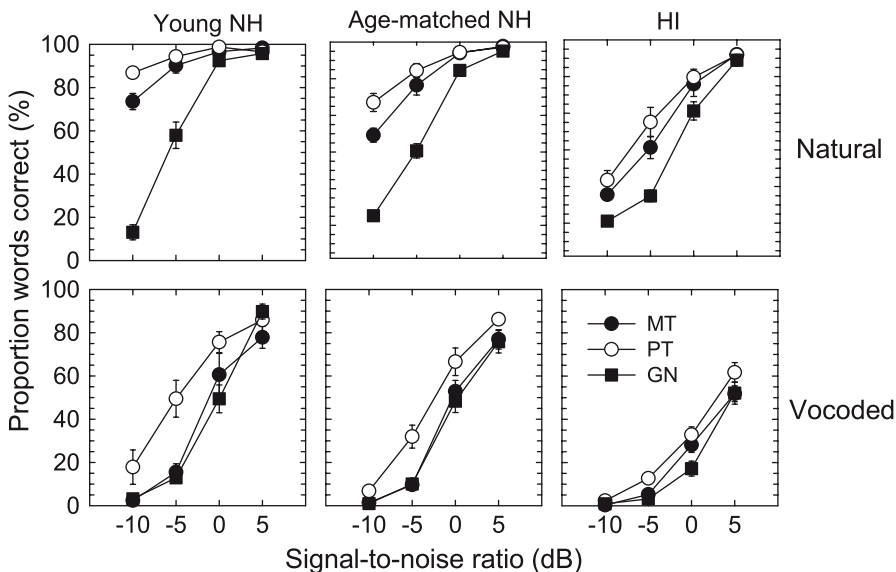


Fig. 3 Speech intelligibility with the speech at roughly equal sensation level (*SL*) across the three groups. Sentence recognition was tested in speech-shaped Gaussian noise (*GN*), pure tones (*PT*), and modulated tones (*MT*)

Consider first the results from the non-vocoded conditions (upper row of Fig. 3). Performance in the presence of Gaussian noise (GN; filled squares) is quite similar across the three groups but, as expected, performance is slightly poorer in the HI group. When the noise masker is replaced by 16 pure tones (PT; open circles), performance is markedly better—all three groups show a large improvement in performance, although there appears to be a systematic decrease in performance from young NH to age-matched NH to the HI group. Imposing noise envelope fluctuations to the pure-tone maskers (MT; filled circles) results in a small reduction in performance for all three groups. It appears that both the spectral sparsity of the tonal maskers (PT and MT), as well as the lack of fluctuations (PT), contributes to the release from masking relative to the GN condition.

Consider next the results from the vocoded conditions (lower row of Fig. 3). Here, the benefits of frequency selectivity have been reduced by limiting spectral resolution to the 16 vocoder channels. As expected, the MT masker produces very similar results to the GN masker, as both produce very similar outputs from the tone vocoder. The young NH and the age-matched NH groups seem able to take similar advantage of the lack of inherent masker fluctuations in the PT condition. In contrast, the differences between the PT and MT seem less pronounced in the HI listeners; although some differences remain (in contrast to CI users), they are smaller than in the NH listeners.

The results using 85-dB SPL speech for all listeners are shown in Fig. 4. Only three speech-to-masker ratios (-5 , 0 , and 5 dB) were tested in this part of the experiment. The relevant data from the HI listeners are simply replotted from Fig. 3. Performance was generally worse overall for both the young and age-matched NH

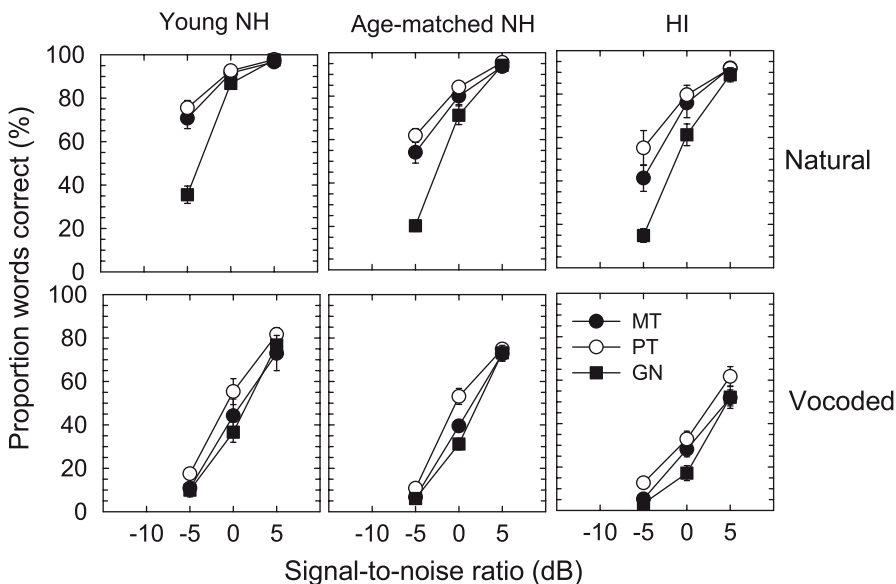


Fig. 4 Speech intelligibility with the speech at equal sound pressure level (85 dB SPL). Data from the HI listeners are replotted from Fig. 3

listeners, compared to results at the lower sound level. In addition, the difference between the PT and MT conditions was reduced. This reduction was observed even in the non-vocoded conditions, but was particularly apparent in the vocoded conditions. In fact, in the vocoded conditions, the differences between the NH groups and the HI group were relatively small.

4 Discussion

Overall, the HI group showed smaller-than-normal differences between maskers with and without inherent fluctuations. This loss of sensitivity to inherent fluctuations was particularly apparent in the vocoded conditions (Fig. 3; lower right panel). However, in contrast to earlier results from CI users (Oxenham and Kreft 2014), some differences remained between conditions with and without inherent fluctuations, suggesting that the effects of poorer frequency selectivity are not as profound as for CI users. This could be for two reasons: First, frequency selectivity in HI listeners with mild-to-moderate hearing loss is not as poor as that in CI users. Second, the difference between the interaction occurring before and after envelope extraction may affect outcomes to some extent, in that the reduction in masker modulation power may be greater for CI users than for HI listeners, even with a similar loss of frequency selectivity.

One interesting outcome was that the differences between the NH groups and the HI group were not very pronounced when the speech was presented to the groups at the same high SPL. It is well known that frequency selectivity becomes poorer in NH listeners at high levels (e.g., Nelson and Freyman 1984; Glasberg and Moore 2000). Apparently the filter broadening with level in NH listeners is sufficient to reduce the effects of inherent masker fluctuations.

Overall, the results show that the importance of inherent masker fluctuations in determining speech intelligibility in noise depends to some extent on the conditions and the listening population. It cannot be claimed that inherent masker fluctuations always limit speech perception, as the effect of the fluctuations is non-existent in CI users (Oxenham and Kreft 2014), and is greatly reduced in HI listeners and in NH listeners at high sound levels. Finally, the reduced effect of inherent fluctuations may also provide a reason for why HI listeners (as well as CI users) exhibit less masking release in the presence of maskers with imposed additional fluctuations.

Acknowledgments This work was supported by NIH grant R01 DC012262.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Dorman MF, Loizou PC, Fitzke J, Tu Z (1998) The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels. *J Acoust Soc Am* 104(6):3583–3585
- Dubbelboer F, Houtgast T (2008) The concept of signal-to-noise ratio in the modulation domain and speech intelligibility. *J Acoust Soc Am* 124(6):3937–3946. doi:10.1121/1.3001713
- Festen JM, Plomp R (1990) Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am* 88(4):1725–1736
- French NR, Steinberg JC (1947) Factors governing the intelligibility of speech sounds. *J Acoust Soc Am* 19:90–119
- George EL, Festen JM, Houtgast T (2008) The combined effects of reverberation and nonstationary noise on sentence intelligibility. *J Acoust Soc Am* 124(2):1269–1277. doi:10.1121/1.2945153
- Glasberg BR, Moore BCJ (2000) Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. *J Acoust Soc Am* 108:2318–2328
- Gregan MJ, Nelson PB, Oxenham AJ (2013) Behavioral measures of cochlear compression and temporal resolution as predictors of speech masking release in hearing-impaired listeners. *J Acoust Soc Am* 134(4):2895–2912. doi:10.1121/1.4818773
- Hu J, Beaulieu NC (2005) Accurate simple closed-form approximations to Rayleigh sum distributions and densities. *IEEE Commun Lett* 9:109–111. doi:10.1109/LCOMM.2005.02003
- Humes LE, Wilson DL, Barlow NN, Garner C (2002) Changes in hearing-aid benefit following 1 or 2 years of hearing-aid use by older adults. *J Speech Lang Hear Res* 45(4):772–782
- Jorgensen S, Dau T (2011) Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. *J Acoust Soc Am* 130(3):1475–1487. doi:10.1121/1.3621502
- Jorgensen S, Ewert SD, Dau T (2013) A multi-resolution envelope-power based model for speech intelligibility. *J Acoust Soc Am* 134(1):436–446. doi:10.1121/1.4807563
- Kryter KD (1962) Methods for the calculation and use of the articulation index. *J Acoust Soc Am* 34:467–477
- Lawson JL, Uhlenbeck GE (1950) *Threshold signals*, vol 24. McGraw Hill, New York
- Nelson DA, Freyman RL (1984) Broadened forward-masked tuning curves from intense masking tones: delay-time and probe level manipulations. *J Acoust Soc Am* 75:1570–1577
- Nelson PB, Jin SH (2004). Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. *J Acoust Soc Am* 115(5 Pt 1):2286–2294
- Oxenham AJ, Krefl HA (2014). Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing. *Trends Hear* 18. doi:10.1177/2331216514553783
- Spahr AJ, Dorman MF, Litvak LM, Van Wie S, Gifford RH, Loizou PC, Loisel LM, Oakes T, Cook S (2012) Development and validation of the AzBio sentence lists. *Ear Hear* 33(1):112–117. doi:10.1097/AUD.0b013e31822c2549
- Stickney GS, Zeng FG, Litovsky R, Assmann P (2004) Cochlear implant speech recognition with speech maskers. *J Acoust Soc Am* 116(2):1081–1091
- Stone MA, Fullgrabe C, Mackinnon RC, Moore BCJ (2011) The importance for speech intelligibility of random fluctuations in “steady” background noise. *J Acoust Soc Am* 130(5):2874–2881. doi:10.1121/1.3641371
- Stone MA, Fullgrabe C, Moore BCJ (2012) Notionally steady background noise acts primarily as a modulation masker of speech. *J Acoust Soc Am* 132(1):317–326. doi:10.1121/1.4725766
- Stone MA, Moore BCJ (2014) On the near non-existence of “pure” energetic masking release for speech. *J Acoust Soc Am* 135(4):1967–1977
- Whitmal NA, Poissant SF, Freyman RL, Helfer KS (2007) Speech intelligibility in cochlear implant simulations: effects of carrier type, interfering noise, and subject experience. *J Acoust Soc Am* 122:2376–2388
- Zeng FG (2004) Trends in cochlear implants. *Trends Amplif* 8(1):1–34