

# Effects of WDRC Release Time and Number of Channels on Output SNR and Speech Recognition

Joshua M. Alexander and Katie Masterson

**Objectives:** The purpose of this study was to investigate the joint effects that wide dynamic range compression (WDRC) release time (RT) and number of channels have on recognition of sentences in the presence of steady and modulated maskers at different signal-to-noise ratios (SNRs). How the different combinations of WDRC parameters affect output SNR and the role this plays in the observed findings were also investigated.

**Design:** Twenty-four listeners with mild to moderate sensorineural hearing loss identified sentences mixed with steady or modulated maskers at three SNRs (-5, 0, and +5 dB) that had been processed using a hearing aid simulator with six combinations of RT (40 and 640 msec) and number of channels (4, 8, and 16). Compression parameters were set using the Desired Sensation Level v5.0a prescriptive fitting method. For each condition, amplified speech and masker levels and the resultant long-term output SNR were measured.

**Results:** Speech recognition with WDRC depended on the combination of RT and number of channels, with the greatest effects observed at 0 dB input SNR, in which mean speech recognition scores varied by 10 to 12% across WDRC manipulations. Overall, effect sizes were generally small. Across both masker types and the three SNRs tested, the best speech recognition was obtained with eight channels, regardless of RT. Increased speech levels, which favor audibility, were associated with the short RT and with an increase in the number of channels. These same conditions also increased masker levels by an even greater amount, for a net decrease in the long-term output SNR. Changes in long-term SNR across WDRC conditions were found to be strongly associated with changes in the temporal envelope shape as quantified by the Envelope Difference Index; however, neither of these factors fully explained the observed differences in speech recognition.

**Conclusions:** A primary finding of this study was that the number of channels had a modest effect when analyzed at each level of RT, with results suggesting that selecting eight channels for a given RT might be the safest choice. Effects were smaller for RT, with results suggesting that short RT was slightly better when only 4 channels were used and that long RT was better when 16 channels were used. Individual differences in how listeners were influenced by audibility, output SNR, temporal distortion, and spectral distortion may have contributed to the size of the effects found in this study. Because only general suppositions could be made for how each of these factors may have influenced the overall results of this study, future research would benefit from exploring the predictive value of these and other factors in selecting the processing parameters that maximize speech recognition for individuals.

**Key words:** Envelope difference index, Output signal-to-noise, Signal-to-noise ratio, Temporal envelope, Wide dynamic range compression.

(*Ear & Hearing* 2015;36:e35–e49)

## INTRODUCTION

For the past 15 to 20 years, one of the greatest tools for amplifying speech in hearing aids has been multichannel wide

dynamic range compression (WDRC). One reason for its ubiquity is that it can increase audibility for weak sounds while maintaining comfort for intense sounds, thereby increasing the dynamic range of sound available to the user. Compression parameters, such as compression release time (RT) and number of channels, are adjustments that are not typically available to audiologists, and, if they are, they are rarely manipulated, despite their potential to alter acoustic cues (Souza 2002). In a survey of almost 300 audiologists, Jenstad et al. (2003) reported that part of the reason these settings are not adjusted is lack of knowledge about how they will influence outcomes. This is important because WDRC has the potential to alter speech recognition by influencing a hearing aid user's access to important speech cues in an environment with competing sounds.

## Compression RT

Gain in a compression circuit varies as a function of input level and is determined by the compression ratio and threshold that define the input–output level function. Typical speech levels fluctuate with varying time courses across frequency. However, adjustments in gain are relatively sluggish because they are controlled by compression time-constants that smooth the gain control signal in order to minimize processing artifacts, such as harmonic and intermodulation distortion (Moore et al. 1999). The attack time controls the rate of decrease in gain associated with increases in input level. A short attack time is generally recommended to help prevent intense sounds with sudden onset from causing loudness discomfort. RT controls the rate of increase in gain associated with decreases in input level. A longer RT means that it will take the compression circuit more time to increase gain and reach the target output level. Therefore, if the input signal level fluctuates at a faster rate than the RT, the target output level will not be fully realized and audibility may be compromised. That is, the output will “undershoot” its target level, resulting in an effective compression ratio that is less than the nominal value for the compression circuit (Verschuure et al. 1996). There are no accepted definitions of fast or slow compression, but RTs less than 200 to 250 msec are commonly accepted as giving fast compression because these are shorter than the period of the dominant modulation rate for speech (4–5 Hz).

Unlike attack time, there is little agreement about what constitutes an appropriate RT (Moore et al. 2010). Acoustically, short RTs put a greater range of the speech input into the user's residual dynamic range because the compression circuit can act more quickly to amplify the lower-intensity segments of speech above their threshold for audibility (Henning & Bentler 2008). Perceptually, short RTs may be favorable for speech recognition in quiet because they can increase audibility for soft consonants (e.g., voiceless stops and fricatives), especially when a more intense preceding speech sound (e.g., a vowel) drives the circuit into compression (Souza 2002).

Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana, USA.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site ([www.ear-hearing.com](http://www.ear-hearing.com)).

Short RTs might also be favorable for listening to speech in a competing background. Listening in the dips refers to the release from masking that comes from listening to speech in a background that is fluctuating in amplitude compared with a background that is steady (Duquesnoy 1983; Festen & Plomp 1990; George et al. 2006; Lorenzi et al. 2006). During brief periods when the fluctuating interference has low intensity (i.e., “dips”), listeners are afforded an opportunity to glimpse information that allows them to better segregate speech from background interference and/or piece together missing context (Moore et al. 1999; Cooke 2006; Hopkins et al. 2008; Moore 2008). Audibility of information in dips can depend greatly on the RT in the WDRC circuit. When input levels drop because of momentary breaks in background interference, short RTs allow gain to recover in time to restore the audibility of the interposing speech from the target talker.

Only a few studies have measured the effects of short RT on both audibility and speech recognition. Souza and Turner (1998 1999) found that two-channel WDRC with short RT improved recognition relative to linear amplification for listeners with mild to severe sensorineural hearing loss (SNHL), but only under conditions where WDRC resulted in an improvement in audibility, namely when presentation levels were low. Using a clinical device with four channels and short RT that could be programmed with WDRC or compression limiting, Davies-Venn et al. (2009) found somewhat different results for two groups of listeners with mild to moderate or severe SNHL. Similar to the study by Souza and Turner, they found that WDRC with short RT yielded the greatest speech recognition advantage over compression limiting when audibility was most affected, which was when presentation levels were low, particularly for listeners with severe SNHL. For average and high presentation levels, WDRC with short RT decreased speech recognition relative to compression limiting despite measurable improvements in audibility. The authors suggest that the observed disassociation between audibility and recognition might occur because “it is possible that once optimum audibility is achieved, susceptibility to WDRC-induced distortion does occur” (p. 500).

It has long been suspected that use of short RTs to amplify speech has limitations. This is because relative to long RTs, short RTs are more effective at equalizing differences in amplitude across time, which can distort information carried by slowly varying amplitude contrasts (the temporal envelope) in the input signal (Plomp 1988; Drullman et al. 1994; Souza & Turner 1998; Stone & Moore 2003 2007; Jenstad & Souza 2005 2007; Souza & Gallun 2010). This effect becomes more pronounced as nominal compression ratios are increased to accommodate reductions in residual dynamic range associated with greater severities of hearing loss (Verschuure et al. 1996; Souza 2002; Jenstad & Souza 2007). Stone and Moore (2003, 2004, 2007, 2008) have shown that to the extent that listeners rely on the temporal envelope of speech, recognition can be compromised by short RTs. These same authors also demonstrated that a short RT can induce cross-modulation (common amplitude fluctuation) between the target speech and background interference, which can make it increasingly difficult to perceptually segregate the two.

### Number of Channels

With multichannel WDRC, the acoustic signal is split into several frequency channels. The short-term level is estimated for each channel and gain is selected accordingly. The effect of RT

on speech recognition may interact with the number of independent compression channels. One hypothesis is that increasing the number of channels improves recognition because gain can be increased more selectively across frequency, thereby increasing the audibility of low-level speech cues (Moore et al. 1999). The use of several channels can help promote audibility by ensuring that low-intensity speech segments in spectral regions with favorable signal-to-noise ratios (SNRs) receive sufficient gain.

Another hypothesis is that increasing the number of channels impairs speech recognition because greater gain is applied to channels that contain valleys in the spectrum than to channels that contain peaks, which tends to flatten the spectral envelope, especially in the presence of background noise (Plomp 1988; Edwards 2004). Listeners with SNHL tend to have more difficulty processing changes in spectral shape and maintaining an internal representation of spectral peak-to-valley contrasts (e.g., Bacon & Brandt 1982; Leek et al. 1987; Turner et al. 1987; Van Tasell et al. 1987; Summers & Leek 1994; Henry & Turner 2003). An additional reduction in spectral contrast from multichannel WDRC can diminish the perception of speech sounds whose phonemic cues are primarily conveyed by spectral shape, for example, vowels, semivowels, nasals, and place of articulation for consonants (Yund & Buckles 1995b; Franck et al. 1999; Bor et al. 2008; Souza et al. 2012). Thus, performance for WDRC with short RTs, which also degrades temporal envelope cues, might decrease once the number of channels is increased beyond the point at which additional channels no longer optimize audibility (Stone & Moore 2008). This reasoning may help explain results from the study by Yund and Buckles (1995a), which found improvements in speech recognition as the number of channels was increased from 4 to 8, but no change as the number of channels was increased from 8 to 16.

### Effects of RT and Number of Channels on Output SNR

In addition to the arguments for and against multichannel WDRC and/or a short RT in terms of audibility and distortion, a few investigators have described changes in SNR after processing with WDRC. When speech and maskers are processed together by multichannel WDRC, their relative levels after amplification (the output SNR) may depend on the RT, number of channels, masker type, and input SNR (Souza et al. 2006; Naylor & Johannesson 2009). Souza et al. (2006) used the phase inversion technique proposed by Hagerman and Olofsson (2004) to compare the output SNR for linear amplification to that for a software simulation of single- and dual-channel WDRC with short RT. They found that when speech with a steady masker was processed with WDRC, the SNR at the output was less than that at the input, especially at higher input SNRs. Naylor and Johannesson (2009) also used the phase inversion technique to investigate how the output SNR was affected by the masker type (steady, modulated, and actual speech) and by the compression parameters in a hearing aid. In addition to replicating the findings by Souza et al., they found minimal differences for modulated versus steady maskers. They also found that the decrease in output SNR was greater (more adverse) as more compression was applied (i.e., higher compression ratios and shorter RT). Interestingly, they did not find an effect of increasing the number of channels from 1 to 8.

The perceptual relevance of changes in output SNR that may occur as RT and/or the number of channels are varied is uncertain. The reason for this is that although sentence recognition

tends to be a monotonic function of SNR, this relationship has generally been shown for conditions in which the only variable is the speech or noise level, with all other characteristics of the speech and masker held constant. However, as noted, multiple changes occur to temporal and spectral cues in the speech signal as WDRC parameters are manipulated, thereby making predictions for speech recognition across conditions based on SNR-alone tenuous (Naylor & Johannesson 2009). To date, very few studies have examined the role of output SNR on speech recognition outcomes with different multichannel WDRC amplification conditions. The closest reports are those by Naylor et al. (2007, 2008) who measured sentence recognition using listeners wearing a hearing aid programmed with linear amplification or programmed with single-channel WDRC and a short RT. They found that although an increase or decrease in the output SNR after WDRC generally predicted whether speech recognition would increase or decrease relative to linear amplification, that recognition for the two amplification conditions were not comparable when evaluated in terms of output SNR. That is, changing the input/output SNR of the linear hearing aid program to match the change in output SNR observed when WDRC was implemented did not change speech recognition scores by the same amount. To examine this issue further, the present study will use performance–intensity (PI) functions to extrapolate speech recognition scores at a fixed output SNR for different multichannel WDRC conditions. Expressed this way, speech recognition as a function of RT and number of channels should be equal if the effects of multichannel WDRC can be explained solely by output SNR.

### Study Rationale

Most of the previously reported investigations of the effects of RT and number of channels on speech recognition and on indices of audibility have examined these factors in isolation and have not considered them jointly. When considered separately, short RT and multiple channels appear to be beneficial for audibility. However, these two factors may also be associated with reduced temporal and spectral contrast, respectively, which could have negative consequences for speech recognition. To the extent that listeners are influenced by audibility and reduced contrast, an interaction is predicted. The best combination of parameters for speech recognition may occur for short RT (good for audibility) combined with few channels (good for spectral contrast) or occur for long RT (good for temporal contrast) combined with many channels (good for audibility). Furthermore, these effects may depend on the SNR. Long RT combined with few channels might lead to decreases in recognition primarily at negative SNRs, when audibility is most affected. Short RT combined with many channels might lead to decreases in recognition primarily at positive SNRs, when audibility is less of a concern (Hornsby & Ricketts 2001; Olsen et al. 2004). Finally, the effects might depend on the modulation characteristics of the masker. For example, combinations of RT and number of channels that promote audibility may be more favorable for a modulated masker than for a steady masker because audibility of speech information will vary rapidly over time and frequency, especially at negative SNRs.

To explore these effects and interactions, the present study used listeners with mild to moderate SNHL to measure speech recognition for sentences mixed with a steady or modulated masker at three SNRs (–5, 0, and +5 dB) that had been

processed using six combinations of RT (40 and 640 msec) and number of channels (4, 8, and 16). Unlike many of the previously reported investigations, which used unrealistically high compression ratios (presumably, for greater effect) or did not customize the compression parameters for individual losses, the compression parameters in this study were set as close as possible to what would be done clinically, yet with very precise control over the manipulations so that relationships between their acoustic characteristics (i.e., relative speech and masker levels) and the observed perceptual outcomes could be made.

## PARTICIPANTS AND METHODS

### Participants

Twenty-four native–English-speaking adults (16 females and 8 males) with mild to moderately-severe SNHL (thresholds between 25 and 70 dB HL) participated as listeners in this study. Listeners had a median age of 62 years (range, 31–83 years). Tympanometry and bone conduction thresholds were consistent with normal middle ear function. The test ear was selected using threshold criteria for moderate high-frequency SNHL (i.e., between 2 and 6 kHz, average thresholds were in the range 40–55 dB HL for all but four listeners, all listeners had at least one threshold  $\geq 40$  dB HL, and none had a threshold  $> 70$  dB HL). If the hearing loss for the two ears met the criteria, the right ear was used as the test ear. Table 1 provides demographic and audiometric information for each listener.

### Stimuli

The target speech consisted of the 720 sentences from the Institute of Electrical and Electronics Engineers (IEEE) sentence database (Rothausser et al. 1969). The sentences have relatively low semantic predictability and five keywords per sentence. They were spoken by a female talker and recorded in a mono channel at a 44.1 kHz sampling rate using a LynxTWO™ sound card (Lynx Studio Technology, Inc., Newport Beach, CA) and a headworn Shure Beta 53 professional subminiature condenser omnidirectional microphone with foam windscreens and filtered protective cap (Niles, IL). The published frequency response of this microphone arrangement is flat within 3 dB up to 10 kHz. Master recordings were downsampled at 22.05 kHz for further processing.

Sentences were mixed with two masker types, steady and modulated, at three SNRs (–5, 0, and +5 dB). The modulated masker was processed using methods specified by the International Collegium of Rehabilitative Audiology (ICRA) 2010

\* The signal was first filtered into three bands with crossover frequencies at 850 and 2500 Hz, using Butterworth filters with 100 dB/octave slopes. According to the International Collegium of Rehabilitative Audiology documentation, these bands were chosen so that most first and second formant frequencies of the vowels would be processed by the first and second bands, respectively, whereas unvoiced fricatives would be processed by the third band. The waveform samples from each band were multiplied by a signal of equal length that consisted entirely of a random sequence of positive and negative ones. Because this preserves the envelope of the original signal in each band, but with a flat spectrum, the modified signals were filtered again using the same filter as before. These band-limited, modulated signals were then scaled to a constant spectral density level so that, when summed, they created signals with a flat spectrum, which were then analyzed every 32 samples by overlapping 256-point fast Fourier transforms (FFTs). The phase derived from each FFT was randomized and then converted back into the time domain using inverse FFT. The phase-altered segments were then reassembled using 7/8 overlap-and-add.

**TABLE 1.** Listeners' age, whether they were current hearing aid users (Y/N), and thresholds in dB HL at each frequency (kHz)

Listener	Age (yrs)	Aid	0.25	0.5	1.0	2.0	3.0	4.0	6.0	8.0
1	68	N	10	15	10	5	30	35	50	55
2	55	N	5	15	10	10	55	55	65	60
3	60	N	5	10	15	20	45	50	60	55
4	63	N	20	20	15	25	40	45	55	45
5	61	Y	10	5	10	40	40	40	70	55
6	64	N	0	15	20	25	30	50	55	35
7	68	Y	5	5	5	35	60	55	70	50
8	47	N	10	15	20	20	60	60	60	55
9	71	N	15	20	35	30	30	40	65	65
10	63	N	30	45	30	30	45	45	45	40
11	83	Y	5	20	15	40	30	50	70	80
12	64	N	10	25	45	40	30	25	55	40
13	59	N	25	25	20	35	50	55	65	60
14	70	N	30	40	25	35		50	60	70
15	53	Y	5	15	20	35	55	60	55	60
16	61	Y	15	15	20	50	60	50	50	60
17	50	Y	15	25	35	45	45	45	65	60
18	54	Y	45	45	40	40	45	50	55	55
19	53	N	15	25	30	35	60	65	65	65
20	61	Y	25	25	15	55	60	65	60	55
21	70	Y	20	25	40	40	45	55	70	65
22	82	Y	25	30	40	45	50	55	60	35
23	31	Y	55	50	50	45	40	50	65	50
24	66	Y	20	35	60	55	50	55	65	60
Mean	61.5		17.5	23.5	26.0	34.8	45.9	50.2	60.6	55.4
SD	10.7		12.9	11.9	14.1	12.4	10.6	9.0	6.8	10.6

Listeners are arranged according to the average thresholds at 1.0, 2.0, and 4.0 kHz.

(Dreschler et al. 2001).<sup>\*</sup> The resultant "ICRA noise" had a similar broadband temporal envelope and spectrum as the original speech but was unintelligible. ICRA noise was generated independently from one male and one female talker (different from the talker used for the IEEE sentences) reading 5 min of prose that was recorded in a similar manner as the target sentences. Silent intervals >250 msec were shortened. Recordings derived from each talker were then band-pass filtered<sup>†</sup> and spectrally shaped to approximate the 1/3-octave band levels from the International Long-Term Average Speech Spectrum (Byrne et al. 1994). Using stored random seeds, random selections derived from each masker talker with durations 500 msec longer than the target sentence were mixed together to form a modulated two-talker masker. The masker started 250 msec before and ended 250 msec after the target sentence. The steady masker for each sentence was produced

<sup>†</sup> The filter design and envelope peak detector in MATLAB were based on code from James Kates. The first channel used a low-pass filter that extended down to 0 Hz, and the last channel used a high-pass filter that extended up to the Nyquist frequency. Finite impulse response filters were used to maintain a constant delay (linear phase) at the output of each filter. The impulse response of each filter lasted 8 msec (176 taps for a 22,050 Hz sampling frequency). Finite Impulse Response (FIR) filters were generated in MATLAB using the FIR command such that half the transition region was added/subtracted from each crossover frequency. The transition region of each filter was 350 Hz. For very low crossover frequencies, the transition region was divided by 4 so that negative frequencies would not be submitted to the filter design function. The filters used to scale masker levels to the International Long-Term Average Speech Spectrum were designed in the same way, except that crossover frequencies were the arithmetic mean of the 1/3-octave frequencies and the transition regions were equal to  $\pm 10\%$  of each crossover frequency.

using the same random selections of ICRA noise, the only difference being that the temporal envelope was flattened by randomizing the phase of a fast Fourier transform of the masker followed by inverse fast Fourier transform. Figure 1 shows the long-term average spectra for a 65 dB SPL presentation level of the target speech (solid line), masker (dotted line, which was approximately the same for both types), and the international average (filled circles) from the study by Byrne et al. (1994). As shown, the long-term average spectrum of the target speech closely approximated the 1/3-octave band levels of the masker (and international average), except for the 100 and 125 Hz bands, where it was significantly lower. This difference is not expected to have influenced listener performance because of the relatively small contribution to overall speech recognition by these bands and it is not expected to have influenced the overall output SNR because of the relatively small amount of amplification applied to this region.

### Procedure

Speech recognition for the two masker types and three SNRs was measured with six WDRC parameter combinations (40 and 640 ms RT with 4, 8, and 16 channels) for a total of 36 test conditions. Sentences were presented in the same order for all listeners, but the condition assigned to each sentence was randomized using a Latin Square design. The experiment was divided into 10 blocks of 72 sentences, each of the 36 conditions being tested twice per block. Listeners completed the 10 blocks in two separate test sessions, each lasting 90 to 120 min. Before data collection, all listeners were given a practice session in which they identified 3 blocks of 24 sentences

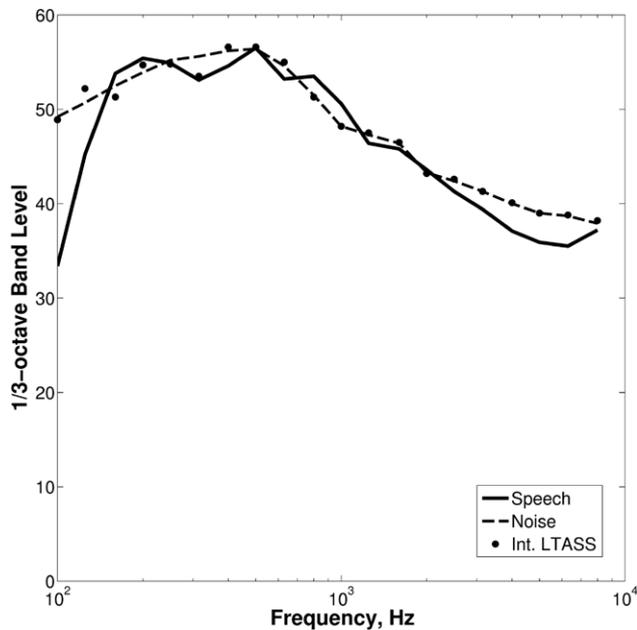


Fig. 1. Long-term average spectra of the target speech (solid line) and masker (dashed line) at 0 dB signal-to-noise ratio, using 1/3-octave band analysis. For comparison, the 1/3-octave band levels (after scaling down 5.5 dB to align with the level of the masker at 1000 Hz) from the International Long-Term Average Speech Spectrum (LTASS; Byrne et al. 1994) are plotted using filled circles.

from the Bamford-Kowal-Bench sentence lists (Bench et al. 1979). In addition, they completed two more blocks of practice from the Bamford-Kowal-Bench sentences lists before the start of the second session. The 24 sentences in each block were processed with 4 or 16 channels and represented a subset of the conditions used in the main experiment. For all sentence testing, listeners were instructed to repeat back any part of the sentence that they heard, even if they were uncertain. A tester, who was blind to the conditions, sat in a sound-treated booth away from the listeners and scored the keywords during the experiment using custom software in MATLAB (MathWorks, Natick, MA).

### Hearing Aid Simulator

To have rigorous control as well as flexibility over the signal processing, a hearing aid simulator designed using MATLAB by the first author (see also, Brennan et al. 2012; Kopun et al. 2012; McCreery et al. 2013; McCreery et al. 2014) was used to amplify speech via a LynxTWO™ sound card and circumaural BeyerDynamic DT150 headphones (Heilbron, Germany). Absolute threshold in dB SPL at the tympanic membrane for each listener was estimated using a transfer function of the TDH 50 earphones (used for the audiometric testing) on a DB100 Zwislocki Coupler and Knowles Electronic Manikin for Acoustic Research. Threshold values and the number of compression channels (4, 8, and 16) were entered in the DSL m(I/O) v5.0a algorithm for adults. Individualized prescriptive values for compression threshold and compression ratio for each channel as well as target values for the real-ear aided responses were generated for a behind-the-ear hearing aid style with WDRC, no venting, and a monaural fitting.

Individualized prescriptions were generated separately for 4, 8, and 16 channels using a 5 msec attack time and a 160 msec RT,

the geometric mean of the 40 and 640 msec RTs used to process the experimental stimuli. Attack time was kept constant at 5 msec throughout the experiment. Gain for each channel was automatically tuned to targets using the “carrot passage” from Audioscan® (Dorchester, ON). Specifically, the 12.8 sec speech passage was analyzed using 1/3-octave filters specified by American National Standards Institute S1.11 (American National Standards Institute 2004). A transfer function of the BeyerDynamic DT150 headphones on a DB100 Zwislocki Coupler and Knowles Electronic Manikin for Acoustic Research was used to estimate the real-ear aided response. The resulting long-term average speech spectrum was then compared with the prescribed DSL m(I/O) v5.0a targets for a 60 dB SPL presentation level. The 1/3-octave values for dB SPL were grouped according to the WDRC channel they overlapped. The average level difference for each group was computed and used to adjust the gain in the corresponding channel. Maximum gain was limited to 55 dB and minimum gain was limited to 0 dB (after accounting for the headphone transfer function). Because of filter overlap and subsequent summation in the low-frequency channels, gain for channels centered below 500 Hz was decreased by an amount equal to  $10 \times \log_{10}(N \text{ channels})$ , where  $N \text{ channels}$  is the total number of channels used in the simulation. Because DSL v5.0a does not prescribe an output target at 8000 Hz, an automatic routine in MATLAB set the channel gains so that the real-ear aided response generated a sensation level (SL) in this frequency region that was equal to SL of the prescribed target at 6000 Hz. To prevent the automatic routine from creating an unusually sharp frequency response, the resultant level was limited to the level of the 6000 Hz target plus 10 dB.

Within MATLAB, stimuli were first scaled so that the input speech level was 65 dB SPL and then band-pass filtered into 4, 8, or 16 channels. Center and crossover frequencies were based on the recommendations of the DSL algorithm. Specifically, when the number of channels is a power of 2, crossover frequencies roughly correspond to a subset of the nominal 1/3-octave frequencies: 250, **315**, 400, **500**, 630, **800**, 1000, **1250**, 1600, **2000**, 2500, **3150**, 4000, **5000**, 6300 Hz (eight-channel crossover frequencies are in bold and four-channel crossover frequencies are italicized).†

Each channel signal was smoothed using an envelope peak detector (Equation 8.1 in Kates 2008).‡ This set the attack and release times so that the values corresponded to what would be measured acoustically using American National Standards Institute S3.22 (American National Standards Institute 2003). The gain below compression threshold was constant, except where the DSL v5.0a algorithm prescribed a compression ratio  $<1.0$  (expansion) (the DSL algorithm for adults often prescribes negative gain with expansion in low-frequency regions of normal hearing).

Output limiting compression with a 10:1 compression ratio, 1 msec attack time, and 50 msec RT was implemented in each channel to keep the estimated real-ear aided response from exceeding the DSL recommended Broadband Output Limiting Targets or 105 dB SPL, whichever was lower. The signals were summed across channels and subjected to a final stage of output limiting compression (10:1 compression ratio, 1 msec attack time, and 50 msec RT) to control the final presentation level and prevent saturation of the soundcard and/or headphones.

† The envelope of each signal was computed by low-pass filtering the full-wave rectified signal using a sixth-order Butterworth filter with a 50-Hz cutoff frequency. The envelopes were then normalized to the same mean amplitude.

## Measurement of Output SNR After Amplification With WDRC

The phase inversion technique described by Hagerman and Olofsson (2004) was used to estimate the output SNR for each listener in each condition. The technique requires that each signal under investigation be processed with WDRC three times: once using the original speech and masker mixture, once with the phase of the masker inverted before mixing, and once with the phase of the speech inverted before mixing. Therefore, when the masker-inverted signal is added to the original after processing, the masker cancels out, leaving only the speech at twice the amplitude (because it is added in phase). The opposite occurs when the speech-inverted signal is added to the original signal. As a test of the methods, when the two inverted signals were added together, everything canceled and there was no signal.

For each listener and each of the 36 conditions, the first 100 sentences of the IEEE database were processed using the same WDRC parameters as in the main experiment, yielding 10,800 sentences per listener and 7,200 sentences per condition. Onset and offset masking noise from the processed sentences were excluded from the computation of speech and masker levels.

## RESULTS

### Speech Recognition for Fixed Input SNRs

Percent correct recognition of keywords in the IEEE sentences for each condition is shown in Figure 2, with each input SNR represented in a different panel. To examine the effects of the WDRC parameters and masker type on speech recognition, a repeated-measures analysis of variance (ANOVA) was conducted with RT (long and short), channels (4, 8, and 16), masker type (steady and modulated), and input SNR (−5, 0, and +5 dB) as within-subjects factors. All statistical analyses were conducted with percent correct transformed to rationalized arcsine units (Studebaker 1985) to normalize the variance across the range. Effect size was computed using the generalized eta-squared statistic,  $\eta_G^2$  (Bakeman 2005).

Results of the ANOVA, which are shown in Table 2, indicated significant main effects of channels, masker, and SNR and

significant interactions of RT  $\times$  Channels and Masker  $\times$  SNR. Post hoc testing of all pairwise comparisons was completed using paired *t* tests and the Benjamini and Hochberg (1995) false discovery rate correction for multiple comparisons. Post hoc results revealed that the Masker  $\times$  SNR interaction arose because the difference in speech recognition for steady and modulated maskers depended on the input SNR. At −5 dB SNR, speech recognition was significantly higher ( $p < 0.001$ ,  $\eta_G^2 = 0.28$ ) with the modulated masker ( $M = 24.4\%$ ,  $SE = 2.3\%$ ) than with the steady masker ( $M = 12.6\%$ ,  $SE = 1.7\%$ ). The opposite was true at +5 dB SNR, in which speech recognition with the steady masker ( $M = 89.4\%$ ,  $SE = 1.2\%$ ) was significantly higher ( $p = 0.001$ ,  $\eta_G^2 = 0.04$ ) than with the modulated masker ( $M = 87.3\%$ ,  $SE = 1.2\%$ ). At 0 dB SNR, speech recognition was also significantly higher ( $p < 0.05$ ,  $\eta_G^2 = 0.01$ ) for the steady masker ( $M = 65.6\%$ ,  $SE = 2.4\%$ ) than for the modulated masker ( $M = 63.9\%$ ,  $SE = 2.5\%$ ) although the effect size was very small.

Post hoc tests of the data pooled across SNR and masker type revealed that the RT  $\times$  Channels interaction arose because with four channels speech recognition was significantly higher ( $p < 0.01$ ,  $\eta_G^2 = 0.02$ ) with the short RT ( $M = 58.4\%$ ,  $SE = 1.6\%$ ) than with the long RT ( $M = 56.0\%$ ,  $SE = 1.9\%$ ). An opposite, but very small, effect of RT was found with 16 channels, in which speech recognition was significantly higher ( $p < 0.01$ ,  $\eta_G^2 = 0.01$ ) with the long RT ( $M = 56.3\%$ ,  $SE = 1.8\%$ ) than with the short RT ( $M = 53.7\%$ ,  $SE = 2.3\%$ ). With eight channels, there was no significant difference ( $p > 0.05$ ) in speech recognition between the short RT ( $M = 59.1\%$ ,  $SE = 1.7\%$ ) and the long RT ( $M = 59.8\%$ ,  $SE = 2.0\%$ ).

To examine the interaction further, separate ANOVAs (one for the long RT and one for the short RT) were conducted using channels as the within-subjects factor. As expected, the ANOVAs revealed a significant effect for number of channels with both the long RT ( $F[2,46] = 6.2$ ,  $p < 0.01$ ,  $\eta_G^2 = 0.03$ ) and the short RT ( $F[2,46] = 12.1$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.06$ ). However, the pattern of results was different. Post hoc *t* tests showed that speech recognition with the long RT with 8 channels was significantly higher than with 4 and 16 channels ( $p < 0.001$ , each;  $\eta_G^2 = 0.04$  and

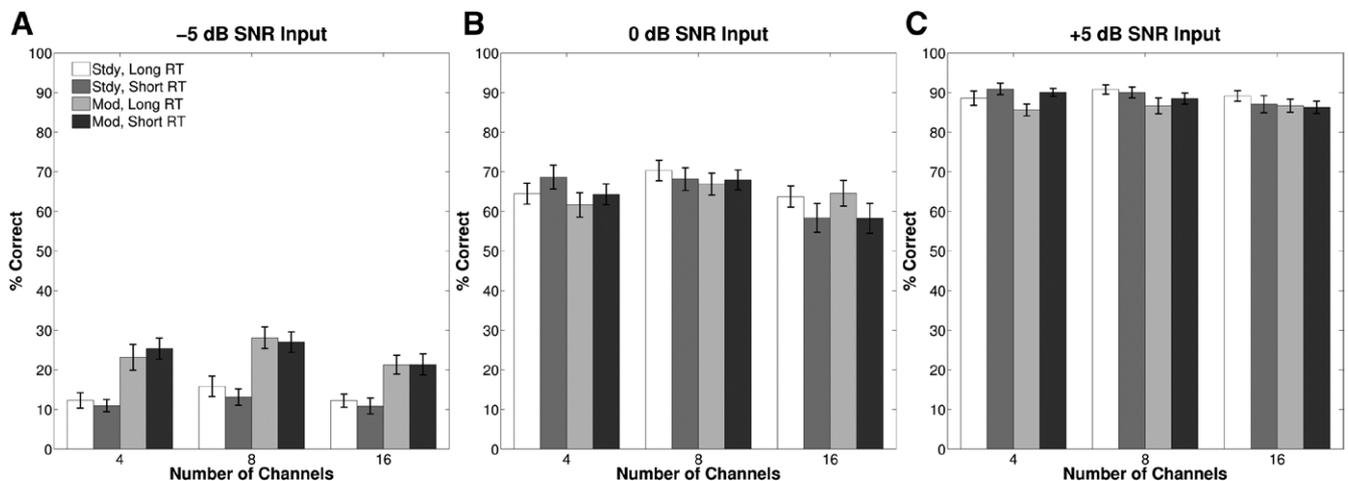


Fig. 2. Mean percent correct recognition of keywords in the Institute of Electrical and Electronics Engineers sentences for each condition. A–C, Shows results for one input signal-to-noise ratio (SNR). Bars are grouped by number of channels. Recognition in the presence of the steady masker is shown by the first pair of bars in each group and recognition in the presence of the modulated masker is shown by the second pair of bars in each group. Lighter-shaded bars in each pair represent the long release times (RTs) and darker-shaded bars represent the short RTs. In this and later figures, error bars represent the standard error of the mean.

**TABLE 2. ANOVA for the main effects of the WDRC parameters (RT and number of channels), masker type, and input SNR and their interactions on speech recognition (rationalized arcsine units)**

Source	df <sub>source</sub>	df <sub>error</sub>	F	p	$\eta_G^2$
RT	1	23	0.04	0.834	
Channels	2	46	8.93	0.001*†	0.03
Masker	1	23	21.08	<0.001*	0.02
SNR	2	46	1721.61	<0.001*	0.85
RT × Channels	2	46	9.41	<0.001*	0.01
RT × Masker	1	23	2.66	0.117	
Channels × Masker	2	46	0.16	0.850	
RT × SNR	2	46	2.55	0.089	
Channels × SNR	4	92	2.03	0.096	
Masker × SNR	2	46	154.56	<0.001*	0.09
RT × Channels × Masker	2	46	0.08	0.924	
RT × Channels × SNR	4	92	0.90	0.470	
RT × Masker × SNR	2	46	0.82	0.448	
Channels × Masker × SNR	4	92	1.90	0.118	
RT × Channels × Masker × SNR	4	92	0.53	0.711	

\*p values with statistical significance <0.05.

†p values after correction using the Greenhouse–Geisser epsilon for within-subjects factors with  $p < 0.05$  on the Mauchly test for Sphericity.  $\eta_G^2$  is the generalized eta-squared statistic for effect size. ANOVA, analysis of variance; RT, release time; SNR, signal-to-noise ratio; WDRC, wide dynamic range compression.

0.03, respectively), which were not significantly different from each other ( $p > 0.05$ ). With the short RT, speech recognition with 16 channels was significantly lower than with 4 and 8 channels ( $p < 0.001$ , each;  $\eta_G^2 = 0.06$  and  $0.07$ , respectively), which were not significantly different from each other ( $p > 0.05$ ).

Exploration of the significant interaction as a function of RT indicated that the effect sizes were very small, meaning that if the number of channels is fixed, then differences in RT are not likely to make a substantial difference for average speech recognition. When the interaction was viewed as a function of number of channels, effect sizes were larger meaning that varying the number of channels and holding RT fixed is more likely have a noticeable effect on speech recognition. However, this does not tell the whole story because when analyzed collectively, paired comparisons with false discovery rate correction among all combinations of channels and RT indicated that the best speech recognition (range, 58.4–59.8%) was obtained with eight channels (long and short RT) and four channels (short RT). Scores for these combinations were all significantly higher (individually,  $p < 0.01$ ; collectively,  $\eta_G^2 = 0.05$ ) than for the other combinations (range, 53.7–56.3%).

In summary, speech recognition with WDRC depended on the combination of RT and number of channels. Across both masker types and the three SNRs tested, the best speech recognition was obtained with eight channels, regardless of RT. Comparable performance was obtained with four channels when RT was short. Long RT with 4 and 16 channels and short RT with 16 channels had negative consequences for speech recognition. The greatest effects occurred at 0 dB SNR, in which the manipulation of RT and number of channels resulted in a 12% range of performance for the steady masker and a 10% range for the modulated masker.

### Effects of WDRC Parameters and Masker Type on Output Speech and Masker Levels

The mean output levels in dB SPL for the speech and maskers, as derived from the phase inversion technique, are plotted in Figure 3, with each input SNR represented in a different panel. To examine the effects of the WDRC parameters and masker

type on speech and masker levels, repeated-measures ANOVAs (one for speech levels and one for masker levels) were conducted with RT (long and short), channels (4, 8, and 16), masker type (steady and modulated), and input SNR (–5, 0, and +5 dB) as within-subjects factors. Interestingly, the individual factors had almost the same ordinal effect on the speech and masker levels for each of the 24 listeners despite differences in customized gain and compression settings. Consequently, almost all the main effects and their interactions, including the three- and four-way interactions, were highly significant. To have a manageable interpretation of these effects, only those with an effect size  $\eta_G^2 \geq 0.01$  will be discussed.

### Effects on Output Speech Levels

The greatest effect on amplified speech levels was the number of channels ( $F[2,46] = 67.5$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.11$ ), which systematically increased from 4 channels ( $M = 84.9$  dB,  $SE = 0.63$ ) to 8 channels ( $M = 85.7$  dB,  $SE = 0.63$ ) to 16 channels ( $M = 87.3$  dB,  $SE = 0.54$ ). RT also significantly influenced the amplified speech levels ( $F[1,23] = 385.4$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.04$ ), which were greater for the short RT ( $M = 86.6$  dB,  $SE = 0.62$ ) than for the long RT ( $M = 85.4$  dB,  $SE = 0.57$ ). Finally, the SNR of the input had a small effect ( $F[2,46] = 550.4$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.02$ ) on the amplified speech levels, despite the fact that the input speech level was held constant and that SNR was varied by changing only the level of the masker. Speech levels were significantly lower when the input SNR was –5 dB ( $M = 85.4$  dB,  $SE = 0.57$ ) than when it was 0 dB ( $M = 86.2$  dB,  $SE = 0.59$ ) or +5 dB ( $M = 86.4$  dB,  $SE = 0.61$ ).

### Effects on Output Masker Levels

As expected, the greatest effect on amplified masker levels was the SNR of the input ( $F[2,46] = 3045.5$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.43$ ). The number of channels had the next greatest effect ( $F[2,46] = 74.7$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.12$ ) followed by RT ( $F[1,23] = 315.3$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.06$ ). Just as with the amplified speech, higher masker levels were associated with 16 channels ( $M = 89.4$  dB,  $SE = 0.68$ ) compared with 8 ( $M = 87.5$  dB,

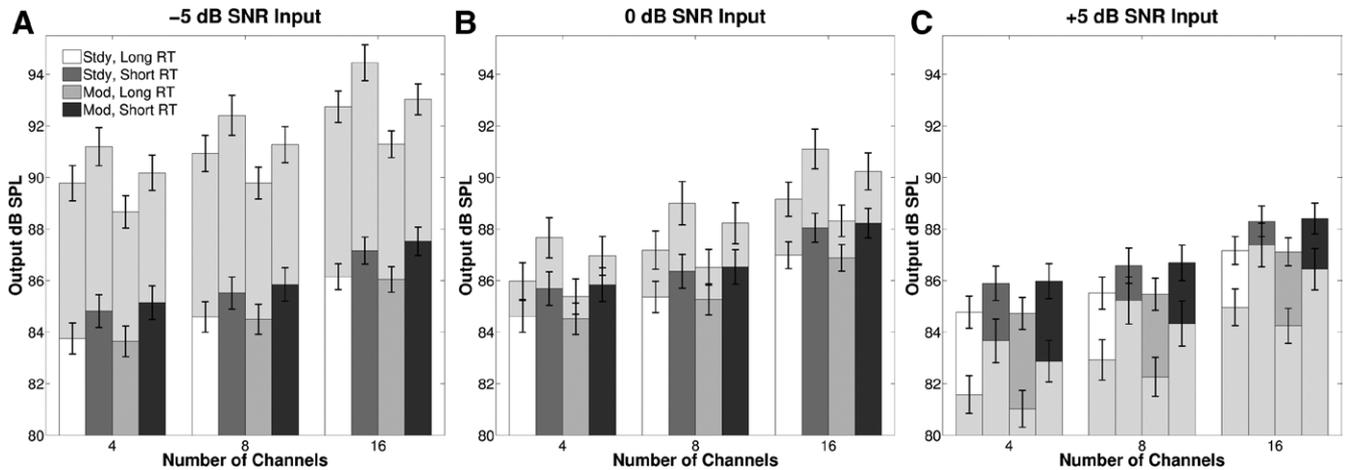


Fig. 3. Mean presentation levels in dB SPL for the speech and masker as derived from the phase inversion technique. Each panel shows results for one input signal-to-noise ratio (SNR). Bars are grouped by number of channels. Speech levels by masker type and release time (RT) are plotted using the same shading as in Figure 2. Masker levels are plotted with the corresponding speech levels using lightly gray-shaded bars in a stacked format. In (A) and (B), the masker levels are greater than the speech levels. Therefore, the relative lengths of the lightly gray-shaded bars also correspond to the amount of negative SNR. In (C), the speech levels are greater than the masker levels, so the relative lengths of the bars for speech correspond to the amount of positive SNR.

SE = 0.76) and 4 channels ( $M = 86.2$  dB, SE = 0.73). Masker levels were also higher with the short RT ( $M = 88.6$  dB, SE = 0.76) than with the long RT ( $M = 86.8$  dB, SE = 0.67). Finally, there was a small effect of masker type ( $F[1,23] = 278.5$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.02$ ); levels were higher with the steady masker ( $M = 88.2$  dB, SE = 0.73) than with the modulated masker ( $M = 87.3$  dB, SE = 0.69).

### Effects on Output SNR

The output SNRs of the amplified signals were derived from the data shown in Figure 3 and are plotted in Figure 4 in the same manner. Apart from the effect of input SNR ( $F[2,46] = 8126.2$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.95$ ), the greatest effect on the output SNR was masker type ( $F[1,23] = 247.4$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.26$ ) followed by number of channels ( $F[2,46] = 34.8$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.15$ ) and RT ( $F[1,23] = 148.0$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.14$ ).

The effect of RT interacted with both input SNR ( $F[2,46] = 227.9$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.03$ ) and masker type ( $F[1,23] = 169.1$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.01$ ). Across all conditions, output SNR was lower (more adverse) for the short RT

than for the long RT; however, the size of the difference was greater for the steady masker ( $M = 0.82$  dB, SE = 0.07) than for the modulated masker ( $M = 0.45$  dB, SE = 0.06), and the difference between RTs increased for both maskers as input SNR increased from  $-5$  dB ( $M = 0.34$  dB, SE = 0.04) to  $0$  dB ( $M = 0.60$  dB, SE = 0.05) to  $+5$  dB ( $M = 0.97$  dB, SE = 0.07). In other words, the smallest effect of RT on output SNR was observed with the modulated masker at the lowest input SNR (almost no difference) and the largest effect was observed with the steady masker at the highest input SNR (about 1 dB SNR difference).

The effect of masker type also showed an interaction with input SNR ( $F[2,46] = 99.1$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.03$ ). The modulated masker consistently led to higher (more favorable) output SNRs than the steady masker, with the difference between masker types being greater for an input SNR of  $-5$  dB ( $M = 1.33$  dB, SE = 0.10) than for  $0$  dB ( $M = 0.78$  dB, SE = 0.05) and  $+5$  dB ( $M = 0.79$  dB, SE = 0.04). Finally, the effect of channels showed an interaction with input SNR ( $F[4,92] = 40.7$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.01$ ). Across all input SNRs, output SNRs decreased (less favorable) as the number of channels increased.

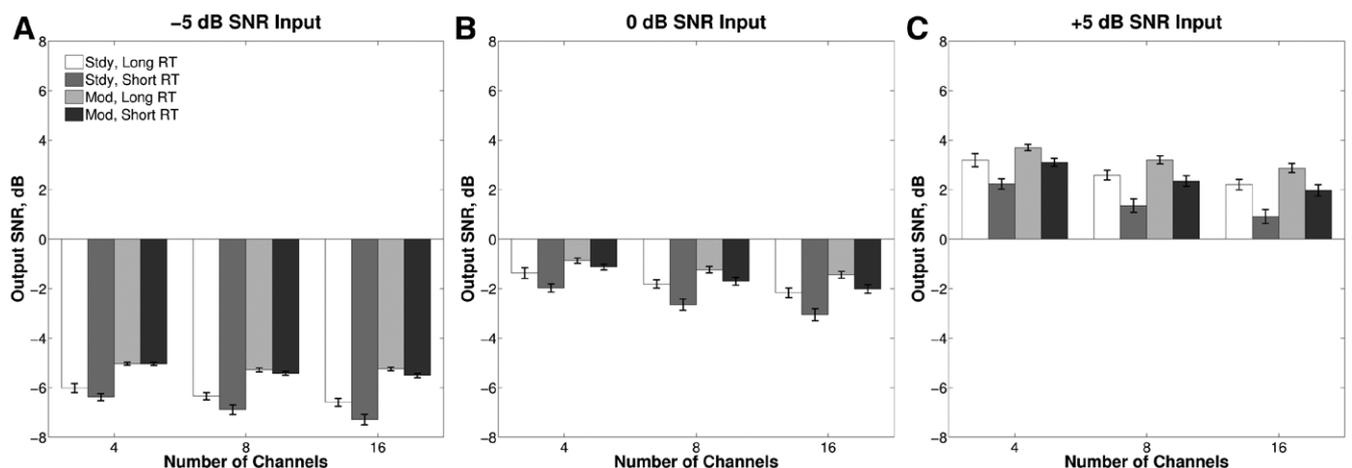


Fig. 4. A–C, Signal-to-noise ratio (SNR) of the amplified signals plotted in the same manner as in Figures 2 and 3. RT indicates release time.

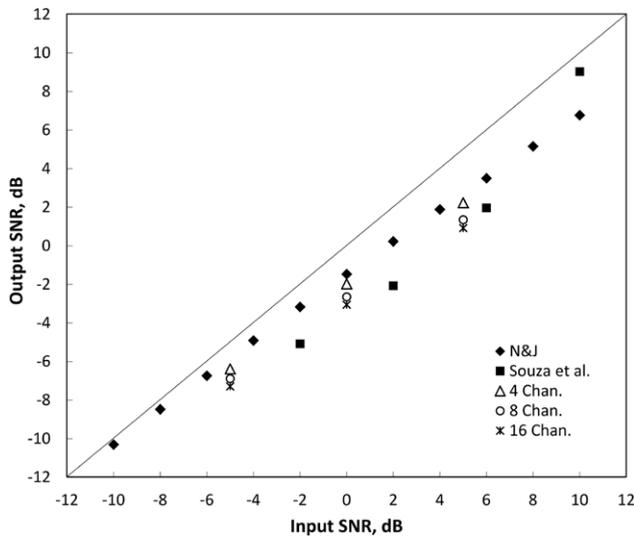


Fig. 5. The filled diamonds represent the input–output signal-to-noise ratio (SNR) differences for speech with a steady masker when processed with single-channel wide dynamic range compression and short release time as reported in the study by Naylor and Johannesson (N&J; 2009). The filled squares represent SNR differences obtained by Souza et al. (2006) in comparable conditions. The open triangles, open circles, and asterisks represent SNR differences measured in this study for speech with a steady masker processed with the short release time and 4, 8, and 16 channels, respectively. The thin diagonal line represents where the data points would lie if there were no changes in SNR.

The difference in outputs SNRs for 4 channels compared with 16 channels increased as the input SNR increased from  $-5$  dB ( $M = 0.54$  dB,  $SE = 0.09$ ) to  $0$  dB ( $M = 0.84$  dB,  $SE = 0.12$ ) to  $+5$  dB ( $M = 1.07$  dB,  $SE = 0.14$ ).§

In summary, output SNR was significantly reduced after amplification with WDRC. The reduction in output SNR relative to the input SNR was less for the modulated masker and for the lowest input SNR, where the reduction across conditions ranged  $0.03$  to  $0.51$  dB for the modulated masker and  $1.02$  to  $1.29$  dB for the steady masker. At the highest input SNR, the reduction ranged  $1.30$  to  $3.03$  dB for the modulated masker and  $1.80$  to  $4.08$  dB for the steady masker. Also, the greatest effects of RT and number of channels occurred at the highest input SNR, with 4 channels and the long RT giving the smallest

§ When input signal-to-noise ratio (SNR) varies across frequency, even by a few dB (cf. Fig. 1), frequency-shaped amplification can lead to differences in overall output SNR (e.g., if relatively greater gain is applied to a frequency region with negative SNR, then overall output SNR will be less). To measure the extent to which this effect influenced the results and subsequent analyses (e.g., Figs. 4–7), output SNR after amplification with linear gain was measured using the same method to measure output SNR after amplification with wide dynamic range compression (WDRC). For each listener and condition, the fixed gain in each channel was computed using the WDRC input–output functions, assuming a  $65$  dB input level. Results indicated that the slight differences between the input speech and masker spectra and subsequent frequency shaping did not substantially affect the output SNR. Output SNR was within  $0.1$  dB of the input SNR and ranged  $\leq 0.2$  dB across numbers of channels at each input SNR, except for the  $-5$  dB SNR modulated masker condition, which was  $0.3$  dB less than the input SNR and ranged  $0.3$  dB across numbers of channels. For each condition, output SNR was highest with 16 channels and lowest with 4 channels, which is the opposite effect that occurred with WDRC amplification. Therefore, the reported effect of numbers of channels on output SNR after WDRC amplification is slightly underestimated.

reduction in output SNR and 16 channels and the short RT giving the greatest reduction.

The reduction in output SNR with increases in input SNR is consistent with earlier work in this area (Souza et al. 2006; Naylor & Johannesson 2009). Figure 5 compares the input–output SNR differences for speech with a steady masker when processed with single-channel WDRC and short RT as reported in the studies by Souza et al. (2006) and Naylor and Johannesson (2009) to the SNR differences for the steady masker measured in this study when processed with 4-, 8-, and 16-channel WDRC and the short RT. The figure shows that measurements obtained in this study, which used clinically realistic WDRC settings, are in the range of those obtained in the previous studies, which used fixed compression kneepoints and ratios. In fact, the data points for single-channel WDRC from Naylor and Johannesson lie slightly above the data points for multichannel WDRC from the present study and together they show a consistent pattern of decreasing output SNR with increasing number of channels.

### Speech Recognition for Fixed Output SNRs

The parameters selected for WDRC processing systematically influenced the output SNR; however, it is not immediately clear the extent to which it explains the observed differences in speech recognition across conditions. If performance was determined by output SNR, then speech recognition should be equal for the different conditions when input SNR is adjusted to produce a fixed output SNR (cf. Naylor et al. 2007, 2008). Alternatively, PI functions can be used to extrapolate speech recognition at a fixed output SNR. This was accomplished by converting the proportions correct at the three input SNRs for each listener into  $z$  scores using a  $p$ -to- $z$  transform, which were then linearly regressed against the corresponding output SNRs. PI functions were derived by converting the obtained fits back into proportions using a  $z$ -to- $p$  transform. Figure 6 shows the PI functions derived using the mean of the individual slopes and intercepts for each condition.

As shown in Figure 6, the greatest differences in speech recognition between conditions occurred when the output levels of the speech and maskers were about equal,  $0$  dB SNR. For a hearing aid user in a real setting, this will occur toward the middle of the PI function when the user can hear some or most of the words, but cannot pick enough of them out of the masker for successful communication. Figure 7 displays the derived percent correct for a fixed output SNR of  $0$  dB using the same shading to represent masker type and RT as in the previous figures. Table 3 provides the results of a repeated-measures ANOVA with RT, channels, and masker type as within-subjects factors. A significant main effect of channels was found, which post hoc tests indicated was due to significantly lower performance with 4 channels ( $M = 71.3\%$ ,  $SE = 1.7\%$ ) than with 8 channels ( $M = 77.6\%$ ,  $SE = 1.7\%$ ) and 16 channels ( $M = 76.2\%$ ,  $SE = 2.2\%$ ). There was no significant difference in performance between 8 and 16 channels ( $p < 0.05$ ). There were also significant main effects of RT and masker, with better performance for the steady masker ( $[M = 82.4\%$ ,  $SE = 1.6\%]$  and  $[M = 75.0\%$ ,  $SE = 1.8\%]$  for the short and long RT, respectively) than for the modulated masker ( $[M = 74.2\%$ ,  $SE = 1.9\%]$  and  $[M = 68.6\%$ ,  $SE = 2.2\%]$  for the short and long RT, respectively). As will be discussed, the significant differences between conditions at a fixed input SNR that are no longer significant at a fixed output

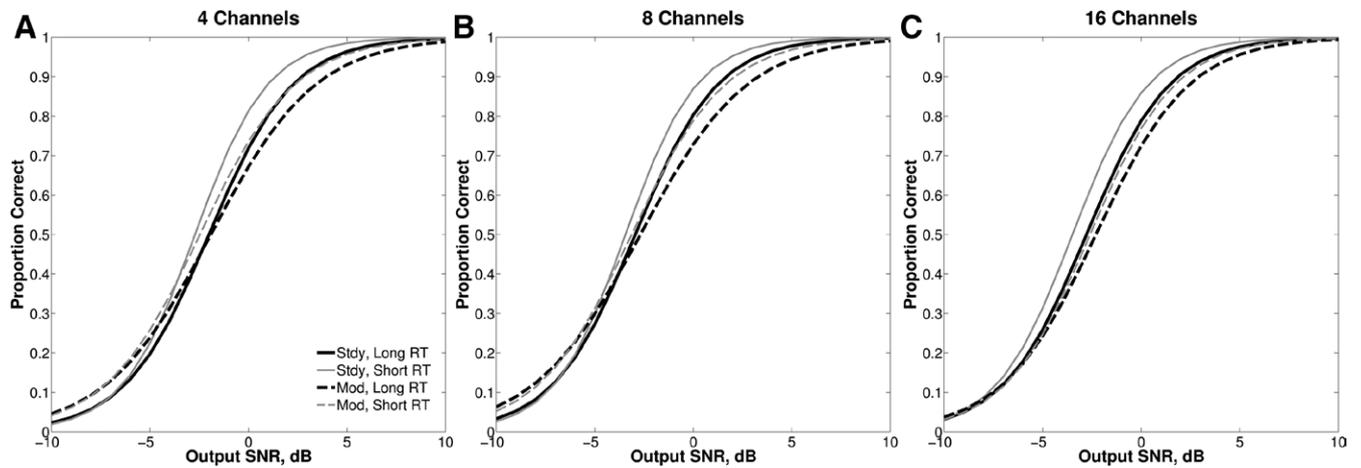


Fig. 6. A–C, Performance–intensity functions, in terms of output signal-to-noise ratio (SNR), derived from the mean of the individual slopes and intercepts for each condition. Results for steady and modulated maskers are plotted with solid and dashed lines, respectively. Results for the long and short release times (RTs) are plotted with thicker, darker lines and with thinner, lighter lines, respectively. Results are grouped by number of channels in the different panels.

SNR suggests that the former finding (cf. Fig. 2B) might be explained by the effects of WDRC processing on output SNR. To the extent that conditions have significantly lower derived scores at a fixed output SNR (e.g., long RT) suggests that other factors are involved.

Finally, an examination of the PI slopes in terms of output SNR is useful for understanding the underlying factors that influence the relative performance between conditions. For example, conditions with steep slopes indicate that the factors that influence the output SNR have a greater effect on speech recognition than conditions with shallow slopes. A repeated-measures ANOVA with RT, channels, and masker type as within-subjects factors was performed on the PI slopes (% correct per dB SNR). Results of the ANOVA, which are shown in Table 4, indicated

a significant main effect of RT, with steeper slopes for conditions with the short RT ( $M = 12.2\%$  per dB,  $SE = 0.3$ ) than with the long RT ( $M = 10.8\%$  per dB,  $SE = 0.2$ ). There was also a significant main effect of masker and a significant Channels  $\times$  Masker interaction. Post hoc tests indicated that for each number of channels, PI slopes were significantly steeper for the steady masker (grand  $M = 12.7\%$  per dB,  $SE = 0.3$ ) than for the modulated masker (grand  $M = 10.3\%$  per dB,  $SE = 0.3$ ),  $p < 0.01$ . ANOVAs conducted separately for each masker type using channels as the within-subjects variable revealed that the source of the interaction was that there was no significant effect of number of channels when the masker was steady ( $F[2,46] < 1.0$ ,  $p > 0.05$ ) but there was when the masker was modulated ( $F[2,46] = 6.5$ ,  $p < 0.01$ ). Post hoc tests showed that with the modulated masker, slopes were significantly steeper for 16 channels ( $M = 11.1\%$  per dB,  $SE = 0.4$ ) than the slopes for 4 channels ( $M = 9.9\%$  per dB,  $SE = 0.4$ ,  $p < 0.01$ ) and for 8 channels ( $M = 9.9\%$  per dB,  $SE = 0.3$ ,  $p < 0.01$ ). As will be discussed, the difference in slopes between masker types influences the relative performance at different SNRs, which can help explain why a release from masking is observed only at negative SNRs.

## DISCUSSION

The purpose of this study was to investigate the joint effects of the number of WDRC channels and their RT on speech recognition in the presence of steady and modulated maskers. For speech recognition at fixed input SNRs, the primary finding was a RT  $\times$  Channels interaction that did not depend on masker type or the input SNR. With the long RT, speech recognition with 8 channels was modestly better than with 4 and 16 channels, which were not significantly different from each other. With the short RT, speech recognition with 4 channels was as good as with 8 channels, both of which were modestly better than with 16 channels. Effect sizes were generally small. Several acoustic factors have been suggested to influence speech recognition with multichannel WDRC, including audibility, long-term output SNR, temporal distortion, and spectral distortion. Therefore, although no definitive explanations can be given, it is possible that each factor contributed to the overall results of

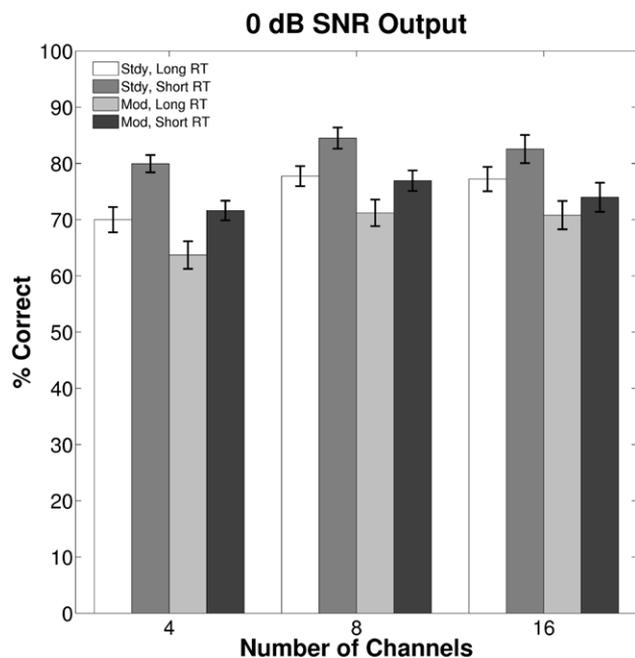


Fig. 7. Derived mean percent correct for speech recognition when output signal-to-noise ratio (SNR) was fixed at 0 dB for each condition and listener. The same shading as in the previous figures is used to represent masker type and release time (RT).

**TABLE 3. ANOVA for the main effects of RT, number of channels, and masker type, and their interactions on speech recognition (rationalized arcsine units) as derived from the performance–intensity functions in Figure 6 for a fixed output SNR of 0 dB**

Source	df <sub>source</sub>	df <sub>error</sub>	F	p	$\eta_G^2$
RT	1	23	53.48	<0.001*	0.09
Channels	2	46	11.39	<0.001*†	0.07
Masker	1	23	59.39	<0.001*	0.12
RT × Channels	2	46	3.05	0.057	
RT × Masker	1	23	6.43	0.018*	<0.01
Channels × Masker	2	46	0.17	0.838	
RT × Channels × Masker	2	46	0.01	0.986	

\*p values with statistical significance <0.05.

†p values after correction using the Greenhouse–Geisser epsilon for within-subjects factors with  $p < 0.05$  on Mauchly test for Sphericity.  
ANOVA, analysis of variance; RT, release time; SNR, signal-to-noise ratio.

this study and that individuals were influenced by the different factors to varying extents. For example, although short RT may have been relatively more beneficial for some listeners due to the rapid increase in gain for low-intensity parts of speech, it may have been relatively more disadvantageous for others by increasing the noise floor (*i.e.*, ambient noise in the recorded signal in this study or ambient room noise combined with microphone noise in the real world), thereby reducing the salience of information conveyed by temporal envelope in the presence of the masker (Plomp 1988; Drullman et al. 1994; Souza & Turner 1998; Stone & Moore 2003, 2007; Jenstad & Souza 2005, 2007; Souza & Gallun 2010). Similarly, although use of multiple channels may have been relatively more beneficial for some listeners because gain for low-intensity speech could be rapidly increased in frequency-specific regions when and where it was needed the most (e.g., Yund & Buckles 1995a; Henning & Bentler 2008), it may have been relatively more disadvantageous for others because the increased gain for low-intensity frequency regions may have reduced the salience of information conveyed by spectral shape, including formant peaks and spectral tilt (Yund & Buckles 1995b; Franck et al. 1999; Bor et al. 2008; Souza et al. 2012). Furthermore, if the low-intensity temporal and spectral segments primarily consisted of noise, increased masking may have resulted (Plomp 1988; Edwards 2004). Consistent with all these effects, an analysis of speech and masker levels using the phase inversion technique indicated that while overall speech levels were higher for conditions with short RT and more channels, so were overall masker levels, for a net decrease in long-term output SNR. Therefore, audibility of certain speech cues for some listeners likely came at the perceptual cost of a more degraded signal (Souza & Turner 1998, 1999; Davies-Venn et al. 2009; Kates 2010).

When considering changes in long-term output SNR, it is important to recognize that single-microphone signal-processing strategies, such as noise reduction and WDRC, cannot change the short-term SNR. With WDRC, because compressors respond to the overall level in each channel regardless of the individual contributions from the speech and masker, changes in masker level occur at the same time and almost to the same extent as changes in speech level (see Supplemental Digital Content, <http://links.lww.com/EANDH/A171>, for a further demonstration of this effect). Consequently, the short-term SNR is essentially unchanged when comparing the acoustic effects of RT and number of channels. As demonstrated by Naylor and Johannesson (2009), changes in the long-term SNR result from the application of nonlinear gain across two mixed signals with varying amounts of modulation. That is, if a time segment contains a mix of lower-intensity speech and a higher-intensity masker, the short-term SNR will be maintained because they will both receive the same amount of gain in dB. However, the absolute change in power will be greater for the latter. Therefore, when the long-term average is computed across a number of these nonlinearly amplified intervals, the overall level of the more modulated signal will be lower than that of the less-modulated signal. Across frequencies, the target speech was more modulated than the (modulated) ICRA noise, both of which were more modulated than the steady masker. Therefore, everything else being equal, less gain would have been applied to the target speech than to the steady and modulated maskers, respectively (*cf.*, Verschuure et al. 1996). Increasing the effectiveness of compression by increasing the compression ratio, decreasing RT, and/or increasing the number of channels enhances the effect (Souza et al. 2006; Naylor et al. 2007; Naylor & Johannesson 2009).

**TABLE 4. ANOVA for the main effects of RT, number of channels, and masker type, and their interactions on the slopes of the performance–intensity functions (percent correct/output dB SNR)**

Source	df <sub>source</sub>	df <sub>error</sub>	F	p	$\eta_G^2$
RT	1	23	24.51	<0.001*	0.08
Channels	2	46	1.06	0.353	
Masker	1	23	90.85	<0.001*	0.20
RT × Channels	2	46	0.30	0.742	
RT × Masker	1	23	0.74	0.400	
Channels × Masker	2	46	5.37	0.008*	0.02
RT × Channels × Masker	2	46	0.30	0.745	

\*p values with statistical significance <0.05.

ANOVA, analysis of variance; RT, release time; SNR, signal-to-noise ratio.

Considering that short-term SNR is unaffected by WDRC, the perceptual relevance of changes in long-term SNR is not clear. As indicated in the previous paragraph, decreases in long-term SNR in a particular channel result from the amplification of time segments that have low-intensity speech and a relatively higher-intensity masker. The low-intensity speech segments likely contribute very little to speech recognition because they represent either the noise floor of the speech signal or weak speech cues that are masked by the noise. On the other hand, applying greater gain to these segments relative to segments that contain the peaks of speech effectively decreases the modulation depth of the temporal envelope in the channel, which can have negative consequences for perception. Therefore, reductions in long-term SNR should be closely associated with reductions in amplitude contrast and temporal envelope shape.

To test the relationship between changes in long-term SNR and temporal envelope shape before and after WDRC, reductions in temporal envelope fidelity were quantified using the Envelope Difference Index (EDI) (e.g., Fortune et al. 1994; Jenstad & Souza 2005, 2007; Souza et al. 2006). EDI has a single value that varies from 0 (maximum similarity) to 1 (maximum difference) and is computed by averaging the absolute differences between the envelopes of two signals.‡ Using the prescriptive fitting parameters for a listener with a hearing loss representative of the group mean (Listener 13 in Table 1), the sentence “A tusk is used to make costly gifts” was processed with the six combinations of RT and number of channels at 0 dB input SNR for each masker type. The phase inversion technique was used to extract the processed speech and masker signals. The EDI between the input speech signal and the amplified speech-in-masker signal and the long-term SNR were computed for each 1/3-octave band from 250 to 8000 Hz. Mean EDI across the 1/3-octave bands for the speech-in-masker followed the same general pattern described for long-term output SNR—it was poorer (greater) for the short RT than for the long RT and became more adverse as the number of channels increased. This is the same pattern observed by Souza et al. (2006), who computed EDI for broadband speech signals (Jenstad & Souza 2007). Individual correlations (32 total) between EDI and long-term output SNR for each of the bands (16) and masker types (2) ranged from  $r(5) = -0.84$  to  $-1.00$  ( $p < 0.05$ ). Consequently, multiple regression of EDI against SNR, using frequency band and masker type as covariates, was highly significant ( $R^2 = 0.93$ ,  $F[1,190] = 2434.8$ ,  $p < 0.001$ ). It is important to recognize that the relationships shown here apply only to WDRC processing when noise is present and that causality between the two is not implied because EDI can be manipulated separately from SNR. For example, changes in EDI, but not in SNR, have been documented for speech-in-quiet processed with nonlinear amplification (Fortune et al. 1994; Jenstad & Souza 2005; Souza et al. 2010).

The association between long-term output SNR and temporal envelope fidelity for the speech-in-masker processed with WDRC can provide insight about factors that may have influenced speech recognition as RT and number of channels varied. For example, when performance across combinations of these WDRC parameters was compared at a fixed output SNR (Fig. 7), fidelity of the temporal envelopes, as quantified by the EDI, was also about equal across conditions. The lack of a significant difference in estimated speech recognition at a fixed output SNR between 8 and 16 channels was in contrast to the significant

difference at a fixed input SNR (worse recognition with 16 channels than with 8 channels). This suggests that the latter finding might be explained by the effects of WDRC processing on temporal envelope and/or long-term output SNR. However, the involvement of other factors is suggested for conditions where estimated speech recognition at equal long-term output SNRs and EDIs was significantly lower compared with other conditions. For example, even though long-term output SNR and EDI were equated across conditions, derived speech recognition was significantly higher for conditions that promoted audibility of short-term speech information and was lower for conditions that were less favorable for audibility. Specifically, derived performance for 4 channels was lower than for 8 and 16 channels and was consistently lower for long RT compared with short RT, which suggests that short-term audibility may have been a factor that also influenced speech recognition.

In addition to changes in the fidelity of temporal envelope shape, Stone and Moore (2007) demonstrated how WDRC could affect relationships between temporal envelopes across bands within the speech signal and relationships between the speech and masker envelopes. Common amplitude fluctuation between the speech and masker is yet another negative acoustic consequence to consider because changes in gain for the two signals covary the most with short RT and many channels (Stone & Moore 2004). This phenomenon, cross-modulation, may lead to decreased speech recognition due to increased perceptual fusion of the speech and masker (Nie et al. 2013) and/or to an increased difficulty detecting amplitude fluctuations in the speech (Stone et al. 2012). Stone and Moore describe a technique, Across-Source Modulation Correlation (ASMC), for measuring cross-modulation in which Pearson correlations are used to quantify the degree of association between the temporal envelopes (in log units) of the speech and masker.§ To examine how cross-modulation may have influenced the results of this study, for each listener, Across-Source Modulation Correlation in each 1/3-octave band from 250 to 8000 Hz was computed for the same speech and masker mix that was used for the computation of EDI.

Results of the ASMC analysis are plotted in the same manner as earlier figures and are separated according to mean data for the 250 to 4000 Hz bands (Fig. 8) and the 5000 to 8000 Hz bands (Fig. 9) because the size and pattern of correlations were noticeably split across these two frequency regions. For this same reason, the scales of the two figures span different ranges to highlight differences between conditions. The different patterns across frequency regions likely occurred because gain and compression ratios were greatest in the high frequencies due to increased thresholds as well as decreased input levels in this region (cf., Fig. 1). The correlations between the high-frequency bands (Fig. 9) were negative because an increase in intensity for one signal would cause a decrease in gain that tended to decrease intensity for the other signal. The effect was most pronounced

§ Another similar measure described by Stone and Moore (2007), the Within-Signal Modulation Correlation, summarizes the relationships between the temporal envelopes in each band of the amplified speech signal. It is hypothesized that multichannel wide dynamic range compression could reduce the degree of correlation between the speech envelopes across frequency, thereby making it more difficult to perceptually integrate them in the presence of a masker. An analysis of Within-Signal Modulation Correlation did not reveal noticeable differences across conditions. As expected, correlations for all conditions were highest for adjacent bands compared with distant bands. Mean correlations were of similar magnitude reported by Stone and Moore and ranged 0.337 to 0.346 across conditions.

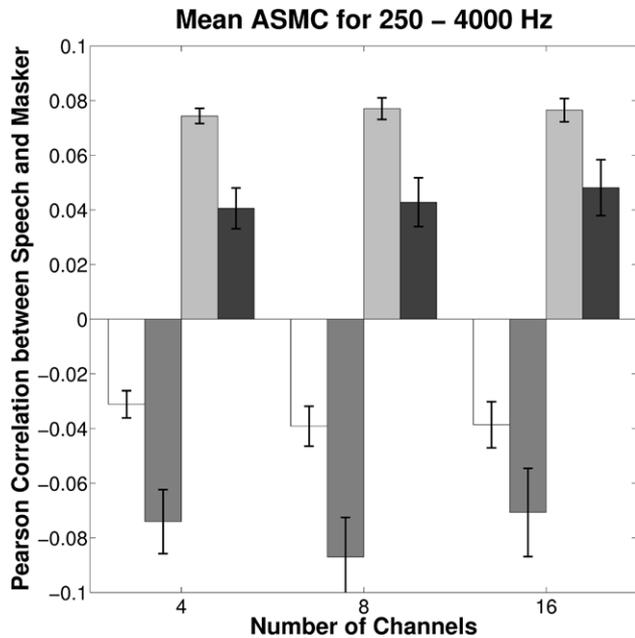


Fig. 8. Across-Source Modulation Correlations (ASMC) averaged over listeners and the 1/3-octave bands from 250 to 4000 Hz between a speech sample and masker at 0 dB input signal-to-noise ratio. For reference, correlations of the input signals before amplification were 0.00 and 0.09 for the steady and modulated maskers, respectively. Therefore, the absolute change in the magnitude of the correlations is greater for the steady masker at each release time. The same shading as in the previous figures is used to represent masker type and release time (see also, Fig. 9). Error bars represent the standard error of the mean across listeners.

for conditions with the steady masker and the short RT because the compressors in each channel were activated primarily by the speech peaks. The onsets of the speech peaks caused channel gain to decrease and the short RT allowed it to recover quickly upon their cessation, thereby modulating the level of the masker out of phase with the changing level of the speech (Stone & Moore 2007). The range of correlations in the high-frequency bands for conditions with the modulated masker and conditions with the steady masker paired with slow RT is of similar magnitude as the mean correlations reported by Stone and Moore (2007), who examined temporal envelope statistics for several of their earlier studies that used another talker as the masker and a positive SNR. In contrast, the magnitude of the correlations for the steady masker with short RT is surprisingly high. Because speech recognition for these conditions was as good as or better than the other conditions, despite the high degree of cross-modulation, and because SLs were relatively lower for the high-frequency band, the influence of this finding on the perceptual outcomes of the present experiment is unclear. In the low-frequency bands where gain and compression were less, the correlations for the steady masker and modulated masker were of comparable magnitude, but with opposite sign, which is partially explained by difference in the correlations of the input signals before amplification (0.00 and 0.09 for the steady and modulated maskers, respectively). As explained by Stone and Moore, differences in the phase of the modulations (sign of the correlations between temporal envelopes) between the speech and masker are hypothesized to have about the same effect on speech recognition.

Finally, the effect of masker type on speech recognition was the same when examined in terms of input or output

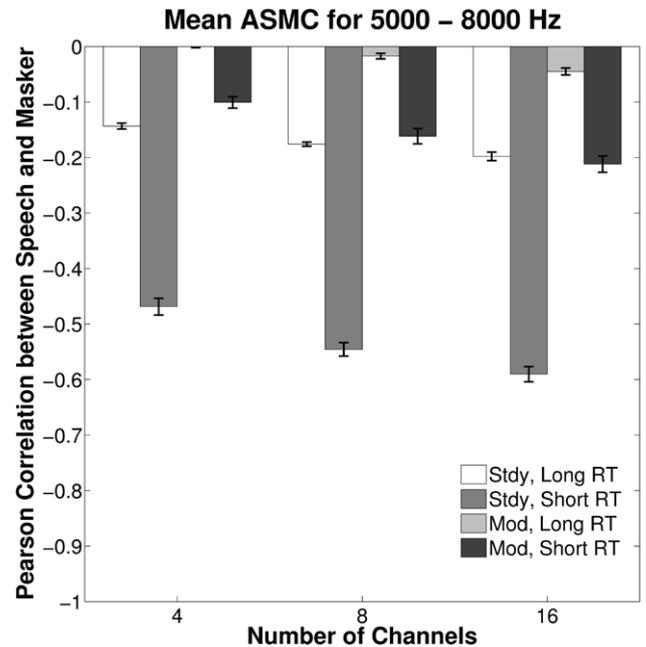


Fig. 9. Across-Source Modulation Correlations (ASMC) averaged over listeners and 1/3-octave bands from 5000 to 8000 Hz. For reference, correlations of the input signals before amplification were 0.02 and 0.07 for the steady and modulated maskers, respectively. RT indicates release time.

SNR, which further suggests that factors other than temporal envelope distortion (specifically, EDI) may have been responsible. The slopes of the PI functions (Fig. 6) were significantly steeper for the steady masker than for the modulated masker. This phenomenon reflected a release from masking when the masker was modulated at low SNRs, where spectro-temporal glimpses of the speech in the masker were most important for perception. At high SNRs, the relationship reversed and speech recognition was more favorable for the steady masker. This same reversal, which has been reported for normal-hearing and hearing-impaired listeners under various conditions (e.g., Bernstein & Grant 2009; Oxenham & Simonson 2009; Smits & Festen 2013), has been explained by perceptual interference or masking caused by similarities between the speech and modulated masker (Brungart 2001; Kwon & Turner 2001; Freyman et al. 2004). In the context of WDRC processing, another explanation is that as SNR increased, the benefit of listening in the dips of the modulated masker progressively lessened, whereas at the same time, the additional peaks in the temporal envelope from the modulated masker were still sufficient to activate the compressor and decrease gain for the speech that followed. In addition, input masker levels were lower at the higher SNR. Therefore, at high frequencies with moderate SNHL, the steady masker may have been inaudible or barely audible in comparison to the peaks of the modulated masker, which had a higher crest factor. This may have led to the percept of random interjections of frication-like noise from the masker that would have made it difficult to detect the presence of a true speech sound in the affected channels.

### Generalizability to Real-World Settings

The results of the present study need to be considered in the context of at least three factors that could influence their generalizability to real-world settings. First, as presentation level

changes, the relative importance of factors related to audibility and envelope distortion may change (Davies-Venn et al. 2009; Kates 2010). At low presentation levels, parameters that promote audibility, namely short RT and many channels, may lead to more favorable speech recognition. The opposite may be true at high presentation levels because minimizing envelope distortion may be more important for speech recognition. Second, the outcomes may differ when listeners experience WDRC with bilaterally fitted hearing aids, where the preservation of interaural level cues may also be important. For example, Moore et al. (2010) found that speech recognition with five-channel WDRC was better for slow compression time-constants than for fast when the target speech came from one side of the head and interfering speech came from the other side. Third, it is not known how the results would change if listeners had opportunity to get used to the different WDRC settings over an extended period of real-world listening with wearable devices (e.g., Yund & Buckles 1995b; Kuk 2001). For example, Yund and Buckles (1995b) provide evidence of long-term learning with 8- and 16-channel WDRC, in which they tracked improvements in listeners' ability to identify nonsense syllables in noise over the course of 1 year of routine hearing aid use outside the laboratory.

### CONCLUSIONS

It is important to consider the effects of WDRC processing on speech recognition because the choice of parameters may result in clinically significant consequences for individuals. For example, in this study, the range of speech recognition scores for the group means was 10 to 12% across the different combinations of RT and number of channels when the input SNR was 0 dB. Across individuals, the median range was twice as large. One of the primary findings of this study was that the number of channels had a small effect when analyzed at each level of RT, with results suggesting that selecting 8 channels for a given RT might be the safest choice. Across conditions, the best speech recognition occurred with 8 channels for 14 of the 24 listeners. The best speech recognition occurred with 4 channels for 6 listeners and with 16 channels for 4 listeners, which includes one tie between the two. Effects were even smaller for RT, with results suggesting that short RT was better when only 4 channels were used and that long RT was slightly better when 16 channels were used. Individual differences in how listeners were influenced by audibility, output SNR, temporal distortion, and spectral distortion may have contributed to the effects found in this study. Because only general suppositions could be made for how each of these factors may have influenced the overall results of this study, future research would benefit from exploring the predictive value of these and other factors in selecting the processing parameters that maximize speech recognition for individuals.

### ACKNOWLEDGMENTS

The hearing aid simulator would not have been possible without the programming efforts of Minseok Kwon, code from James Kates, conversations with Susan Scollie and Steve Beaulac about the DSL graphical user interface, and program testing by the Hearing and Amplification Research Lab at Boys Town National Research Hospital. The authors thank Jessica Warnshuis, Yonit Shames, Kelsey McDonald, and Patricia Rhymer for their help with data collection. They also thank Evelyn Addae-Mensah, Alyson Harmon, Brian Moore, and Benjamin Hornsby for their useful suggestions on an earlier version of this article.

This work was supported by RC1 DC010601 from the National Institutes of Health—National Institute on Deafness and Other Communication.

The authors declare no other conflict of interest.

Address for correspondence: Joshua M. Alexander, Department of Speech, Language, and Hearing Sciences, Purdue University, Lyles-Porter Hall, 715 Clinic Drive, West Lafayette, IN 47907, USA. E-mail: alexan14@purdue.edu

Received October 29, 2012; accepted September 10, 2014.

### REFERENCES

- American National Standards Institute. (2003). ANSI S3.22–2003, Specification of Hearing Aid Characteristics. New York, NY: American National Standards Institute.
- American National Standards Institute. (2004). ANSI S1.11–2004, Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters. New York, NY: American National Standards Institute.
- Bacon, S. P., & Brandt, J. F. (1982). Auditory processing of vowels by normal-hearing and hearing-impaired listeners. *J Speech Hear Res*, *25*, 339–347.
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behav Res Methods*, *37*, 379–384.
- Bench, J., Kowal, A., Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br J Audiol*, *13*, 108–112.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*, *57*, 289–300.
- Bernstein, J. G., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*, *125*, 3358–3372.
- Bor, S., Souza, P., Wright, R. (2008). Multichannel compression: Effects of reduced spectral contrast on vowel identification. *J Speech Lang Hear Res*, *51*, 1315–1327.
- Brennan, M. A., McCreery, R. W., Kopun, J., et al. (2012). Signal processing preference for music and speech in adults and children with hearing loss. 39th Annual Meeting of the American Auditory Society, March 8–10, 2012, Scottsdale, AZ.
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am*, *109*, 1101–1109.
- Byrne, D., Dillon, H., Tran, K., et al. (1994). An international comparison of long-term average speech spectra. *J Acoust Soc Am*, *96*, 2108–2120.
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *J Acoust Soc Am*, *119*, 1562–1573.
- Davies-Venn, E., Souza, P., Brennan, M., et al. (2009). Effects of audibility and multichannel wide dynamic range compression on consonant recognition for listeners with severe hearing loss. *Ear Hear*, *30*, 494–504.
- Dreschler, W. A., Verschuure, H., Ludvigsen, C., et al. (2001). ICRA Noises: Artificial noise signals with speech-like spectral and temporal properties for hearing aid assessment. *Audiology*, *40*, 148–157.
- Drullman, R., Festen, J. M., Plomp, R. (1994). Effect of temporal envelope smearing on speech recognition. *J Acoust Soc Am*, *95*, 1053–1064.
- Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *J Acoust Soc Am*, *74*, 739–743.
- Edwards, B. (2004). Hearing aids and hearing impairment. In S. Greenberg, A. N. Popper, R. R. Fay (Eds.), *Speech Processing in the Auditory System*. New York, NY: Springer-Verlag.
- Festen, J. M., & 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*, 1725–1736.
- Fortune, T. W., Woodruff, B. D., Preves, D. A. (1994). A new technique for quantifying temporal envelope contrasts. *Ear Hear*, *15*, 93–99.
- Franck, B. A. M., van Kreveld-Bos, C. S. G. M., Dreschler, W. (1999). Evaluation of spectral compression in hearing aids, combined with phonemic compression. *J Acoust Soc Am*, *106*, 1452–1464.
- Freyman, R., Balakrishnan, U., Helfer, K. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J Acoust Soc Am*, *115*, 2246–2256.
- George, E. L., Festen, J. M., Houtgast, T. (2006). Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*, *120*, 2295–2311.
- Hagerman, B., & Olofsson, A. (2004). A method to measure the effect of noise reduction algorithms using simultaneous speech and noise. *Acta Acoustica*, *90*, 356–361.

- Henning, R. L., & Bentler, R. A. (2008). The effects of hearing aid compression parameters on the short-term dynamic range of continuous speech. *J Speech Lang Hear Res*, *51*, 471–484.
- Henry, B. A., & Turner, C. W. (2003). The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *J Acoust Soc Am*, *113*, 2861–2873.
- Hopkins, K., Moore, B. C., Stone, M. A. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *J Acoust Soc Am*, *123*, 1140–1153.
- Hornsby, B. W., & Ricketts, T. A. (2001). The effects of compression ratio, signal-to-noise ratio, and level on speech recognition in normal-hearing listeners. *J Acoust Soc Am*, *109*, 2964–2973.
- International Collegium of Rehabilitative Audiology. (2010). The International Collegium of Rehabilitative Audiology. Retrieved October 27, 2010, from [http://www.icra.nu/?page\\_id=58](http://www.icra.nu/?page_id=58).
- Jenstad, L. M., & Souza, P. E. (2005). Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility. *J Speech Lang Hear Res*, *48*, 651–667.
- Jenstad, L. M., & Souza, P. E. (2007). Temporal envelope changes of compression and speech rate: Combined effects on recognition for older adults. *J Speech Lang Hear Res*, *50*, 1123–1138.
- Jenstad, L. M., Van Tasell, D. J., Ewert, C. (2003). Hearing aid troubleshooting based on patients' descriptions. *J Am Acad Audiol*, *14*, 347–360.
- Kates, J. M. (2008). *Digital Hearing Aids*. San Diego, CA: Plural.
- Kates, J. M. (2010). Understanding compression: Modeling the effects of dynamic-range compression in hearing aids. *Int J Audiol*, *49*, 395–409.
- Kopun, J., McCreery, R. W., Brennan, M. A., et al. (2012). Effects of Exposure on Speech Recognition with Nonlinear Frequency Compression. 39th Annual Meeting of the American Auditory Society, March 8–10, 2012, Scottsdale, AZ.
- Kuk, F. K. (2001). Adaptation to enhanced dynamic range compression (EDRC): Examples from the Senso P38 digital hearing aid. *Sem Hear*, *22*, 161–171.
- Kwon, B. J., & Turner, C. W. (2001). Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference? *J Acoust Soc Am*, *110*, 1130–1140.
- Leek, M. R., Dorman, M. F., Summerfield, Q. (1987). Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*, *81*, 148–154.
- Lorenzi, C., Gilbert, G., Carn, C., et al. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc Natl Acad Sci U S A*, *103*, 18866–18869.
- McCreery, R. W., Alexander, J. M., Brennan, M. A., et al. (2014). The influence of audibility and short-term audio-visual exposure on speech recognition with nonlinear frequency compression for children and adults with hearing loss. *Ear Hear*, *35*, 440–447.
- McCreery, R. W., Brennan, M. A., Hoover, B., et al. (2013). Maximizing audibility and speech recognition with nonlinear frequency compression by estimating audible bandwidth. *Ear Hear*, *34*, e24–e27.
- Moore, B. C. (2008). The choice of compression speed in hearing aids: Theoretical and practical considerations and the role of individual differences. *Trends Amplif*, *12*, 103–112.
- Moore, B. C., Füllgrabe, C., Stone, M. A. (2010). Effect of spatial separation, extended bandwidth, and compression speed on intelligibility in a competing-speech task. *J Acoust Soc Am*, *128*, 360–371.
- Moore, B. C., Peters, R. W., Stone, M. A. (1999). Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. *J Acoust Soc Am*, *105*, 400–411.
- Naylor, G., & Johannesson, R. B. (2009). Long-term signal-to-noise ratio at the input and output of amplitude-compression systems. *J Am Acad Audiol*, *20*, 161–171.
- Naylor, G., Johannesson, R. B., Rønne, F. M. (2007). An investigation of effective SNR-change through amplitude-compression hearing aids. *Proc ISAAR*.
- Naylor, G., Rønne, F. M., Johannesson, R. B. (2008). Perceptual correlates of the long-term SNR change caused by fast-acting compression. 2008 International Hearing Aid Conference, August 13–17, 2008, Lake Tahoe, CA.
- Nie, Y., Svec, A., Nelson, P. (2013). The effects of temporal envelope confusion on listeners' phoneme and word recognition. *Proc Meet Acoust*, *19*, 060111.
- Olsen, H. L., Olofsson, A., Hagerman, B. (2004). The effect of presentation level and compression characteristics on sentence recognition in modulated noise. *Int J Audiol*, *43*, 283–294.
- Oxenham, A. J., & Simonson, A. M. (2009). Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference. *J Acoust Soc Am*, *125*, 457–468.
- Plomp, R. (1988). The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *J Acoust Soc Am*, *83*, 2322–2327.
- Rothauser, E. H., Chapman, W. D., Guttman, N., et al. (1969). IEEE recommended practice for speech quality measurements. *IEEE Trans Audio Electroacoust*, *17*, 227–246.
- Smits, C., & Festen, J. M. (2013). The interpretation of speech reception threshold data in normal-hearing and hearing-impaired listeners: II. Fluctuating noise. *J Acoust Soc Am*, *133*, 3004–3015.
- Souza, P. E. (2002). Effects of compression on speech acoustics, intelligibility, and sound quality. *Trends Amplif*, *6*, 131–165.
- Souza, P. E., & Gallun, F. (2010). Amplification and consonant modulation spectra. *Ear Hear*, *31*, 268–276.
- Souza, P., Hoover, E., Gallun, F., et al. (2010). Assessing envelope distortion in clinically-fit hearing aids. *International Hearing Aid Conference*, Tahoe City, CA.
- Souza, P. E., Jenstad, L. M., Boike, K. T. (2006). Measuring the acoustic effects of compression amplification on speech in noise. *J Acoust Soc Am*, *119*, 41–44.
- Souza, P. E., & Turner, C. W. (1998). Multichannel compression, temporal cues, and audibility. *J Speech Lang Hear Res*, *41*, 315–326.
- Souza, P. E., & Turner, C. W. (1999). Quantifying the contribution of audibility to recognition of compression-amplified speech. *Ear Hear*, *20*, 12–20.
- Souza, P., Wright, R., Bor, S. (2012). Consequences of broad auditory filters for identification of multichannel-compressed vowels. *J Speech Lang Hear Res*, *55*, 474–486.
- Stone, M. A., Füllgrabe, C., Moore, B. C. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *J Acoust Soc Am*, *132*, 317–326.
- Stone, M. A., & Moore, B. C. (2003). Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task. *J Acoust Soc Am*, *114*, 1023–1034.
- Stone, M. A., & Moore, B. C. (2004). Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task. *J Acoust Soc Am*, *116*, 2311–2323.
- Stone, M. A., & Moore, B. C. (2007). Quantifying the effects of fast-acting compression on the envelope of speech. *J Acoust Soc Am*, *121*, 1654–1664.
- Stone, M. A., & Moore, B. C. (2008). Effects of spectro-temporal modulation changes produced by multi-channel compression on intelligibility in a competing-speech task. *J Acoust Soc Am*, *123*, 1063–1076.
- Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J Speech Hear Res*, *28*, 455–462.
- Summers, V., & Leek, M. R. (1994). The internal representation of spectral contrast in hearing-impaired listeners. *J Acoust Soc Am*, *95*, 3518–3528.
- Turner, C. W., Holte, L. A., Relkin, E. (1987). Auditory filtering and the discrimination of spectral shapes by normal and hearing-impaired subjects. *J Rehabil Res Dev*, *24*, 229–238.
- Van Tasell, D. J., Fabry, D. A., Thibodeau, L. M. (1987). Vowel identification and vowel masking patterns of hearing-impaired subjects. *J Acoust Soc Am*, *81*, 1586–1597.
- Verschuure, J., Mass, A. J. J., Stikvoort, E., et al. (1996). Compression and its effect on the speech signal. *Ear Hear*, *17*, 162–175.
- Yund, E. W., & Buckles, K. M. (1995a). Multichannel compression hearing aids: Effect of number of channels on speech discrimination in noise. *J Acoust Soc Am*, *97*, 1206–1223.
- Yund, E. W., & Buckles, K. M. (1995b). Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids. *J Acoust Soc Am*, *97*, 1224–1240.

### **ILLUSTRATIVE EXAMPLE OF THE EFFECTS OF WDRC ON SHORT-TERM SPEECH AND MASKER LEVELS**

Using the prescriptive fitting parameters for a listener with a hearing loss representative of the group mean (Listener 13 in Table I), the sentence “A tusk is used to make costly gifts” was processed with the 6 combinations of RT and number of channels at 0 dB input SNR. The phase inversion technique was used to extract the processed speech and masker signals. The extracted signals for the different processing conditions were compared with another to better understand how they affect the short-term output levels and consequently influence perception. Most of the level differences in the processed signals between conditions occurred at high frequencies, where nonlinear amplification was greatest for this particular example. Therefore, illustrative examples for the band centered at 8000 Hz are shown in Figures S1 – S6, which compare the 1/3-octave band levels using a 20-ms Hann window with 50% overlap. The effects that will be described also apply to lower frequencies to varying extents depending on the amount of nonlinear gain and the input SNR. Figures S1 and S2 plot the speech and steady masker levels for 4- and 16-channel WDRC, respectively, using solid black (speech) and dotted gray (masker) lines. Output levels for the long and short RT are plotted in panels (a) and (b), respectively, and differences between the two are plotted in panel (c), along with the differences in short-term SNR (light dotted black line). Output levels with the modulated maskers are plotted in the same way for 4- and 16-channel WDRC in Figures S3 and S4, respectively. A phonetic transcription of the speech at each time segment is shown at the top of the panels. Consistent with the ANOVA results from the main study, levels for both the speech and maskers in each example were higher for the short RT than for the long RT, in part because the short RT allowed gain in the high-frequency channel to recover more quickly for lower-intensity segments after going into compression for the fricatives /s/ and /z/ and the intense bursts of the stops /t/ and /d/.

For the steady and modulated maskers, respectively, Figures S5 and S6 compare the level differences for 16- and 4-channel WDRC for the long RT (panel a) and the short RT (panel b). As shown, speech and masker levels increased as the number of channels increased, especially following the fricatives and stop consonants. In

*Ear and Hearing Supplement to Alexander and Masterson*  
*“Effects of WDRC release time and number of channels on output SNR and speech recognition”*

fact, the ANOVA results from the main study had indicated that the number of channels had the greatest effect on the amplified speech and masker levels (e.g., compare the effect of going from 4 to 16 channels in Figure S5a to the effect of changing RT in Figure S1c). This occurred because with fewer and wider channels, lower-intensity segments of the spectrum often shared the same channel as higher-intensity segments. Because higher-intensity segments controlled the compressor, less-than-ideal gain was applied to lower-intensity segments. As more and narrower channels were added, the likelihood increased that lower-intensity segments fell within their own channel and received additional gain. Figures S5 and S6 show that the differences in level for 16- and 4-channel WDRC were greater with the short RT (panel b) than with the long RT (panel a). The lack of a RT x channels interaction for the amplified signal levels indicates that the influence of these two factors was independent.

Reduced temporal contrast is a byproduct of compression because the noise floor of the speech is increased relative to its peaks, which effectively decreases the modulation depth of the temporal envelope. In the examples shown, this is indicated by the elevated levels of the speech signal in the high-frequency band during the production of vowels and low- to mid-frequency consonants. The presence of a masker increases the effect and leads to concomitant changes in the long-term SNR. This effect on temporal contrast is most clearly illustrated in Figures S1 and S2 by comparing the levels of the peaks of speech, which correspond to the fricatives and stop consonants, to the levels of the interposing steady masker. On the high end of the continuum, Contrast was approximately 15-20 dB with 4 channels and the long RT (Figure S1a). When RT was made short (Figure S1b), contrast reduced by about 5 dB and when the number of channels increased to 16 with the long RT (Figure S2a), contrast reduced further, especially toward the end of the RT when masker levels were highest. On the low end of the continuum, contrast was only about 7 dB with 16 channels and the short RT (Figure S2b).

#### **SUPPLEMENT FIGURE CAPTIONS**

**Figure S1.** Following 4-channel WDRC with the long RT (a) and the short RT (b), the phase inversion technique was used to estimate the short-term, real-ear output levels for the 1/3-octave band centered at 8000 Hz for the sentence, "A tusk is used to make costly gifts" (solid black line) in the presence of a steady masker (dotted gray line) mixed at 0 dB input SNR. The prescriptive fitting parameters for a listener with a hearing loss representative of the group mean (Listener 13 in Table 1) were used for this example. (c) displays the difference in levels for the speech and masker when processed with the short and the long RT as well as the difference in short-term output SNR (dotted black line).

**Figure S2.** Same as in Figure S1 for 16-channel WDRC with a steady masker.

**Figure S3.** Same as in Figure S1 for 4-channel WDRC with a modulated, masker.

**Figure S4.** Same as in Figure S1 for 16-channel WDRC with a modulated masker.

**Figure S5.** The difference in levels for the speech (solid black line) and steady masker (dotted gray line) for 16-channel WDRC (Figure S2) and those for 4-channel WDRC (Figure S1) are plotted separately for the long RT (a) and the short RT (b). The difference in short-term output SNR (dotted black line) is also plotted.

**Figure S6.** Same as in Figure S5, except for with a modulated masker.

Ear and Hearing Supplement to Alexander and Masterson  
"Effects of WDRC release time and number of channels on output SNR and speech recognition"

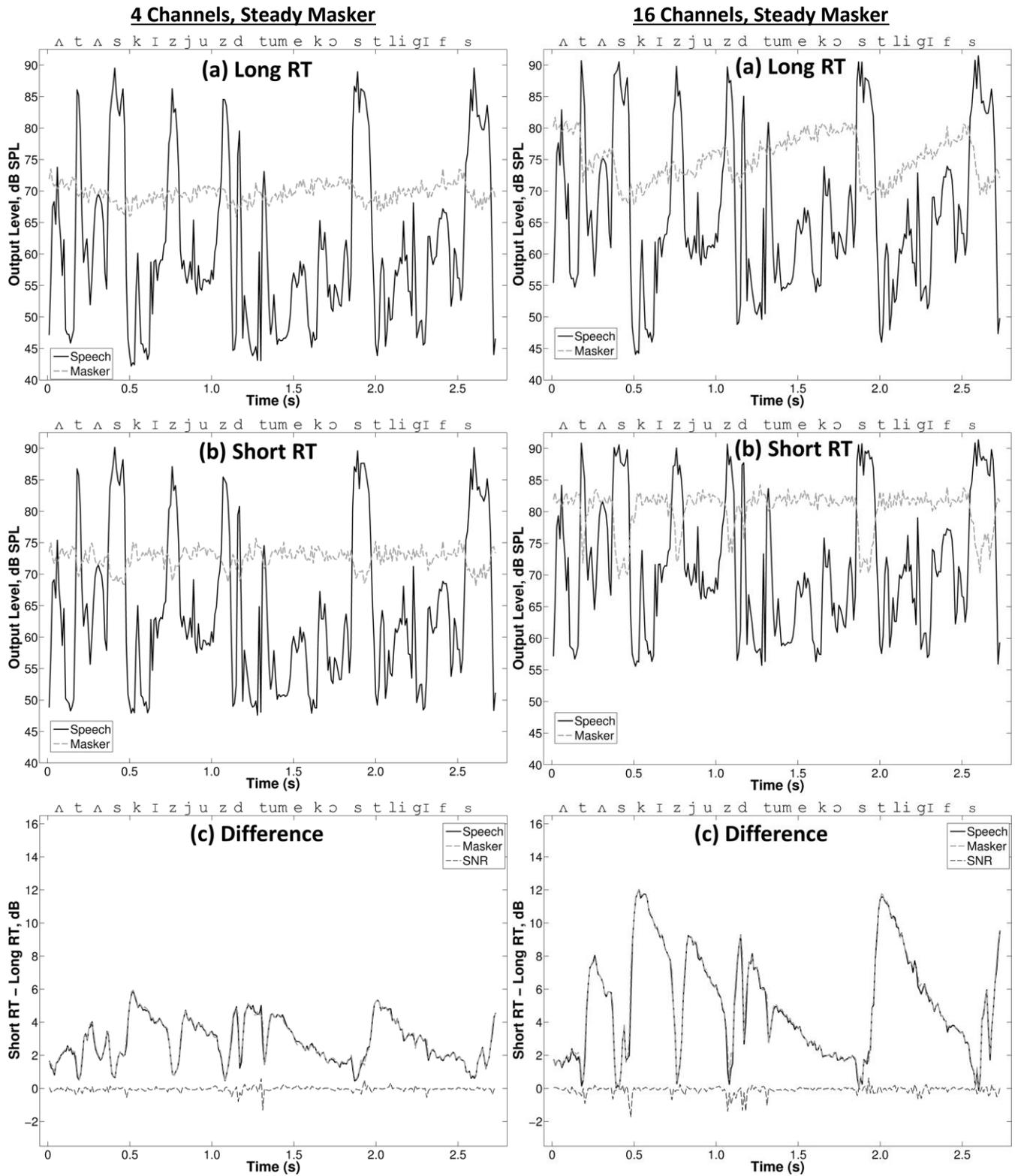
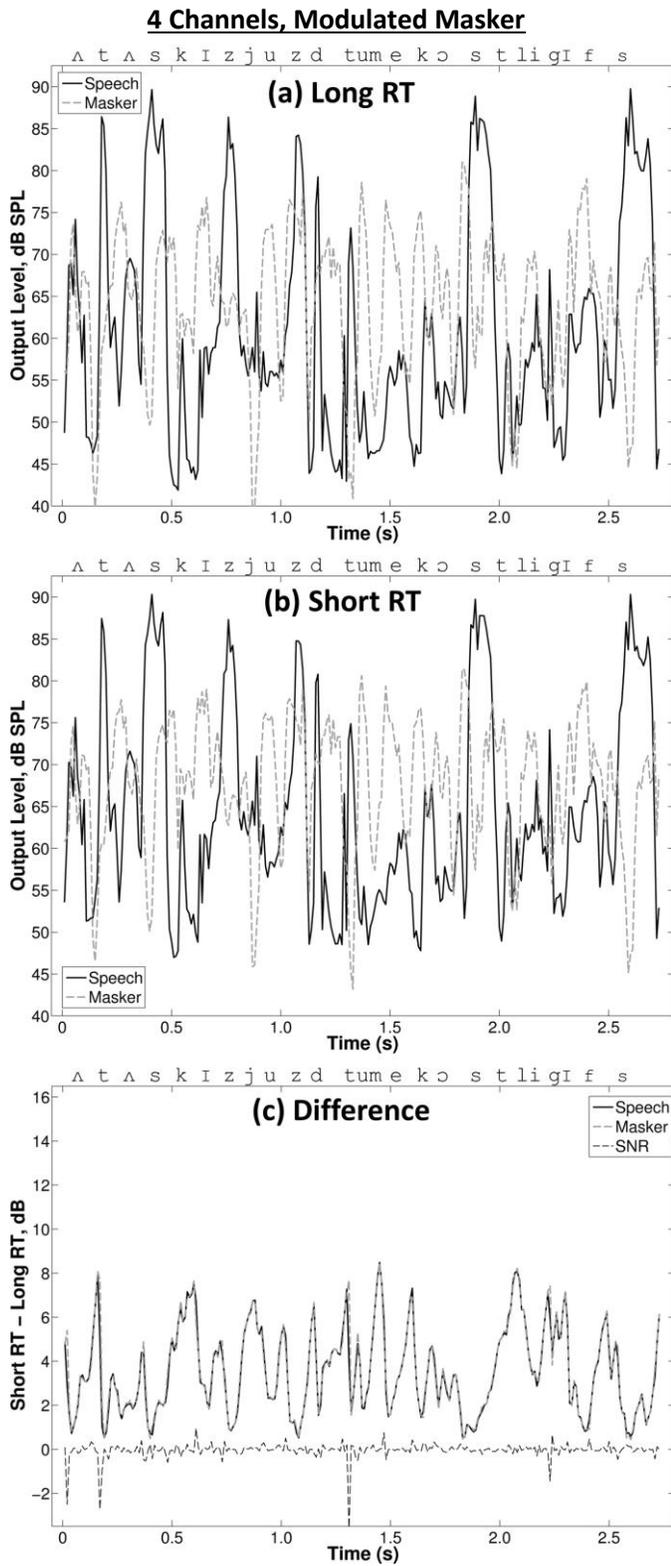
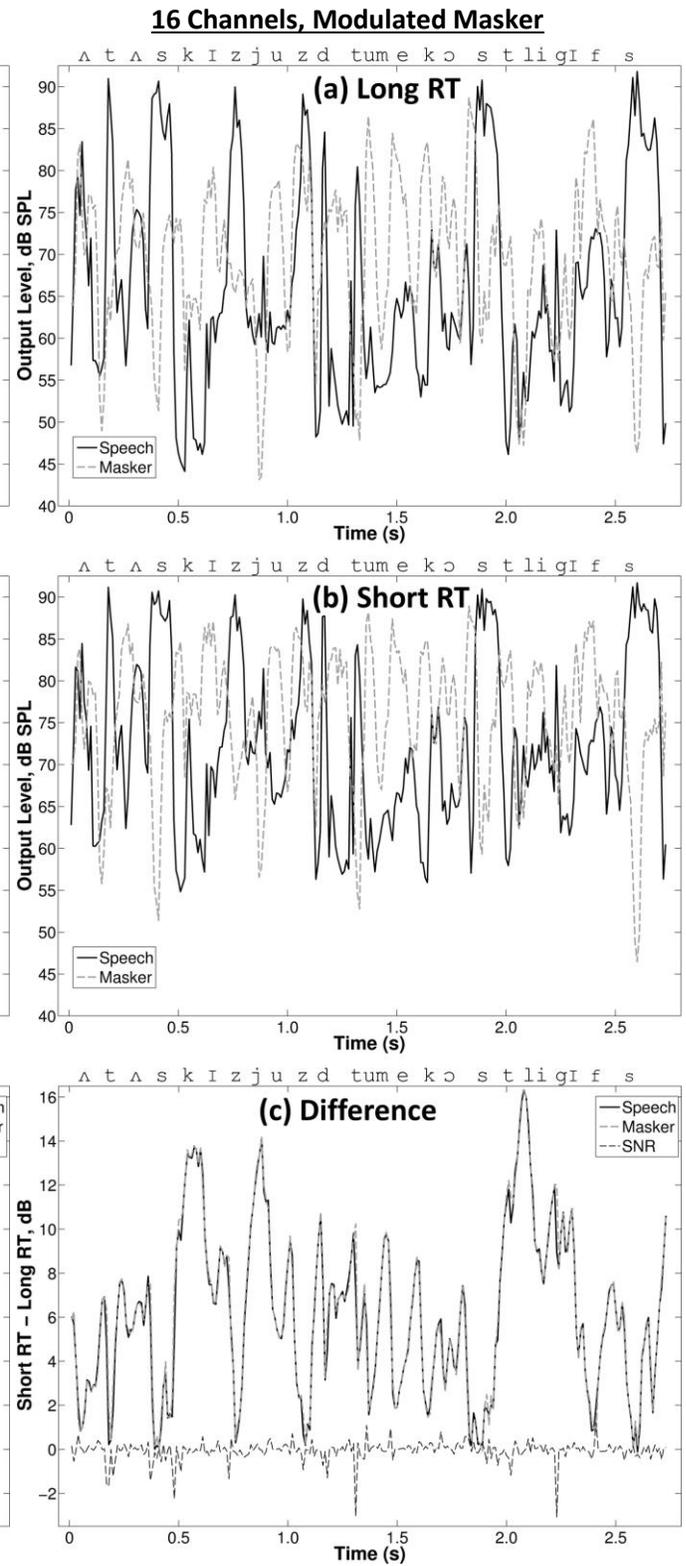


Figure 1

Figure 2

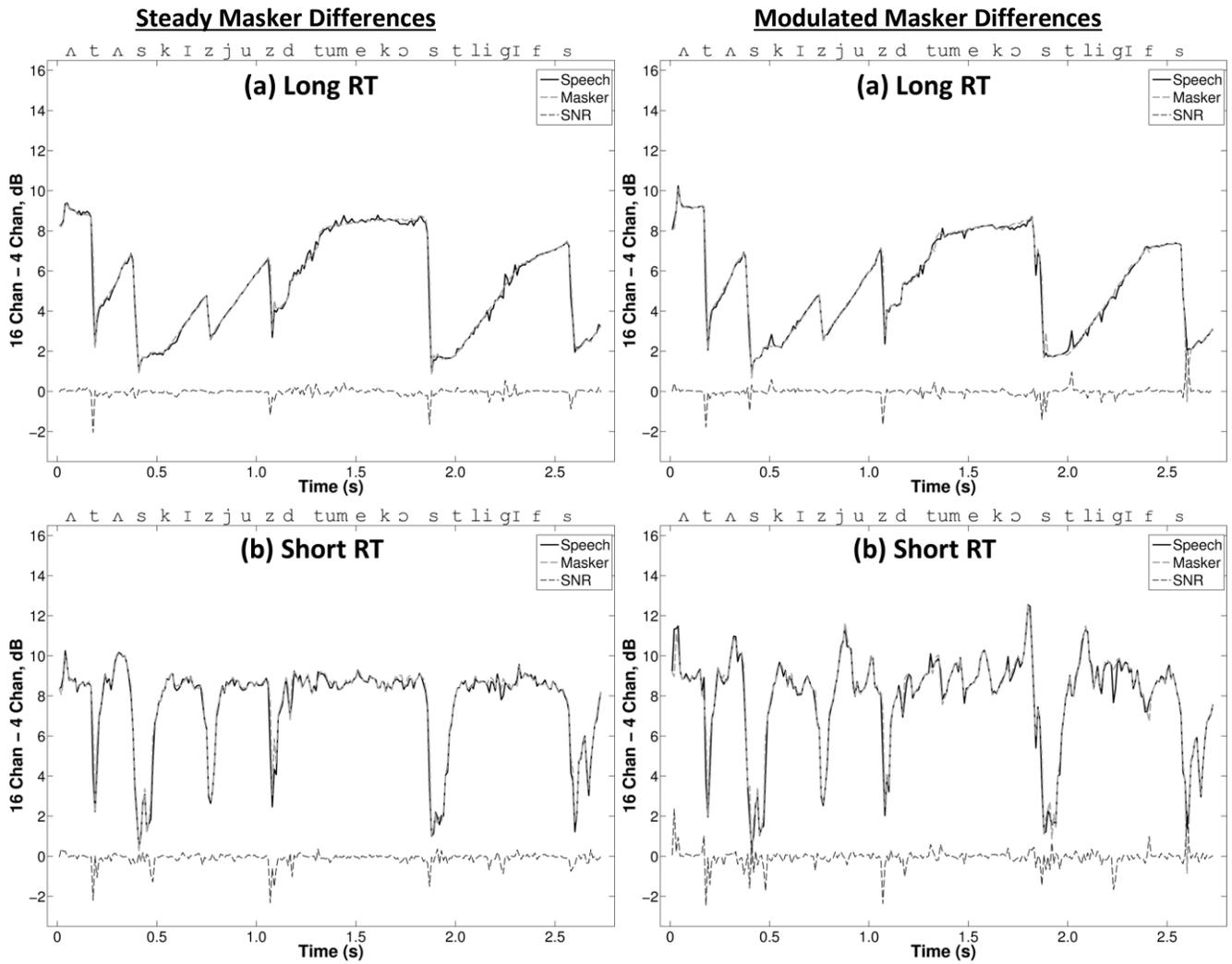


**Figure 3**



**Figure 4**

*Ear and Hearing Supplement to Alexander and Masterson  
"Effects of WDRC release time and number of channels on output SNR and speech recognition"*



**Figure 5**

**Figure 6**