

2016

The effects of auditory contrast tuning upon speech intelligibility

Nathan J. Killian

Washington University in St. Louis

Paul V. Watkins

Washington University in St. Louis

Lisa S. Davidson

Washington University School of Medicine in St. Louis

Dennis L. Barbour

Washington University in St. Louis

Follow this and additional works at: http://digitalcommons.wustl.edu/open_access_pubs

Recommended Citation

Killian, Nathan J.; Watkins, Paul V.; Davidson, Lisa S.; and Barbour, Dennis L., "The effects of auditory contrast tuning upon speech intelligibility." *Frontiers in Psychology*.7, 1145. (2016).

http://digitalcommons.wustl.edu/open_access_pubs/5231



The Effects of Auditory Contrast Tuning upon Speech Intelligibility

Nathan J. Killian^{1†}, Paul V. Watkins^{1†}, Lisa S. Davidson² and Dennis L. Barbour^{1*}

¹ Laboratory of Sensory Neuroscience and Neuroengineering, Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, USA, ² Central Institute for the Deaf, Washington University School of Medicine, St. Louis, MO, USA

OPEN ACCESS

Edited by:

Elia Formisano,
Maastricht University, Netherlands

Reviewed by:

Maria Chait,
University College London, UK
Samuel Evans,
University College London, UK

*Correspondence:

Dennis L. Barbour
dbarbour@wustl.edu
orcid.org/0000-0003-0851-0665

† Present Address:

Nathan J. Killian,
Department of Neurosurgery,
Massachusetts General Hospital,
Boston, USA
Paul V. Watkins,
NIH/NINDS/Branch Contractor, Global
Solutions Network, Inc., Bethesda,
USA

Specialty section:

This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Psychology

Received: 30 March 2016

Accepted: 18 July 2016

Published: 09 August 2016

Citation:

Killian NJ, Watkins PV, Davidson LS
and Barbour DL (2016) The Effects of
Auditory Contrast Tuning upon
Speech Intelligibility.
Front. Psychol. 7:1145.
doi: 10.3389/fpsyg.2016.01145

We have previously identified neurons tuned to spectral contrast of wideband sounds in auditory cortex of awake marmoset monkeys. Because additive noise alters the spectral contrast of speech, contrast-tuned neurons, if present in human auditory cortex, may aid in extracting speech from noise. Given that this cortical function may be underdeveloped in individuals with sensorineural hearing loss, incorporating biologically-inspired algorithms into external signal processing devices could provide speech enhancement benefits to cochlear implantees. In this study we first constructed a computational signal processing algorithm to mimic auditory cortex contrast tuning. We then manipulated the shape of contrast channels and evaluated the intelligibility of reconstructed noisy speech using a metric to predict cochlear implant user perception. Candidate speech enhancement strategies were then tested in cochlear implantees with a hearing-in-noise test. Accentuation of intermediate contrast values or all contrast values improved computed intelligibility. Cochlear implant subjects showed significant improvement in noisy speech intelligibility with a contrast shaping procedure.

Keywords: auditory cortex, noise reduction, human, cochlear implant, primate

INTRODUCTION

Hearing-impaired individuals in general, and cochlear implantees in particular, can often comprehend speech comparably to normal-hearing listeners in the absence of noise, but then exhibit poorer-than-normal performance in noisy environments (Friesen et al., 2001; Nie et al., 2005). In patients with sensorineural hearing loss, this outcome is due at least partly to the decreased perceived spectral contrast of speech sounds associated with diminished frequency selectivity in the auditory periphery (Leek et al., 1987; ter Keurs et al., 1993; Summers and Leek, 1994; Leek and Summers, 1996; Dreisbach et al., 2005). Spectral contrast represents the variation in sound energy distribution across frequency. Neuronal processing that utilizes spectral contrast to reduce noise is likely to be impaired in these individuals. Research into artificially compensating for spectral contrast deficits has resulted in a class of algorithms that perform a global spectral contrast enhancement. These algorithms represent signal processing strategies where peaks in the short-term signal spectrum are stretched relative to the troughs (i.e., the spectral variance is uniformly increased as a function of frequency). Previous studies evaluating the ability of spectral contrast enhancement to improve the intelligibility of noisy speech have demonstrated some potential benefit of this noise-reduction strategy for cochlear implant users and variable results for other subjects (Bunnell, 1990; Simpson et al., 1990; Stone and Moore, 1992; Baer et al., 1993; Ribic et al., 1996; Munoz et al., 1999; Loizou and Poroy, 2001; Lyzenga et al., 2002; Yang et al., 2003).

Contrast enhancement likely achieves its effect by accentuating the information-carrying spectral peaks of vocalizations relative to the nearly uniform spectrum of stationary noise. Based upon our electrophysiological findings in the primary auditory cortex of a vocal primate species, we have developed a unique, biologically-inspired approach to compensate for deficits in spectral contrast processing.

We defined spectral contrast as the standard deviation of signal intensity in dB over short periods of time and relatively local ranges of frequency. Speech corrupted with additive white, Gaussian noise shows a considerable variety of spectral contrast values (**Figure 1**). Spectrotemporal signal regions of low contrast are likely to contain relatively more noise power than speech power; conversely, regions of high contrast are likely to have a greater contribution from the speech to the overall signal. Intermediate contrast values are characteristic of regions with a mixture of speech and noise at similar powers. Contrast enhancement algorithms do not explicitly compute spectral contrast and are thus applied uniformly to low-, intermediate-, and high-contrast regions. Emphasizing the low-contrast regions, however, would tend to accentuate the noise, while emphasizing the high-contrast regions would tend to distort the speech—either of which would likely decrease intelligibility. Furthermore, compressing low-contrast regions or preferentially emphasizing regions of intermediate contrast could aid a listener in discerning the differences between the noise and the speech, potentially leading to improved speech intelligibility. In this study, we expand upon previous contrast enhancement studies by systematically exploring the effects of enhancing or compressing specific contrast bands. We find that emphasizing high contrasts or low contrasts alone or in combination degrades the intelligibility of noisy speech in a variety of noise conditions. We also examine the effects of preferential emphasis of intermediate contrasts and enhancement of all contrast components.

Our motivation for decomposing stimulus contrast into separate bands comes from electrophysiological findings. We have identified neurons in the auditory cortex of a vocal primate species that encode spectral contrast non-linearly, such that their peak firing rate as a function of spectral contrast occurs at intermediate contrast values (Barbour and Wang, 2003). Among these relatively common contrast-tuned neurons, the peak of the distribution of preferred contrasts is in the range of 4–8 decibels of standard deviation (dB SD). Because regions of intermediate contrast represent a likely target for differentiating between speech and noise (**Figure 1**, middle), these neurons may be useful to the auditory system for extracting signal information (such as vocalization features) in the presence of noise. Based upon these neurophysiological findings, we created a computational signal processing array containing contrast-sensitive elements in order to mimic the observed neuronal contrast tuning. By analogy with the type of transformations in which contrast-tuned neurons appear to be involved, we then utilized this signal processing array to shape the spectral contrast of noisy speech. We evaluated the effects of this contrast shaping procedure on the perception of simulated cochlear implantees by calculating intelligibility of the processed speech relative to unprocessed speech. The quantitative measure we used to evaluate the result

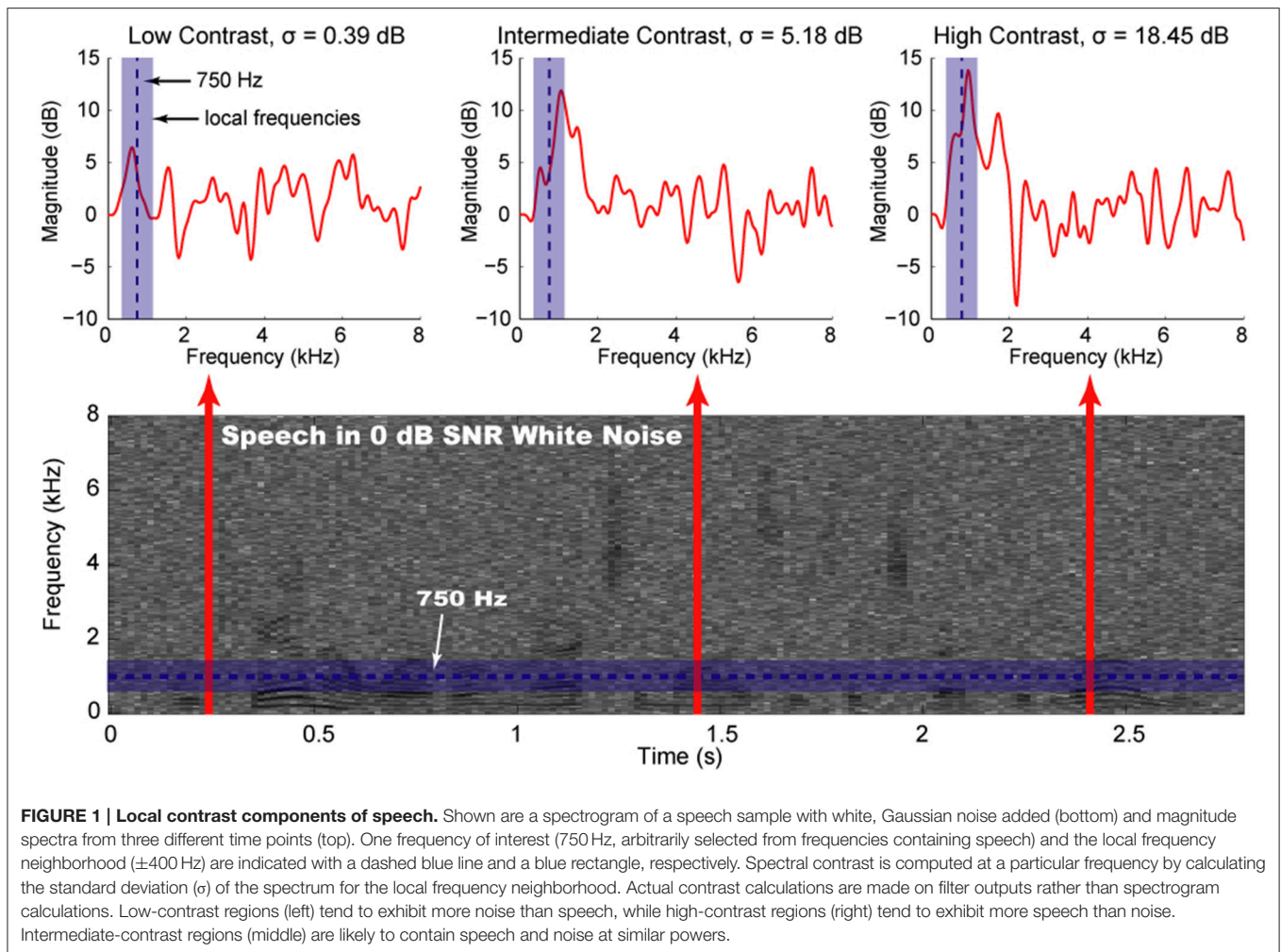
was shown previously to be appropriate for predicting noisy speech intelligibility following non-linear operations with both normal-hearing listeners and cochlear implant users as test subjects (Goldsworthy, 2005). The results indicated that contrast-shaping strategies may provide benefits to cochlear implant users listening to speech in noise. Noise-reduction strategies often show little benefit for non-implantees, even for individuals with considerable hearing loss (Kuk et al., 1990; Moore, 2003). Thus, we sought to test noise reduction in the population most likely to derive benefits: cochlear implantees. We tested one promising strategy by administering a hearing-in-noise test (HINT) to 5 cochlear implantees. Improved speech recognition in noise was seen, supporting the computational results.

MATERIALS AND METHODS

Our methods can be broken down into distinct steps: (i) signal generation followed by preprocessing and application of contrast filters; (ii) weighting of the contrast filter outputs; (iii) reconstruction of the signal; and (iv) evaluation of the intelligibility of the reconstructed signal with either computational metrics or cochlear implantees. All computations in this study were performed using custom code written for MATLAB software.

Signal Generation

We obtained pre-recorded sentences spoken by native English speakers from the TIMIT speech corpus (used in computational analyses) and the HINT sets (used in both computational analyses and in cochlear implant testing) (Garofolo et al., 1993; Nilsson et al., 1994). The signals were degraded using additive noise that was pseudorandomly generated and scaled in the time-domain to be at the desired signal-to-noise ratio (SNR)—computed using the root mean square values of the separate speech and noise signals. The noise was separately computed to have desired characteristics and then added to the speech to create the sounds in the same form as they would be presented to a listener. Samples of noisy and processed noisy speech are included in the online Supplementary Materials. Sound levels in the supplementary audio files were normalized to the maximum level across all conditions at each SNR. Noise was generated by specifying the shape of the power spectrum up to the Nyquist frequency, combining this with phase values pseudorandomly sampled from a uniform distribution on 0 to 2π radians, making the frequency domain signal symmetric around the Nyquist frequency, and finally taking the inverse Fourier transform to obtain the noise in the time domain. White noise had a uniform power spectrum, pink noise had a $1/f$ spectrum, and spectrally-shaped noise had a spectrum that matched the speech. For the filter-bank processing, signals may be either normalized or unnormalized as it does not affect the computations performed on decibel values. In the cochlear implantee tests, no normalization was performed so that speech would be at the same level in the noisy and quiet conditions. However, to eliminate any bias due to changes in signal power after the noise-reduction strategies were applied, noisy speech that had processing applied was scaled to have the same overall signal power as the noisy speech without processing.



Contrast-Shaping Algorithm and Experimental Design

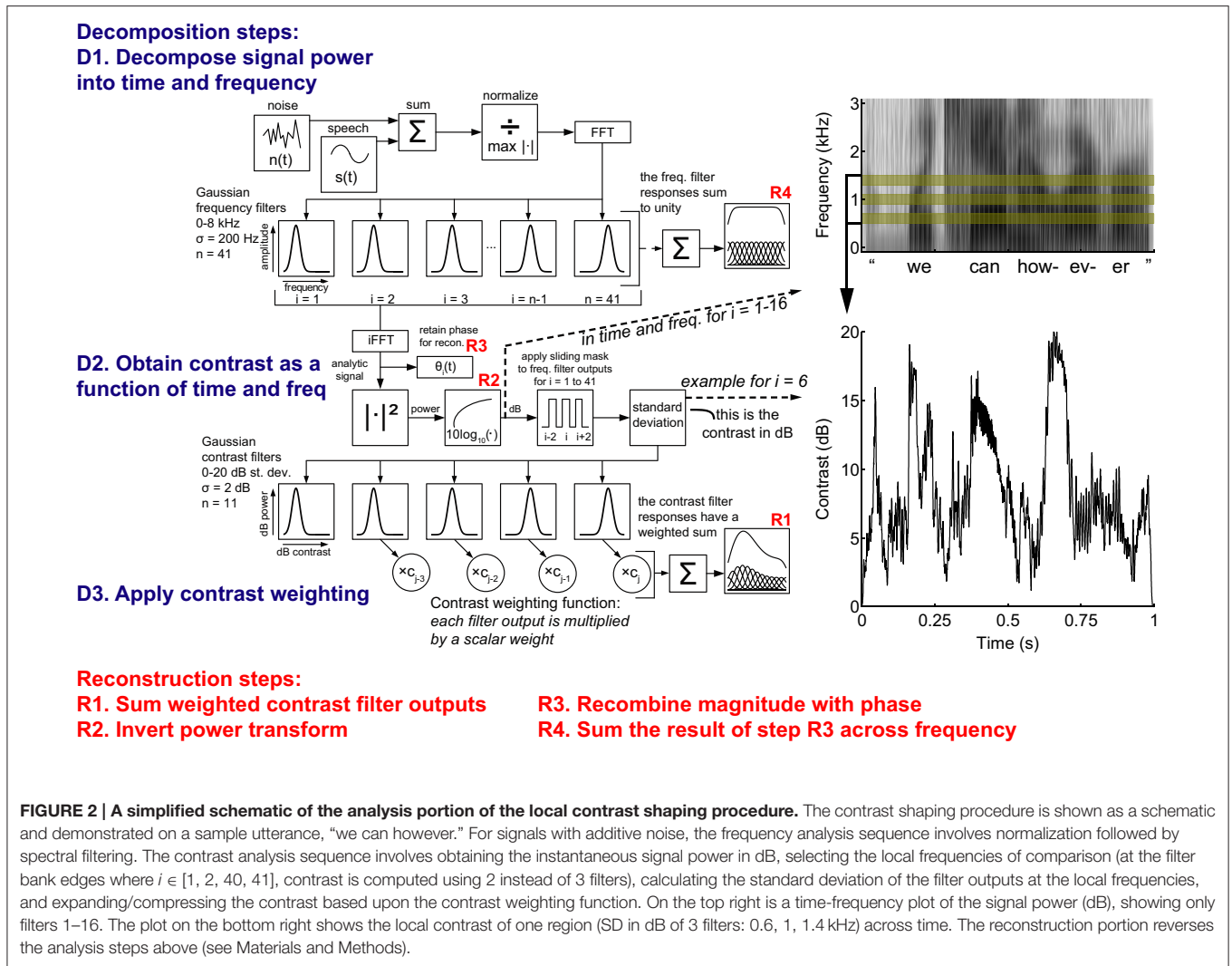
Contrast shaping was performed after filtering the noisy speech signals into different frequency (spectral) components and then into components associated with contrast across subsets of frequencies (spectral contrast). In other words, the contrast-shaping algorithm (Figure 2) contained two basic components of sound decomposition—frequency and contrast. Both of these components were designed to be reversible for reconstruction of the original time-domain signal. Both frequency and contrast decompositions were performed by filtering with an array of squared exponential, i.e., Gaussian, filters. For computational efficiency and so that the original signal could be reconstructed by summing the overlapping time segments, signals were windowed in time with 513-sample Hanning windows, overlapping by 257 samples. Sounds were sampled at 16 k samples/s, so that each time window duration was approximately 32 ms. Frequency and contrast decompositions were computed for each time window.

The frequency decomposition layer of the contrast-shaping algorithm partitioned the signal into linearly spaced frequency bands by transforming the time-domain signal into the frequency domain with a Fourier transform (using a Fast Fourier Transform

algorithm) and multiplying the frequency-domain signal by the frequency response of each Gaussian frequency filter. These filters are given by Equation (1), where m is the number of frequency filters, the X_i are the frequency responses of the filters, and the μ_i are the means of the filters.

$$X_i(\omega) = e^{-\frac{(\omega - \mu_i)^2}{2\sigma^2}} \quad \sigma = \mu_{i+1} - \mu_i \quad i = 1 \dots m \quad (1)$$

Forty-one Gaussian frequency filters ($\sigma = 200$ Hz) linearly spaced from 0 to 8 kHz were used. The filters were equally spaced and the standard deviations set to the constant difference between the means so that the original signal could be reconstructed simply by summing the filter outputs (Theunissen and Doupe, 1998). We selected 41 linearly spaced Gaussian frequency filters empirically, as speech intelligibility scores for this arrangement were generally similar to or superior than intelligibility calculated when using other filter bank arrangements, such as logarithmically spaced triangular filters. The formulas presented in Equation (2) summarize the information retained from the frequency filters. The frequency response of each Gaussian filter was not symmetric, as the negative frequencies were set to zero. An inverse Fourier



transform magnitude of each filter output is analogous to the amplitude envelope of that particular frequency band, and the power (squared amplitude envelope) of each of the 41 such signals, f_i , was used as an eventual input into the contrast decomposition layer. The phase, θ_i , of each analytic signal was retained for subsequent signal reconstruction. The input to the contrast layer was converted to dB power by applying Equation (3), yielding the f_dB_i , which will be referred to collectively as the frequency bands.

$$f_i(t) = |2FFT^{-1}(FFT(x(t))X_i(t))|^2$$

$$\theta_i(t) = \angle 2FFT^{-1}(FFT(x(t))X_i(t)) \quad (2)$$

$$f_dB_i(t) = 10 \log_{10}(f_i(t)), \quad (3)$$

The contrast decomposition layer of the contrast-shaping algorithm acted upon each frequency band and its neighboring frequencies. In other words, the local spectral contrast centered on each frequency band was calculated separately. Contrast was calculated as the standard deviation of a local subset of

the frequency bands, which included the frequency band being decomposed and the second neighbor at both higher and lower frequencies, for a total of three frequency bands over which the standard deviation was computed. The equations in (4) describe this calculation, where the c_i are the contrasts centered at each frequency band, m is the number of frequency bands, p is the number of frequency bands used to compute local contrast, and the f_dB_i are the frequency bands themselves.

$$c_i(t) = \sqrt{\frac{1}{p-1} \left(\sum_k (f_dB_k(t) - \overline{f_dB_i(t)})^2 \right)}$$

$$\overline{f_dB_i(t)} = \frac{1}{p} \sum_k f_dB_k(t) \quad (4)$$

$$k = i - 2, i, i + 2 \quad p = 3 \quad i = 1 \dots m$$

Although the subset of frequencies used to compute local contrast could be the focus of further investigations, preliminary results making use of different subsets (i.e., the local frequency mask

in **Figure 2**, indexed by k in above equations) did not yield systematically improved intelligibility scores. These standard deviations were transformed by a set of Gaussian-shaped contrast channels linearly spaced in dB of standard deviation (dB SD), the units of our contrast measure. This Gaussian transformation described in Equation (5) shared the same form as the frequency filters described in Equation (1). Here m is the number of frequency bands, n is the number of contrast channels, the $S_{i,j}$ are the contrast responses of each contrast channel (indexed by j) in each frequency band (indexed by i), and the μ_j are the means of the contrast channels.

$$S_{i,j}(c_i(t)) = e^{-\frac{(c_i(t) - \mu_j)^2}{2\sigma^2}} \quad \sigma = \mu_{j+1} - \mu_j \quad (5)$$

$$j = 1 \dots n \quad i = 1 \dots m$$

Eleven contrast channels ($\sigma = 2$ dB SD) with means linearly spaced from 0 to 20 dB SD were used. As with the frequency filters, the contrast channels were equally spaced and the standard deviations set to the constant difference between the means so that the original signal could be reconstructed by summing the channel outputs (Theunissen and Doupe, 1998). Practical compromises with computational resources led to a maximum of 11 contrast channels. The outputs of each contrast channel j were then multiplied by the respective frequency band i (the dB power output from the respective frequency filter) to obtain a set of contrast channel outputs for each frequency band, $s_{i,j}$ (Equation 6).

$$s_{i,j}(t) = f_{dB_i}(t) S_{i,j}(t), \quad (6)$$

The original time-domain signal was reconstructed by inverting the above steps for the contrast and frequency decompositions. Before reconstructing the original signal, we multiplied the contrast outputs by scalar weights, using a predetermined weight w_j for each contrast channel j . This step, given by Equation (7), where the $s'_{i,j}$ are the weighted outputs of the contrast channels, is referred to as the contrast-shaping step, and was added in order to emphasize or de-emphasize particular contrast ranges of interest.

$$s'_{i,j}(t) = w_j s_{i,j}(t), \quad (7)$$

The weights were applied directly to the log-transformed values, resulting in a non-linear effect. Because our contrast-shaping algorithm was invertible, if all contrast outputs received a unity weight, the reconstructed time-domain signal was virtually identical to the input signal. Non-unity weights, even when uniformly applied, have non-linear effects. These contrast-shaping manipulations would typically lead to an increase in local spectral contrast when a contrast channel was weighted > 1 and a decrease when a contrast channel was weighted < 1 .

Reconstructed Signal Metrics

We related the intelligibility of the speech-in-noise (either before or after processing was applied) to clean speech using an approach that provides a measure of the correlation between the two signals. The normalized correlation metric (NCM)

provided quantitative estimates of intelligibility of the processed signal (Goldsworthy, 2005). The NCM calculates a transmission index by a normalized correlation ρ^2 between the envelopes of two signals (either clean and degraded signals or clean and processed) for a fixed number of frequency bands, N . Each of these N sub-band transmission indices is multiplied by a weight w_i , determined by psychophysically evaluating the frequencies determined to be most important for speech intelligibility. Summing these values yields an NCM score between 0 and 1 (**Figure 3**). The NCM accounts for average envelope power in addition to temporal envelope fluctuations. It represents a metric empirically determined to be suitable for non-linear speech processing, such as in cochlear implants (Goldsworthy and Greenberg, 2004).

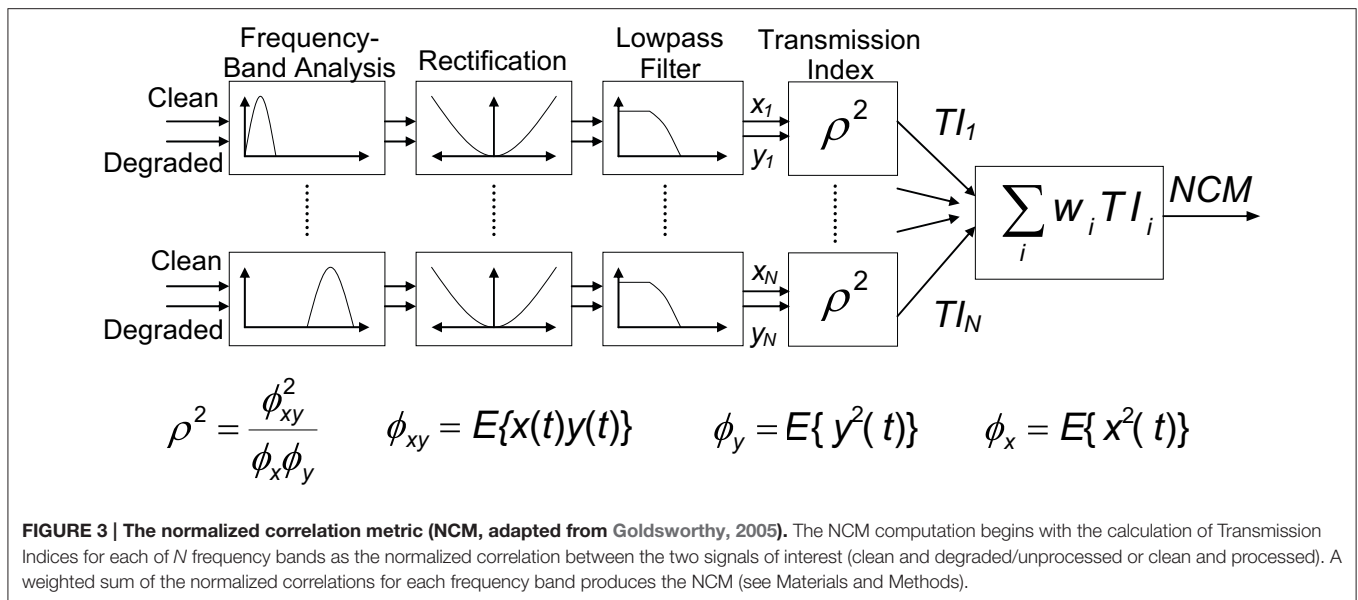
Mutual information (MI) (Cover and Thomas, 1991), specifically normalized MI (NMI) (Strehl and Ghosh, 2003), provided quantitative estimates of relative information content between noisy and clean speech independent of auditory system properties. Although contrast shaping cannot increase MI (as that would violate the data processing inequality, Cover and Thomas, 1991), MI is still informative because (i) finding changes analogous to NCM changes may support the generality of contrast shaping, and (ii) the results of contrast manipulations can be compared using the degree of signal degradation (the magnitude of the decrease in NMI after contrast shaping is applied). NMI was computed from Equation (8), where X and Y are the signals of interest (e.g., clean speech sample and contrast-shaped noisy speech), $p(x)$ and $p(y)$ are the amplitude histograms of the signals and $p(x,y)$ is the joint amplitude histogram of the two signals. MI represents the sum of the marginal entropies of the signals, minus their joint entropy. While the MI is non-negative, it does not necessarily have an upper bound, making direct comparisons of information content among multiple experimental conditions challenging. The NMI, however, takes on values between 0 and 1, with a value of 0 for two statistically independent signals and a value of 1 for two identical signals. We selected this method of normalizing MI so that the metric is guaranteed to be upper bounded by 1, since the mutual information between a signal and itself is the entropy of the signal. Because the NMI uses each sample and the full range of each signal to estimate the probability distributions, sharp peaks of extreme high or low values may bias the result. To remove this bias, we used the signal envelopes at a resolution of N levels (i.e., the probability distributions were estimated with N bins).

$$MI(X, Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}$$

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x) \quad (8)$$

$$NMI(X, Y) = \frac{MI(X, Y)}{\sqrt{H(X)H(Y)}}$$

Both the NCM and NMI were computed with $N = 24$, unless otherwise indicated. The signal envelopes were low-pass filtered with a cutoff frequency of 50 Hz as an initial preprocessing step. Thus, for both metrics we used slow fluctuations in the signals



as input, but the metrics differed in how they used these values; the NCM used the correlation of envelopes in N frequency bands, weighted based on their importance in human speech recognition, and the NMI used the input signals (here we used the broadband envelopes) at N level bins with no assumptions about which signal components are important for speech recognition.

Systematic Searches of the Contrast Weight Space

We investigated the effects on speech intelligibility of performing manipulations such as enhancing or compressing distinct spectral contrast components. The possible weights on binned contrast (the contrast weight space) were explored in an unbiased, systematic manner. The contrast weight space was first explored in detail on a single TIMIT speech utterance (“We can, however, maximize the expected value”). Because the detailed search uses a great deal of computational resources, we performed this only on a single sentence. However, as drawing conclusions based upon this analysis alone might be confounded by overfitting, we then identified principal contrast regions of interest and evaluated them on a set of sentences from the HINT database.

The detailed systematic search of the contrast weight space was performed by iterating the contrast-shaping algorithm over all possible permutations of the weights 0.5 (compression), 1 (no change), and 2 (enhancement) for the 11 contrast channels (3^{11} weighting schemes). All weights were plotted as log base 2 values, falling within the range $[-1, 1]$. Images of the systematic search results for the contrast weight space were created by sorting the contrast channel weight vectors by the computed NCM score. For visualization, these sorted weight vectors were then interpolated linearly in two dimensions at 1025 points uniformly spaced across each dimension width, between 0 and 20 dB contrast for the contrast weight dimension (abscissa)

and between the minimum and maximum metric value for the metric dimension (ordinate). This 2D space was then smoothed with a Gaussian kernel whose standard deviations for both dimensions were set to 1/100 of the dimension range.

By visual examination of these images, specific contrast regions yielded clustered effects on intelligibility when weighting in a certain manner such as compression or enhancement. Based upon this analysis we used three contrast regions for subsequent analyses: low contrasts, the 0 and 2 dB bins; intermediate contrasts, the 4, 6, and 8 dB bins; and high contrasts, bins covering 10–20 dB. We again explored permutations of the logarithmically spaced weights on these three contrast regions ($3^3 = 27$ weighting schemes). As this is computationally tractable, we performed this analysis on a set of 10 sentences from the HINT database (these sets were later used for testing with cochlear implantees).

Psychophysical Testing of Algorithms in Cochlear Implantees

All human testing procedures were approved by the IRB at Washington University with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. After providing written informed consent, five subjects with unilateral cochlear implants participated in this study. Etiology of hearing loss was most often reported as progressive ($n = 3$) with one subject experiencing a sudden hearing loss likely due to autoimmune disease and one subject reporting genetic hearing loss. All subjects used spoken English as their primary mode of communication. The mean length of implant use was 5.6 years and ranged from 2 to 11 years. The mean age at test was 69.4 years and ranged from 55 to 79 years. All subjects had full insertion of the electrode array and used a Cochlear Nucleus implant system

(i.e., N22, N24). Speech recognition scores on the Consonant-Nucleus-Consonant (CNC) (Peterson and Lehiste, 1962) in quiet ranged from 43 to 84% correct. Demographic information for the subjects is provided in **Table 1**.

All testing was performed in a single-walled test booth typically used for clinical audiology research. Loudspeakers were positioned at 0° azimuth and placed one meter from the subjects. Calibration for the audiometer and binaural sound-field warble tones was conducted using a Type 1 sound level meter and ANSI Standards for Audiometers (ANSI S3.6-1996). All speech and noise stimuli were stored on a Dell Desktop computer. The computer was used to deliver the speech stimuli via an audiometer and loudspeaker. The HINT sentences were presented at a level of 65 dB SPL (slow RMS, C-weighting). The presentation level of the sentences was calibrated using a Type 2 sound level meter (Quest) and a 1000 Hz calibration tone.

Aided sound-field detection thresholds were obtained to assure audibility using frequency modulated (FM) stimuli at the following frequencies 250, 500, 1000, 2000, 3000, 4000, and 6000 Hz. A standard modified Hughson-Westlake procedure (Jerger et al., 1959) in increments of 2 dB was used to obtain thresholds. All subjects had aided thresholds at 30 dB HL or better from 250 to 6000 Hz.

Hearing in Noise Test (Nilsson et al., 1994) sentences were pre-recorded in the presence of stationary white noise at two signal-to-noise ratios (SNR, +5 and +10 dB). All sentences at each SNR were processed offline both with and without the noise-reduction strategy. Altogether, 16 sentence lists were used for the study with sentences within each list randomized across 5 test conditions (quiet, +5, +10 dB SNR unprocessed, and +5, +10 dB SNR with flat weights).

The HINT sentences were presented at an overall level of 65 dB SPL. Sentences in quiet and at each SNR for each condition (noise reduction/non-noise reduction) were presented randomly to each subject. All subjects used maps and processor settings (volume and sensitivity) that were worn in everyday listening conditions. After hearing each sentence, the subjects were asked to repeat the sentence back to the examiner. The number of words correctly repeated was recorded and the overall percent accuracy was each subject's speech recognition performance.

Twenty sentences were presented at each of the five conditions for a total of 100 sentences. After a rest period, the subjects were presented with an independent set of 100 sentences. The NCM was used to predict performance using 260 sentences in the HINT database, comparing NCM*100 with percent correct speech recognition performance. This excludes prediction of quiet speech without additive noise or other processing where $NCM = 1$ (100% correct).

RESULTS

Contrast Shaping Algorithm Applied to Noisy Speech

The overall spectral contrast distribution of the TIMIT speech utterance ("We can, however, maximize the expected value"), as determined by the output of the contrast elements in the signal processing array, peaked at around 5 dB SD of contrast (**Figure 4**). When white, Gaussian noise (0 dB SNR) was added, this curve adjusted toward lower contrast values, such that more portions of the noisy signal had lower contrast values and fewer had higher contrast values. The analytical portion of the signal processing array therefore outputs signal characteristics consistent with intuition and demonstrates that much of the dynamic range of local spectral contrast tends to be at intermediate contrast values. Subsets of trials run on other speech utterances and in other conditions verified the generality of the observations made with this speech sample.

We hypothesized based upon our previous neurophysiological findings that emphasizing mainly the intermediate contrast values could result in a substantial increase in speech intelligibility. To test this hypothesis, we modified noisy speech with the signal processing array. Signals were reconstructed from the outputs of 11 contrast-sensitive elements centered from 0 dB SD to 20 dB SD (see Materials and Methods). For a single noisy speech utterance, we systematically searched the space of all possible combinations of accentuation ($\log_2(\text{weight}) > 0$), no change ($\log_2(\text{weight}) \approx 0$) and attenuation ($\log_2(\text{weight}) < 0$) of each contrast element output by measuring the NCM of the reconstructed signal as a computational estimate of speech intelligibility (see Materials and Methods).

TABLE 1 | Subject demographic information.

Sex	Etiology	Age at study (years)	Length of CI use (years)	CI system	Processor	Implanted ear	CI coding strategy	CNC at test
M	Progressive	57	2	Nucleus 24	Freedom	L	ACE	43
F	Progressive	79	5	Nucleus 24	Sprint	R	ACE	60
F	Sudden, unknown	77	6	Nucleus 24	Freedom	R	ACE	62
F	Progressive	79	4	Nucleus 24	Freedom	L	ACE	62
M	Genetic	55	11	Nucleus 22	3G	R	SPEAK	84

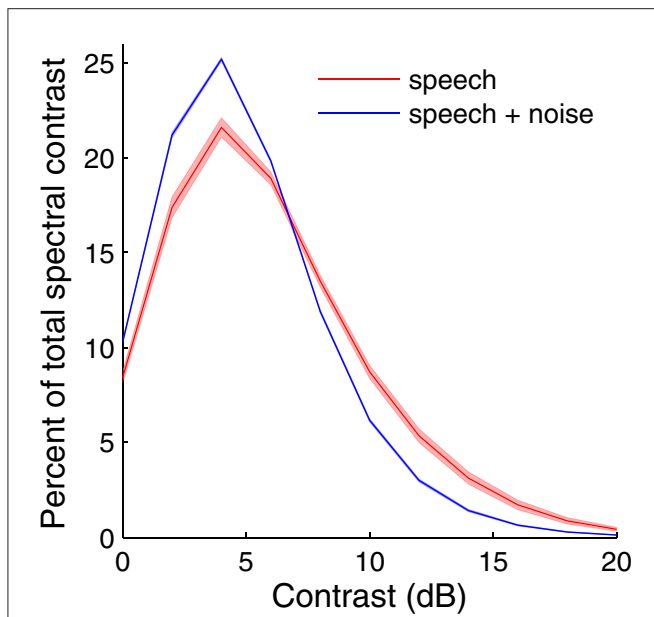


FIGURE 4 | Overlay of contrast distributions as calculated by the signal processing array for 10 sentences. The overall spectral contrast distribution of a speech utterance indicates that most of the contrast values fall into the range of 4–8 dB. When the same speech utterance was corrupted with white, Gaussian noise at 5 dB SNR, the contrast distribution shifted to lower contrast values. Curves represent the mean across 10 sentences in the HINT database and the shading represents the mean \pm standard deviation at each contrast value. The curves demonstrate that intermediate contrast values are common in both clean and noisy speech. Furthermore, there is little variability in contrast across sentences in the HINT database.

The patterns of contrast element weight vectors (i.e., contrast shapes) for the speech sample corrupted by pink noise (i.e., 1/f noise), white noise, and spectrally-shaped noise at three different signal-to-noise ratios (SNR) are shown in **Figure 5**. Contrast shapes in each panel are sorted by the change in NCM score of the processed noisy speech sample relative to the unprocessed noisy speech, yielding the most intelligible samples at the top of each panel and the least intelligible samples at the bottom. For the best improvements in computed intelligibility (top of each panel), modifying the highest and the lowest contrast values appears to influence the NCM scores little (i.e., mean weights near 0), while enhancement of the intermediate contrasts centered around 5 dB SD (i.e., weights near 1) appears to yield a substantial improvement in computed intelligibility. Enhancing the highest contrasts at the expense of the intermediate contrasts substantially worsens computed intelligibility scores (bottom of each panel). The same general patterns of contrast shaping functions for improving or degrading NCM scores are apparent across all types of noise and SNRs tested, as well as sample tests with other speech utterances (data not shown).

The results shown in **Figure 5** demonstrate some clustering in effects at roughly 0–2 dB, 4–8 dB, and 10–20 dB contrast ranges. To simplify the results shown in **Figure 5**, each of these three contrast bands was weighted independently of the others with either compression (weight of -0.5) or

enhancement (weight of $+0.5$) and the resulting changes in the NCM and NMI scores were calculated. These data are displayed in **Figure 6**. As expected, compressing low contrasts or enhancing intermediate contrasts enhanced the predicted intelligibility. Expanding low or high contrasts strongly degraded performance.

When ranking the intelligibility scores for all permutations of the weights -0.5 , 0 , and 0.5 applied to these three bands, it becomes clear that enhancement of intermediate contrasts is generally beneficial regardless of the weighting applied to low and high contrasts (**Figure 7**). Furthermore, enhancing low or high contrasts is only beneficial when intermediate contrasts are also enhanced. Interestingly, if taking a linear summation approach, the results presented in **Figure 6** suggest that low compression, intermediate enhancement, and no change to high contrasts would be beneficial. While beneficial relative to no enhancement, this strategy gives on average the 8th best NCM and NMI scores (out of 27 weighting schemes). It is actually the case of enhancing all contrasts that is the optimal weighting scheme. Thus, some uniform skewing of the contrast distribution to the right produces a signal that most closely approaches the original speech signal.

To further examine the effects of flat/uniform weightings, weights from -0.25 to 1.25 were applied to noisy speech (**Figure 8**). For the sentences tested, the optimal weight (greatest relative increase in NCM score, **Figure 8A**, top) was 0.5 , which coincided with optimum intermediate contrast area (**Figure 8B**). The NMI was also computed (**Figure 8A**, bottom) and showed analogous effects: enhancement degraded the signal less than compression up to an optimal weight, and larger weights degraded the signal more. **Figure 8B** shows that as the uniform weight value was increased, the amount of low contrast decreased, the amount of high contrast increased, and the amount of intermediate contrast showed a maximum at the optimal uniform weight value. The results shown in **Figure 8** suggest that the fractional amount of different contrast regions might correlate with the intelligibility. Indeed, the fractional amount of contrast at 6 dB had a strong positive correlation with the NCM score (**Figure 9**). Other intermediate contrast values had weak positive correlations, while low and high contrast values had negative correlations (with the slight positive correlation at 10 dB being the only exception). Interestingly, this significant correlation between fractional intermediate contrast composition and intelligibility was concurrent with relatively little change in the distribution near 6 dB for any of the flat weightings (**Figures 9, 10B**). Increasing (decreasing) uniform weights resulted in right (left) shifting of the contrast distribution for all sentences tested (**Figure 10A**). At the best weighting ($+0.5$), the greatest proportional changes in the distribution (relative to no weighting) are at low and high contrasts, with little change near intermediate contrasts (**Figure 10B**).

To evaluate how these computational results compared with neurophysiological data, we compared the average contrast shapes that elicited the greatest increase in NCM scores to the neuronal contrast data. **Figure 11** shows the top 10% of NCM score improvements for the speech sample corrupted with white noise at three different SNRs. The shapes of these

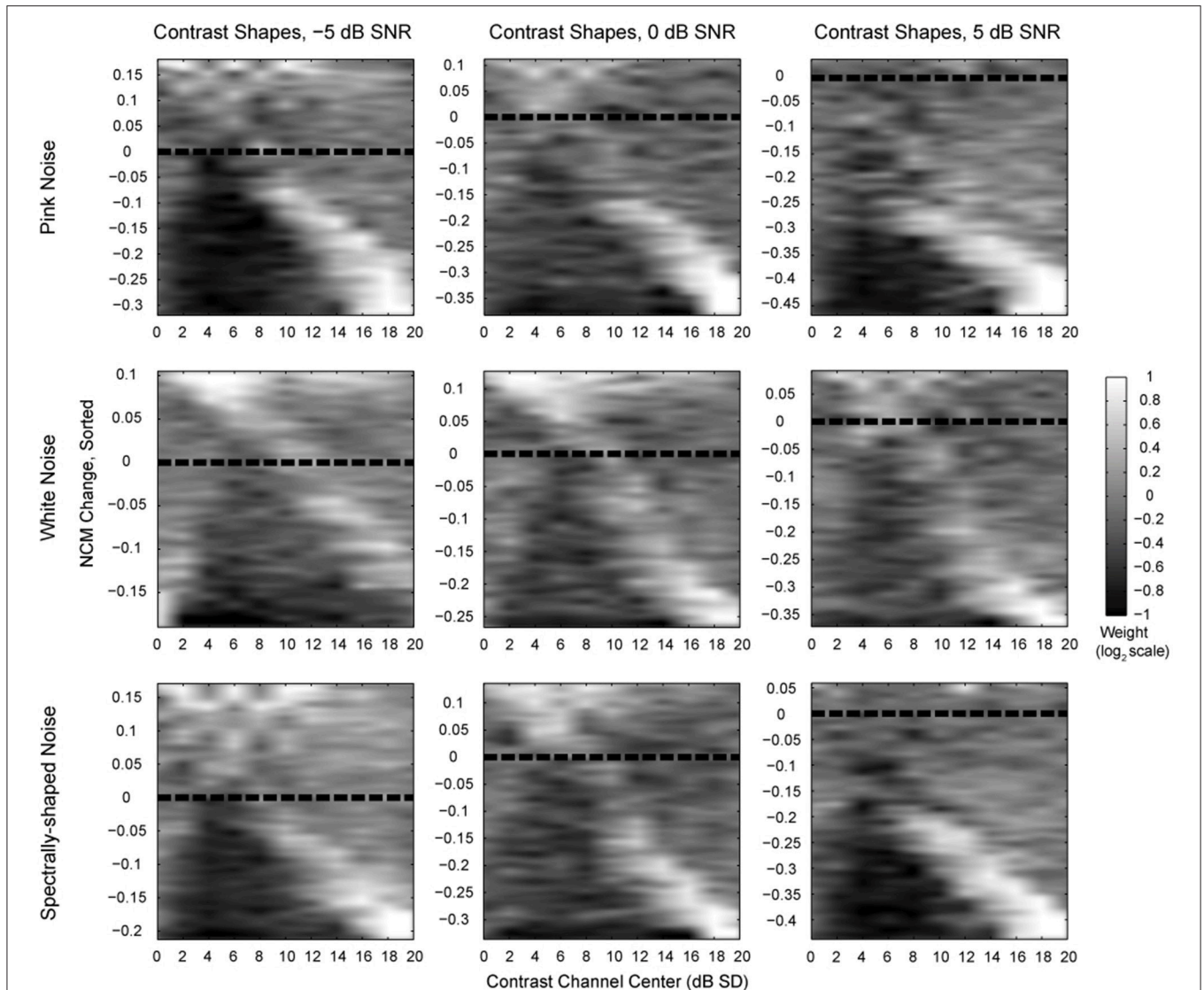


FIGURE 5 | The effects of various contrast weight vector shapes on noisy speech computed intelligibility. Each panel represents a complete systematic search of the contrast weight space with different additive noise conditions and at different SNRs. The difference in NCM scores, calculated with $N = 8$ frequency bands, between the processed and original noisy speech for each contrast weight vector shape was used to sort the contrast shapes themselves. Each systematic search consisted of all permutations for weights of 2^{-1} , 2^0 , and 2^1 on each of the 11 contrast channels. The sorted contrast weights were then interpolated and smoothed to produce the display images. Lighter values correspond to higher weights, showing a convergence on intermediate contrast values at the best NCM scores. For pink noise, improvements achieved +0.04 (+5% relative to 0.87) at 5 dB SNR, +0.11 (+15% relative to 0.75) at 0 dB SNR and +0.18 (+30% relative to 0.6) at -5 dB SNR. For white, Gaussian noise, improvements in the NCM score achieved +0.09 (+12% relative to 0.77) at 5 dB SNR, +0.13 (+20% relative to 0.64) at 0 dB and +0.11 (+21% relative to 0.52) at -5 dB. For spectrally shaped noise, improvements achieved +0.06 (+7% relative to 0.84) at 5 dB SNR, +0.14 (+20% relative to 0.69) at 0 dB SNR and +0.17 (+33% relative to 0.51) at -5 dB SNR.

curves generally match the similar curves from other noise conditions. Overlaid on these plots is the distribution of preferred contrasts across the population of contrast-tuned neurons in primate auditory cortex (Barbour and Wang, 2003). The largest percentage of contrast-tuned neurons prefers contrasts that are also the most relevant, as determined by our contrast-shaping algorithm, for improving the intelligibility of noisy speech. Note that even though a total range of 10 dB SNR is represented in the speech signals tested, the best contrast

shapes in all cases tended to peak at around 5 dB SD. The major difference between the best contrast shapes for different noise conditions appears to be predominantly the total range of contrast values emphasized rather than a shift in their peaks (i.e., lower noise levels lead to a broadened range around the peak at 5 dB). This similarity in shape underscores our empirical observation that contrast shaping can successfully improve computed intelligibility without any estimate of signal or noise properties.

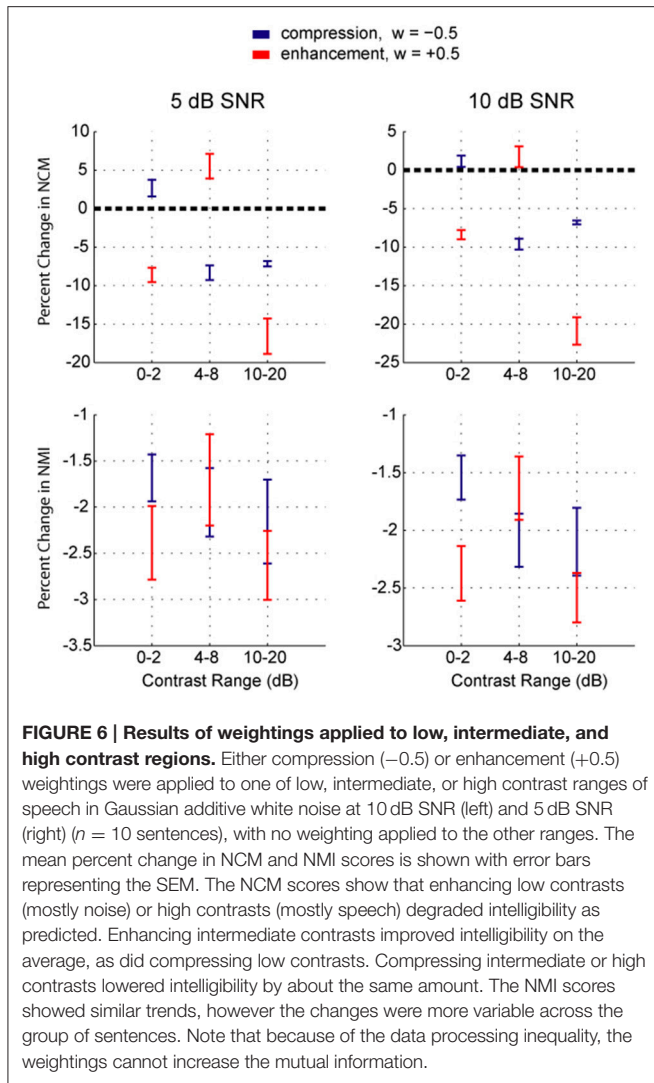


FIGURE 6 | Results of weightings applied to low, intermediate, and high contrast regions. Either compression (-0.5) or enhancement ($+0.5$) weightings were applied to one of low, intermediate, or high contrast ranges of speech in Gaussian additive white noise at 10 dB SNR (left) and 5 dB SNR (right) ($n = 10$ sentences), with no weighting applied to the other ranges. The mean percent change in NCM and NMI scores is shown with error bars representing the SEM. The NCM scores show that enhancing low contrasts (mostly noise) or high contrasts (mostly speech) degraded intelligibility as predicted. Enhancing intermediate contrasts improved intelligibility on the average, as did compressing low contrasts. Compressing intermediate or high contrasts lowered intelligibility by about the same amount. The NMI scores showed similar trends, however the changes were more variable across the group of sentences. Note that because of the data processing inequality, the weightings cannot increase the mutual information.

Evaluation of Contrast Shaping Effectiveness in Cochlear Implantees

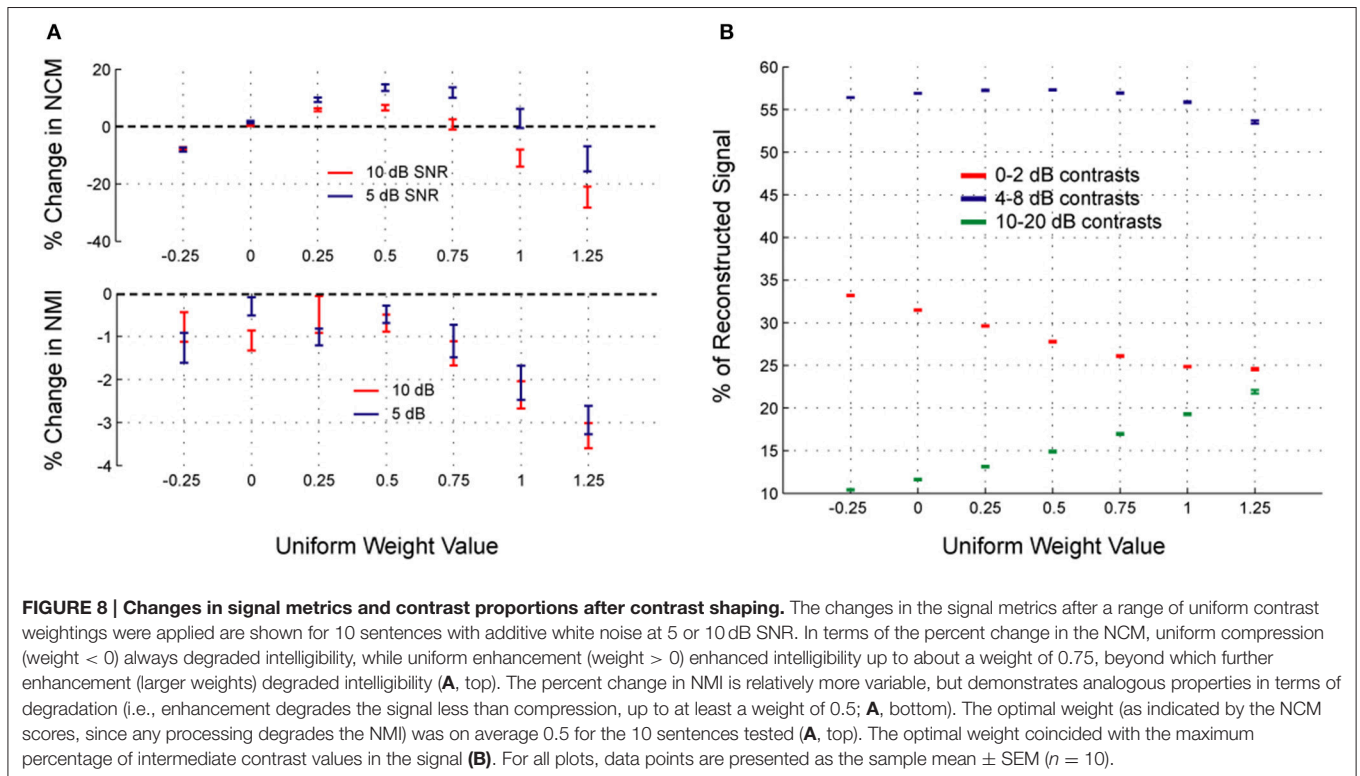
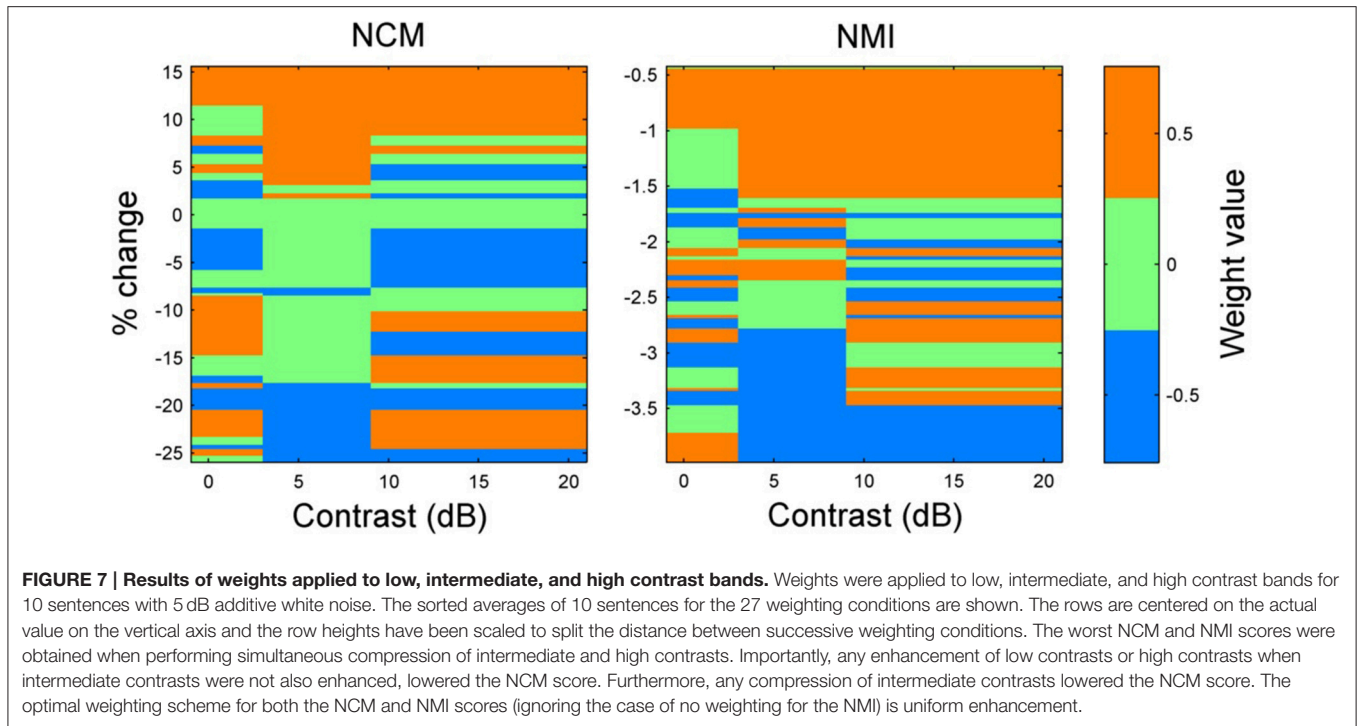
To evaluate the effectiveness of our contrast filter bank and weighting methods, we tested the best weighting scheme from simulations ($+0.5$ uniform weight) on 5 cochlear implant users. The subjects exhibited systematically degraded speech recognition performance as noise was added in a hearing-in-noise test (Figure 12). Contrast shaping that uniformly accentuated all contrasts between 0 dB and 20 dB (i.e., with a flat contrast shaping function) resulted in considerable predicted and actual intelligibility improvement (10 dB SNR: $+7\%$ predicted, $+6\%$ actual; 5 dB SNR: $+9\%$ predicted, $+15\%$ actual). Predicted performance was within 1% of actual performance on average for processed speech in noise and within 10% on average for noisy speech. The average improvement relative to the performance on speech in noise was greater for the 5 dB SNR condition (24%) than for the 10 dB condition (7%). The variable effects of degradation by the two SNR levels and performance of individual subjects on both test sessions are depicted in Figure 12B.

Hearing performance was analyzed with logistic regression using the following predictors: noise (5 or 10 dB SNR), contrast shaping, and test order. By performing a chi-squared test on the reduction in deviance with the inclusion of each predictor, we found that lower SNR was correlated with decreased performance [$\chi^2(1) = 16, p = 5.9 \times 10^{-5}$], as expected. There was no effect of the first or second set of sentences on performance [$\chi^2(1) = 0.19, p = 0.67$], suggesting that subjects did not have any learned changes in their ability to decode speech in noise over the course of the experiment. Finally, as predicted, contrast shaping was correlated with improved speech recognition performance [$\chi^2(1) = 11, p = 8.4 \times 10^{-4}$]. To examine the interaction between processing and noise level, the product of values representing noise level and contrast shaping was included as a predictor. The inclusion of this additional predictor did not result in statistically significant improvements in the model fit [$\chi^2(1) = 1.4, p = 0.23$], but noise level and contrast shaping remained significant predictors ($p < 0.05$). Given the small sample size of five cochlear implant users, further verification of these results with a larger sample of cochlear implantees is necessary.

DISCUSSION

Contrast Shaping Algorithm Applied to Noisy Speech

The contrast distribution of clean speech, as determined by the outputs of the contrast elements in our contrast-shaping algorithm, peaks at intermediate contrast values (Figure 4). This peak shifts toward lower contrast values when noise is added to the speech. If contrast-tuned neurons do aid in coding vocalizations or other sounds in the presence of noise, one might therefore expect that intermediate contrast values around 5 dB (the peak in the contrast-tuned neuronal population) could be particularly important for speech intelligibility. Stemming from recognition that lower spectral contrast results in lowered speech intelligibility, contrast enhancement algorithms attempt to shift the contrast distribution of noisy speech toward higher contrasts. This procedure can yield a higher proportion of higher contrast regions in the enhanced noisy signal—a result that tends to improve speech intelligibility. Using our contrast-shaping algorithm we found that a substantial increase in computed noisy speech intelligibility can potentially be obtained by accentuating primarily the intermediate contrasts (4–8 dB SD) (Figures 5–7). Accentuating the lowest contrasts (<4 dB SD) or the highest contrasts (>8 dB SD) alone consistently degraded computed intelligibility. Compression of low contrasts enhanced computed intelligibility whereas compression of intermediate or high contrasts degraded intelligibility. On average, applying a uniform enhancement yields the best computed intelligibility (Figure 7). Interestingly, over a range of uniform weighting schemes, the fractional amount of intermediate contrast in the signal is simultaneously linearly correlated with computed intelligibility (Figure 9) and changes relatively little across that range (Figures 9, 10). Thus, the optimal contrast-shaping strategy decreases the relative amounts of low contrast components, and



increases the relevant amounts of intermediate and high contrast components. These trends were consistent for the different noise types and SNRs tested. The greatest magnitude of computed intelligibility improvements occurred at the lowest SNRs tested,

implying that contrast manipulations may be more effective for improving speech intelligibility under conditions of greater noisy interference. While we did not identify a significant interaction between SNR and contrast shaping, the results from testing

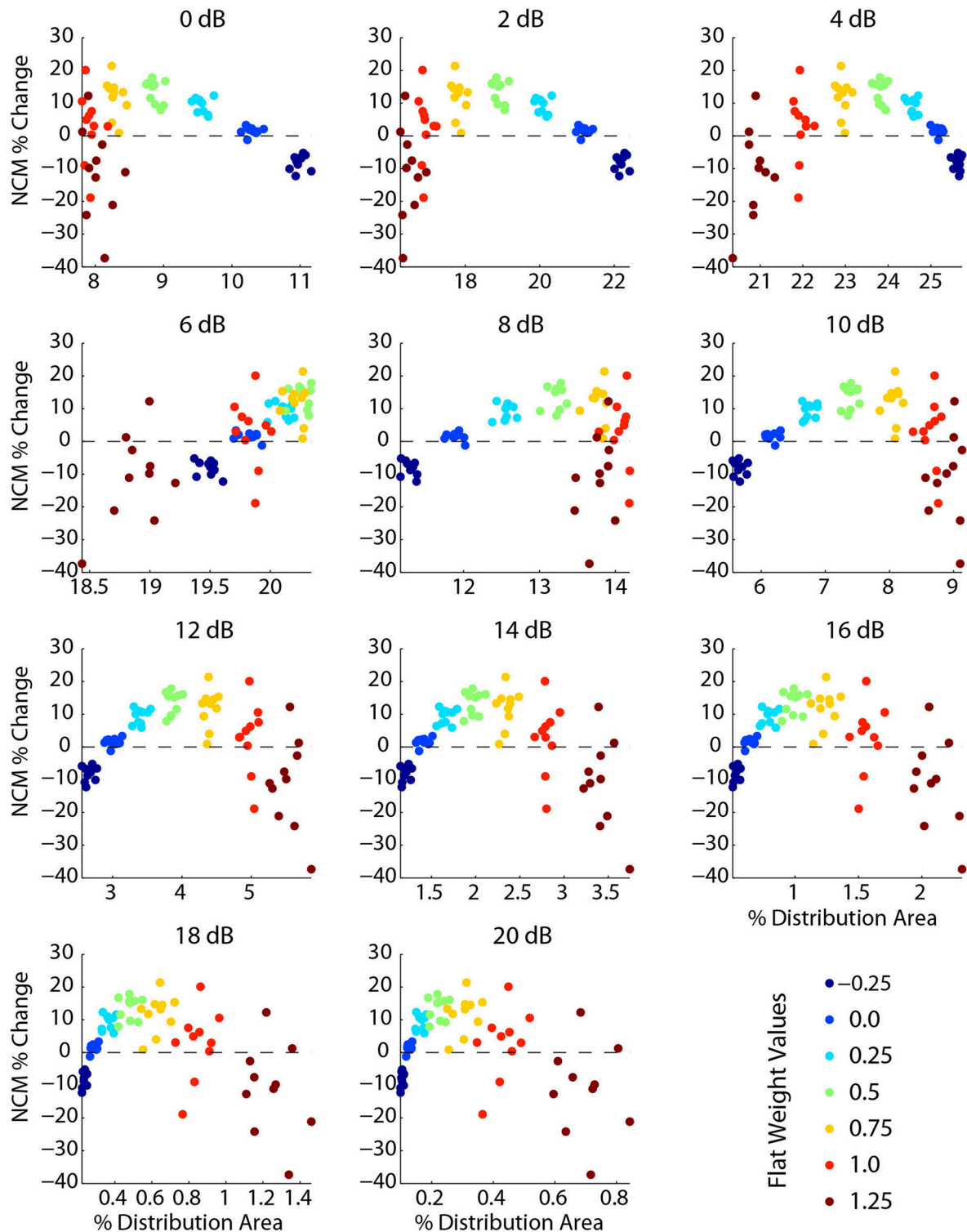
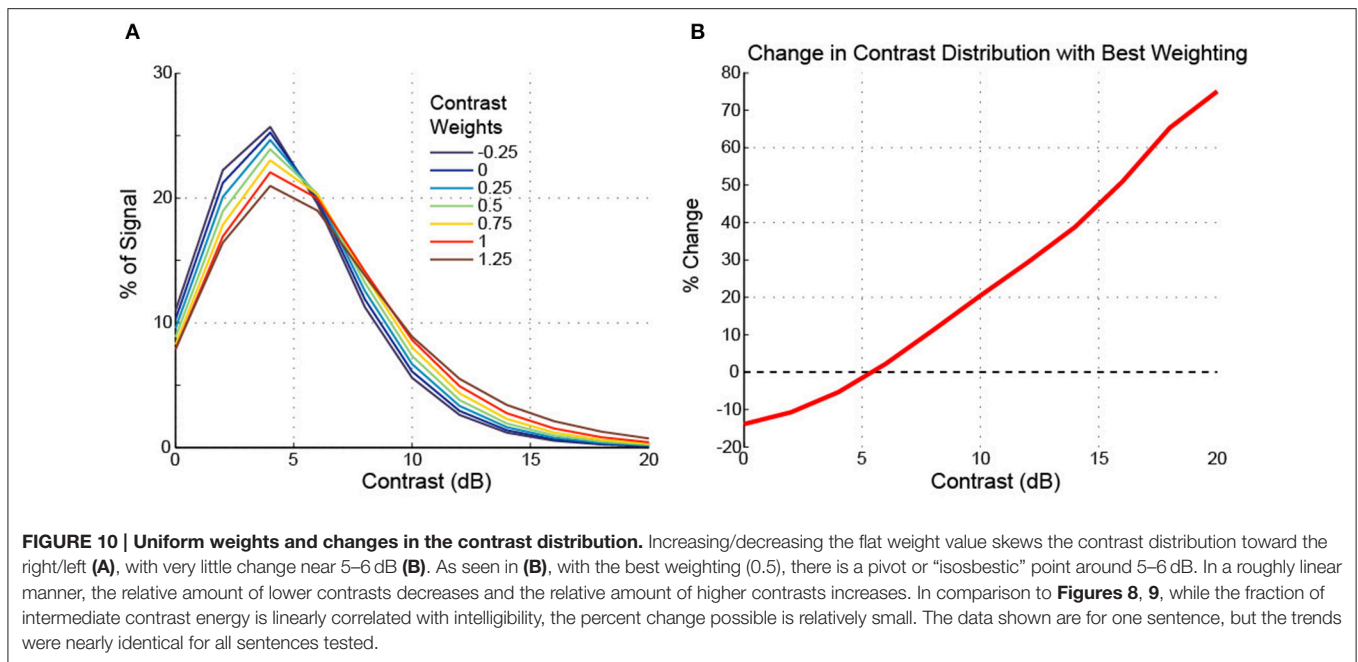


FIGURE 9 | Intelligibility is significantly correlated with the fractional amount of intermediate contrast in the signal. Each plot shows the percent change in NCM score vs. the percent of signal content (percent of histogram area) for one of the 11 contrast bins. Each point represents one of 10 sentences (additive Gaussian white noise, 5 dB SNR) and one flat contrast weighting. The color of each data point represents the weight level (blue \rightarrow red corresponds to -0.25 to 1.25 , green is the optimal weight of 0.5). The 6 dB contrast bin has a strong positive correlation of intelligibility to fractional signal content. Larger weights lead to less fractional signal content at contrasts below 6 dB and more content for contrasts above 6 dB, reflecting that positive contrast weighting skews the spectral contrast distribution toward higher contrasts.



in cochlear implantees are consistent with the computational findings. Cochlear implantees performed relatively better when processing was applied to 5 dB SNR speech in noise compared to 10 dB SNR speech in noise (24% vs. 7% improvement), with one patient consistently having a three-fold improvement at 5 dB compared to a more modest improvement at 10 dB (see **Figure 12B**).

The Role of Contrast-Tuned Neurons

The effects of contrast shaping on a computed intelligibility score, the NCM, indicated that the contrast shapes most effective at improving computed speech intelligibility on average closely matched the distribution of best contrasts in the auditory cortex of marmoset monkeys (**Figure 11**). Furthermore, the peaks of the contrast weighting functions fell near 5 dB for a range of SNR values, indicating that a similar noise-reduction procedure might be effective across a variety of SNRs. We computed the intelligibility of the processed speech at many combinations of accentuation and attenuation of contrast, sorting by intelligibility score and taking the mean of the best contrast-weighting functions. The resulting functions compared favorably with the representations of contrast-tuned neurons in primate auditory cortex. These computational data were derived from an algorithm designed to improve computed speech intelligibility and not by fitting physiological data. This independent result lends plausibility to the hypothesis that contrast-tuned neurons may participate in a biological contrast-shaping algorithm whose purpose is to extract vocalizations or other sounds from a noisy environment. While no causal relationship has yet been established between contrast-tuned neurons and the effect of contrast shaping on altering computed speech intelligibility as reported here, this scenario represents a compelling example where the responses of neurons in the

central nervous system prompted exploration that led to a candidate algorithm for performing a useful engineering task.

Despite the similarities in top weight combinations and tuning curves, it is an open question as to how a firing rate tuning curve might effectively perform local contrast enhancement. From the perspective of what contrast tuning could be doing to the brain's representation of the signal, a simple explanation might be that contrast tuning effectively skews the neuronal representation of the stimulus contrasts. The majority of neurons, which are un-tuned to contrast, could be rate-coding spectral contrast information with a maximizing information capacity strategy (Laughlin, 1981), such that their spiking activity produces a representation of spectral contrast distribution in the sensory environment. The tuned neurons could use this representation to skew the contrast distribution toward higher contrasts, similar to what happens artificially with the best flat weighting scheme. This could happen if the shape of the rate function of the tuned neurons relative to the input contrast distribution shape represented how much of the respective contrast components are passed through to successive levels in the neuronal network. So, if after some normalization (e.g., relative to the maximum firing rate), the firing rate function matched the input distribution, faithful reproduction could occur. Whereas if the rate deviates lower for low contrasts and higher for high contrasts, the effective output distribution may be closer to what occurs in simulation with the best flat weighting.

The design and implementation of the contrast-shaping signal processing array draws inspiration from neurophysiological findings, although the computational components are not intended to model the physiological implementation of the underlying algorithms. Coupling engineering approaches directly with neurophysiological findings represents a powerful tool for devising new, effective algorithms for use in signal

processing. Such a strategy of biologically-inspired engineering derived from cochlear physiology has been quite successful in designing devices to alleviate hearing loss. A broader application of these principles to central nervous system physiology may

yield further improvements in these devices, extending analytical results demonstrating that modeling of auditory cortex neuronal responses can be useful for estimating speech intelligibility in normal subjects (Chi et al., 1999).

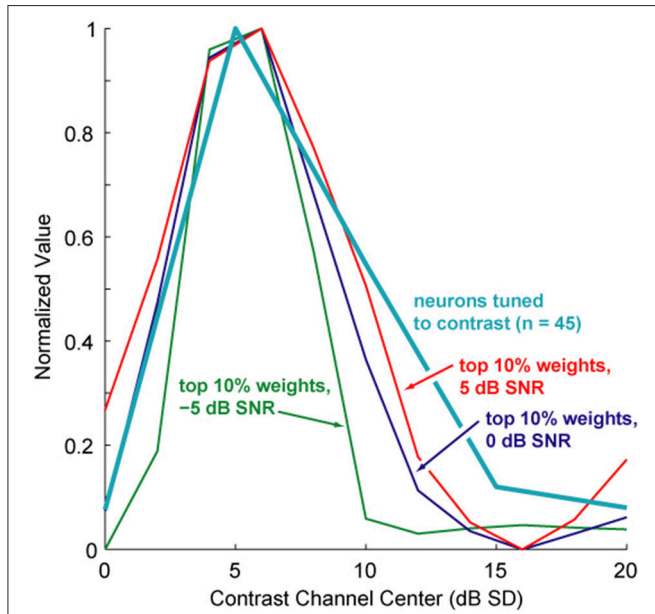


FIGURE 11 | Comparison of best contrast weighting shapes to preferred contrasts of contrast-tuned neurons. Shown is an overlay of averaged contrast shapes achieving the best computed speech intelligibility improvements (from the systematic weight search, see **Figure 5**) compared with the distribution of peak contrast responses from a population of 45 contrast-tuned auditory cortex neurons. The normalized shapes of the best 10% of contrast shapes for 5, 0, and -5 dB SNR of white, Gaussian noise closely resemble this neuronal response distribution. Additionally, these contrast weighting distributions resemble one another even more, each one peaking at around 5 dB SD, despite a total SNR range of 10 dB across all three signals.

Human Performance with Contrast Shaping

Cochlear implantees employing their own implant devices and performing a hearing-in-noise test (HINT) exhibited improved noisy speech intelligibility under the uniform weighting contrast-shaping manipulation. Subjects reported subjectively that these signals sounded much better than the unprocessed noisy signals. Although contrast shaping was not compared directly with other noise-reduction strategies, it appears to be a promising candidate for assisting cochlear implant wearers in noisy situations.

Effective noise-reduction strategies in theoretical tests often show little measurable benefit to human noisy speech intelligibility for non-implantees, even for individuals with considerable hearing loss and even when listeners subjectively indicate a preference for the noise-reduction (Kuk et al., 1990; Moore, 2003). It is possible that real-world tests demonstrate little intelligibility improvement for most individuals because native neuronal noise-reduction circuitry normally does such a good job that trying to improve upon it with an engineered system generally fails to provide additional capabilities (Trine and Tasell, 2002). Even if the above is true, however, noise-reduction strategies could be particularly useful in other situations, such as (1) in subjects such as cochlear implantees, whose native noise-reduction circuitry may not be functioning normally because of the dramatic alteration in signal input relative to normal (Hochberg et al., 1992; Weiss, 1993; Dorman and Loizou, 1996; Loizou and Poroy, 2001; Henry and Turner, 2003; Turner et al., 2004; Henry et al., 2005); (2) in situations where no humans are listening, such as automatic speech recognition (Virag, 1999);

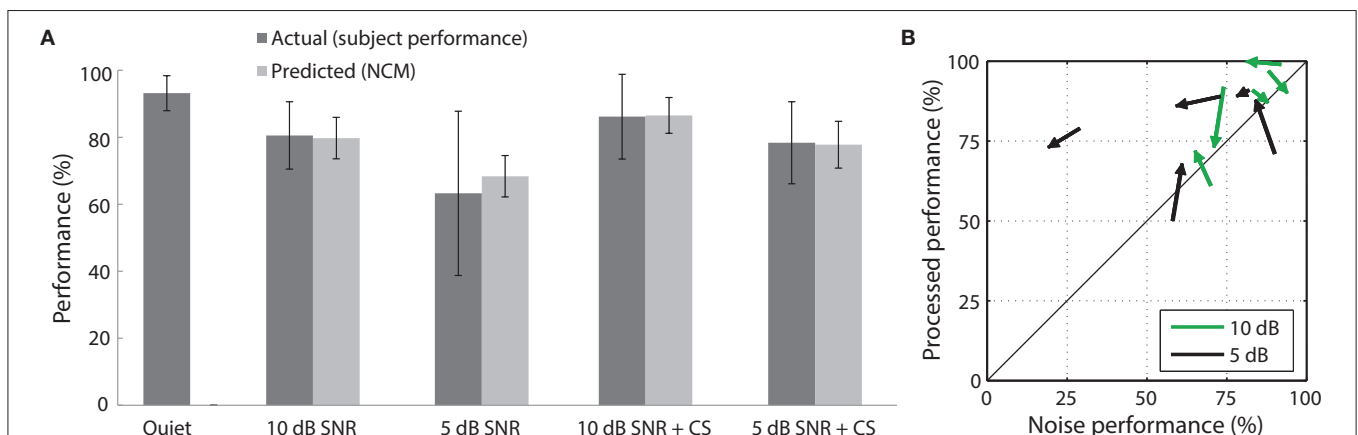


FIGURE 12 | Speech recognition performance of cochlear implant subjects evaluated with the HINT in 5 conditions. (A) Displayed are the raw performance scores (mean \pm standard deviation) of cochlear implantees ($n = 5$ subjects) and the predicted performance (mean \pm standard deviation; $n = 260$ sentences). Improved speech recognition was seen in the 5 and 10 dB SNR flat weighting conditions. (B) Recognition performance (% correct) with contrast shaping is plotted against performance on speech in noise without processing (% correct) for each test session ($n = 5$ subjects, 2 test sessions per subject). The arrows point toward the scores from the second test session.

or (3) to reduce the effort of understanding speech in noise, even if maximal listening effort does not yield improved speech recognition (Kuk et al., 1990; Moore, 2003; Sarampalis et al., 2006).

AUTHOR CONTRIBUTIONS

NK, PW, and DB designed the computational model and performed analyses. All authors designed the tests in cochlear implantees. LD conducted the tests in cochlear implantees. NK and DB wrote the paper.

REFERENCES

- Baer, T., Moore, B. C., and Gatehouse, S. (1993). Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality, and response times. *J. Rehabil. Res. Dev.* 30, 49–72.
- Barbour, D. L., and Wang, X. (2003). Contrast tuning in auditory cortex. *Science* 299, 1073–1075. doi: 10.1126/science.1080425
- Bunnell, H. T. (1990). On enhancement of spectral contrast in speech for hearing-impaired listeners. *J. Acoust. Soc. Am.* 88, 2546–2556. doi: 10.1121/1.399976
- Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (1999). Spectro-temporal modulation transfer functions and speech intelligibility. *J. Acoust. Soc. Am.* 106, 2719–2732. doi: 10.1121/1.428100
- Cover, T. M., and Thomas, J. A. (1991). *Elements of Information Theory*. New York, NY: Wiley.
- Dorman, M. F., and Loizou, P. C. (1996). Improving consonant intelligibility for Ineraid patients fit with continuous interleaved sampling (CIS) processors by enhancing contrast among channel outputs. *Ear Hear.* 17, 308–313. doi: 10.1097/00003446-199608000-00002
- Dreisbach, L. E., Leek, M. R., and Lentz, J. J. (2005). Perception of spectral contrast by hearing-impaired listeners. *J. Speech Lang. Hear. Res.* 48, 910–921. doi: 10.1044/1092-4388(2005/063)
- Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J. Acoust. Soc. Am.* 110, 1150–1163. doi: 10.1121/1.1381538
- Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., Dahlgren, N., et al. (1993). *TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1*. Philadelphia, PA: Linguistic Data Consortium. Available online at: <https://catalog.ldc.upenn.edu/LDC93S1>
- Goldsworthy, R. L. (2005). *Noise Reduction Algorithms and Performance Metrics for Improving Speech Reception in Noise by Cochlear-Implant Users*. Ph.D., Massachusetts Institute of Technology.
- Goldsworthy, R. L., and Greenberg, J. E. (2004). Analysis of speech-based Speech Transmission Index methods with implications for nonlinear operations. *J. Acoust. Soc. Am.* 116, 3679–3689. doi: 10.1121/1.1804628
- Henry, B. A., and Turner, C. W. (2003). The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *J. Acoust. Soc. Am.* 113, 2861–2873. doi: 10.1121/1.1561900
- Henry, B. A., Turner, C. W., and Behrens, A. (2005). Spectral peak resolution and speech recognition in quiet: normal hearing, hearing impaired, and cochlear implant listeners. *J. Acoust. Soc. Am.* 118, 1111–1121. doi: 10.1121/1.1944567
- Hochberg, I., Boothroyd, A., Weiss, M., and Hellman, S. (1992). Effects of noise and noise suppression on speech perception by cochlear implant users. *Ear Hear.* 13, 263–271. doi: 10.1097/00003446-199208000-00008
- Jerger, J. F., Carhart, R., Tillman, T. W., and Peterson, J. L. (1959). Some relations between normal hearing for pure tones and for speech. *J. Speech Hear. Res.* 2, 126–140. doi: 10.1044/jshr.0202.126
- Kuk, F. K., Tyler, R. S., and Mims, L. (1990). Subjective ratings of noise-reduction hearing aids. *Scand. Audiol.* 19, 237–244. doi: 10.3109/01050399009070778
- Laughlin, S. (1981). A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch. C* 36, 910–912.

ACKNOWLEDGMENTS

Supported by The Wallace H. Coulter Foundation and National Institutes of Health grant R01-DC009215 and The Center for Integration of Medicine and Innovative Technology.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01145>

- Leek, M. R., Dorman, M. F., and Summerfield, Q. (1987). Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 81, 148–154. doi: 10.1121/1.395024
- Leek, M. R., and Summers, V. (1996). Reduced frequency selectivity and the preservation of spectral contrast in noise. *J. Acoust. Soc. Am.* 100, 1796–1806. doi: 10.1121/1.415999
- Loizou, P. C., and Poroy, O. (2001). Minimum spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners. *J. Acoust. Soc. Am.* 110(3 Pt 1), 1619–1627. doi: 10.1121/1.1388004
- Lyzenga, J., Festen, J. M., and Houtgast, T. (2002). A speech enhancement scheme incorporating spectral expansion evaluated with simulated loss of frequency selectivity. *J. Acoust. Soc. Am.* 112(3 Pt 1), 1145–1157. doi: 10.1121/1.497619
- Moore, B. C. J. (2003). Speech processing for the hearing-impaired: successes, failures, and implications for speech mechanisms. *Speech Commun.* 41, 81–91. doi: 10.1016/S0167-6393(02)00095-X
- Munoz, C., Nelson, P., and Ruledge, J. (1999). “Enhancement of spectral contrast in speech for hearing impaired listeners,” in *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing* (Antalya).
- Nie, K., Stickney, G., and Zeng, F. G. (2005). Encoding frequency modulation to improve cochlear implant performance in noise. *IEEE Trans. Biomed. Eng.* 52, 64–73. doi: 10.1109/TBME.2004.839799
- Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am.* 95, 1085–1099. doi: 10.1121/1.408469
- Peterson, G. E., and Lehiste, I. (1962). Revised CNC lists for auditory tests. *J. Speech Hear. Disord.* 27, 62–70. doi: 10.1044/jshd.2701.62
- Ribic, Z., Yang, J., and Latzel, M. (1996). “Adaptive spectral contrast enhancement based on masking effect for the hearing impaired,” in *International Conference on Acoustics, Speech and Signal Processing* (Atlanta, GA), 937–940.
- Sarampalis, A., Kalluri, S., Edwards, B. W., and Hafter, E. R. (2006). Cognitive effects of noise reduction strategies. *J. Acoust. Soc. Am.* 119, 3239. doi: 10.1121/1.4786003
- Simpson, A. M., Moore, B. C., and Glasberg, B. R. (1990). Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners. *Acta Otolaryngol. Suppl.* 469, 101–107.
- Stone, M. A., and Moore, B. C. (1992). Spectral feature enhancement for people with sensorineural hearing impairment: effects on speech intelligibility and quality. *J. Rehabil. Res. Dev.* 29, 39–56. doi: 10.1682/JRRD.1992.04.0039
- Strehl, A., and Ghosh, J. (2003). Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.* 3, 583–617. doi: 10.1162/15324430321897735. Available online at: <http://dl.acm.org/citation.cfm?id=944935>
- Summers, V., and Leek, M. R. (1994). The internal representation of spectral contrast in hearing-impaired listeners. *J. Acoust. Soc. Am.* 95, 3518–3528. doi: 10.1121/1.409969
- ter Keurs, M., Festen, J. M., and Plomp, R. (1993). Limited resolution of spectral contrast and hearing loss for speech in noise. *J. Acoust. Soc. Am.* 94(3 Pt 1), 1307–1314. doi: 10.1121/1.408158

- Theunissen, F. E., and Doupe, A. J. (1998). Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVC of male zebra finches. *J. Neurosci.* 18, 3786–3802.
- Trine, T. D., and Tasell, D. V. (2002). Digital hearing aid design: fact vs. fantasy. *Hear. J.* 55, 36–42. doi: 10.1097/01.HJ.0000292490.67222.a2
- Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (2004). Speech recognition in noise for cochlear implant listeners: benefits of residual acoustic hearing. *J. Acoust. Soc. Am.* 115, 1729–1735. doi: 10.1121/1.1687425
- Virag, N. (1999). Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Trans. Speech Audio Process.* 7, 126–137. doi: 10.1109/89.748118
- Weiss, M. R. (1993). Effects of noise and noise reduction processing on the operation of the Nucleus-22 cochlear implant processor. *J. Rehabil. Res. Dev.* 30, 117–128.
- Yang, J., Luo, F.-L., and Nehorai, A. (2003). Spectral contrast enhancement: algorithms and comparisons. *Speech Commun.* 39, 33–46. doi: 10.1016/S0167-6393(02)00057-2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Killian, Watkins, Davidson and Barbour. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.