This is a final version of an article published in final form in: Gustafson, Samantha[1,3]; McCreery, Ryan[2]; Hoover, Brenda[2]; Kopun, Judy G.[2]; Stelmachowicz, Pat[2] Listening Effort and Perceived Clarity for Normal-Hearing Children With the Use of Digital Noise Reduction, Ear and Hearing: March/April 2014 - Volume 35 - Issue 2 - p 183-194
doi: 10.1097/01.aud.0000440715.85844.b8

Listening effort and perceived clarity for normal hearing children with the use of digital noise reduction

Samantha Gustafson, AuD[1,3]

Ryan McCreery, PhD[2]

Brenda Hoover, MA[2]

Judy G Kopun, MA[2]

Pat Stelmachowicz, PhD[2]

1- Arizona State University

2- Boys Town National Research Hospital

3 – Vanderbilt University

Correspondence:
Samantha Gustafson
Vanderbilt Bill Wilkerson Center
Department of Hearing and Speech Sciences
1215 21st Ave. So.
Medical Center East, South Tower
Nashville TN, 37323-8242
(480) 540-8687 (Phone)
Samantha.Gustafson@Vanderbilt.edu

Research was supported by NIH grants RO1 DC4300 10, R03 DC009334, T35 DC8757, F31 DC010505, P30 DC4662.

Abstract

Objectives. The goal of this study was to evaluate how digital noise reduction (DNR) impacts listening effort and judgment of sound clarity in children. It was hypothesized that, when two DNR algorithms differing in signal-to-noise ratio (SNR) output are compared, the algorithm which provides the greatest improvement in overall output SNR will reduce listening effort and receive a better clarity rating from child listeners. A secondary goal was to evaluate the relation between the inversion method measurements and listening effort with DNR processing.

Design. Twenty-four normal hearing children (ages 7-12 years) participated in a speech recognition task in which consonant-vowel-consonant nonwords were presented in broadband background noise. Test stimuli were recorded through two hearing aids with DNR-off and DNR-on at 0 dB and +5 dB input SNR. Stimuli were presented to listeners and verbal response time (VRT) and phoneme recognition scores were measured. The underlying assumption was that an increase in VRT reflects an in increase in listening effort. Children rated the sound clarity for each condition.

The two commercially available HAs were chosen based on: 1) an inversion technique which was used to quantify the magnitude of change in SNR with the activation of DNR, and 2) a measure of magnitude-squared coherence which was used to ensure that DNR in both devices preserved the spectrum.

Results. One device provided a greater improvement in overall output SNR than the other. Both DNR algorithms resulted in minimal spectral distortion as measured using coherence. For both devices, VRT decreased for the DNR-on condition suggesting that listening effort decreased with

DNR in both devices. Clarity ratings were also better in the DNR-on condition for both devices. The device showing the greatest improvement in output SNR with DNR engaged improved phoneme recognition scores. The magnitude of this improved phoneme recognition was not accurately predicted with measurements of output SNR. Measured output SNR varied in the ability to predict other outcomes.

Conclusions. Overall, results suggest that DNR effectively reduces listening effort and improves subjective clarity ratings in children but that these improvements are not necessarily related to the output SNR improvements or preserved speech spectra provided by the DNR.

Introduction

Difficulty understanding speech in background noise is a common complaint among listeners with hearing loss (HL). In complex listening situations, adult listeners can utilize top-down processing; drawing upon their knowledge of lexical, semantic, and syntactic cues in conversational speech, to maintain speech intelligibility (Stelmachowicz et al. 2004; Jerger 2007). Reducing environmental noise can create a more favorable signal-to-noise ratio (SNR), which may improve the intelligibility of speech, particularly for children. When the level of environmental noise cannot be reduced at the input of the hearing aid (HA), digital noise reduction (DNR) algorithms have been used to limit the influence of noise within the output of the HA in an attempt to maintain speech intelligibility and improve listener comfort in noise. While previous studies have demonstrated that DNR does not improve or degrade speech recognition (Stelmachowicz et al. 2010) or word learning (Pittman 2011a) in children under 10 years old, the influence of DNR on listening effort and sound quality in children remains unresolved. It is important that the effects of this signal processing be understood for this population because children are not often able to adjust their environment for optimal listening, especially in classroom settings.

Speech Intelligibility

Research with adults has demonstrated either no improvement (Boymans and Dreschler 2000; Alcantara et al. 2003; Ricketts and Hornsby 2005) or degradation (Jamieson et al. 1995; Kates 2008) in speech intelligibility with DNR compared to conditions without DNR. A study by Stelmachowicz et al. (2010) examined the effect of DNR on the ability of children with HL (ages 5-10 years) to perceive nonsense syllables, monosyllabic words, and sentences at 0, +5, and +10 dB input SNRs. Results suggested that the DNR algorithm used neither improved nor degraded speech perception in children at any input SNR, confirming previous results obtained with adults

(Ricketts and Hornsby 2005; Mueller et al. 2006). Pittman (2011b) used word categorization in the presence of auditory and visual competitors to examine performance of children with normal hearing (NH) and children with HL (ages 8-12 years) in a difficult listening environment (0 dB SNR). Consistent with results reported by Stelmachowicz et al. (2010) and those from previous adult studies, Pittman found the children's performance was unaffected by DNR.

Subjective Ratings of Sound Quality with DNR

While DNR has not been shown to consistently improve or reduce speech recognition in previous studies, several studies have reported improved listening comfort (Jamieson et al. 1995; Boymans and Dreschler 2000; Mueller et al. 2006) and/or sound quality (Ricketts and Hornsby 2005; Hu and Loizou 2007) in noise. Subjective reports from a multicenter study show that adults with NH and with HL report that listening to speech processed with various DNR algorithms is less effortful as the SNR improves (Luts et al. 2010).

Listening Effort

A listener's preference for DNR may be related to a decrease in listening effort caused by the improvement in overall SNR. Listening effort refers to the cognitive resources required to understand speech in a given acoustic environment. Given the model of limited cognitive capacity (Kahneman 1973), listeners of a degraded speech signal require more cognitive resources and experience increased demands on top-down processing in an attempt to maintain optimal performance (Pichora-Fuller et al. 1995). This degradation in the speech signal may be due to a HL, to background noise, or to a combination of the two factors. Listeners with NH and with HL experience increased listening effort in the presence of noise (Larsby et al. 2005). However, having a HL increases the difficulty of listening in background noise, requiring more effort from the listener with HL when compared to listeners with NH (Rakerd et al. 1996; Hicks and Tharpe 2002; Larsby et al. 2005; Zekveld et al. 2011). The implications of this increase in effort for children with HL are likely to be substantial considering the less than optimal noise present in typical classrooms (Arnold and Canning 1999; Crandell and Smaldino 1995, 2000).

Objective Measures of Listening Effort. Verbal response time (VRT) has been used in previous studies of cognitive processing to assess language comprehension (e.g. Stanovich and West 1983) and memory (e.g., Cowan et al. 2003). Without requiring the listener to consciously allocate cognitive resources, cognitive processing is captured in the time delay before a response. Cowan et al. (2003) proposed that increases in VRT in nonword recognition tasks reflect an increase in the amount of time that children need to decode the signal, with longer response time reflecting greater processing effort. Additionally, using VRT as a measure of listening effort may better reflect the increased cognitive load encountered when listening conditions become increasingly difficult. In a study examining the effect of amplification on both word and sentence recognition and VRT in adult listeners with HL, Gatehouse and Gordon (1990) concluded that VRT revealed changes in conditions where speech recognition tasks are subject to learning and ceiling effects. Although the mechanism underlying VRTs in individuals with HL is likely related to the effects of significant HL, the introduction of distortion via signal processing or the addition of background noise would also be expected to increase response time. Hallgren et al. (2001) suggested that processing of information is slower in background noise when compared to quiet conditions, reflecting the need for greater cognitive processing for more challenging tasks. Other studies have noted similar findings of increased listening effort as measured by response time as listening conditions become more difficult (Mackersie et al. 1999; Hallgren et al. 2005; Larsby et al. 2005; Pisoni et al. 2010).

With improvements in SNR showing reduced response times, an additional reduction in VRT should be expected if DNR processing algorithms can further improve the overall SNR. This may have been the case in a study by Sarampalis et al. (2009) who examined the effect of a simulated DNR algorithm on listening effort in young adults with NH at a variety of SNRs. Results revealed that this DNR processing reduced listening effort in the condition with the poorest SNR. However, similar to previous studies of DNR, the processed signal was not analyzed acoustically, so the amount of reduction in level with DNR and how that relates to the finding of improved effort cannot be determined.

Quantifying Signal Changes

Because the effect of the DNR circuit on the output signal varies widely across HA manufacturers (Bentler and Chiou 2006; Hu and Loizou 2007), accurately predicting the effect of DNR on overall changes in SNR requires acoustic analysis of the hearing aid output. Hoetink and colleagues (2009) examined gain reduction in twelve HAs with DNR processing, obtaining data over a range of presentation levels, SNRs, and HA characteristics for different configurations of HL. Their results showed large variations in gain reduction across frequencies and devices and for each variable (i.e., input signal level, input SNR, and HL configuration). Specifically, different DNR algorithms provided varying amounts of reduction in gain across frequencies for the same configuration of HL and the same noise. Additionally, one DNR algorithm provided an overall reduction of gain, while another reduced gain only within a narrow frequency range corresponding to the frequency of the noise. Even for stimuli with the same SNR prior to amplification (input SNR), different signal classification systems and DNR algorithms have varying effects on the SNR after processing. Such differences in the effect of different DNR algorithms create challenges in predicting how gain reduction will impact the wearer's performance with DNR.

Quantifying the affect of DNR on the relative levels of speech and noise signals presents multiple challenges. First, the input signal to the hearing aid is often composed of both speech and noise, which often have similar spectra. In this environment, modulation-based DNR algorithms rely on differences in modulation between speech and noise to accurately distinguish between the two and to reduce the level of the noise while maintaining the spectrum of the speech signal (Bentler and Chiou 2006; Kates 2008). To change the instantaneous SNR of a signal, a DNR algorithm would need to extract both amplitude and phase details of the noise from the combined speech and noise signal. The processing of modern DNR algorithms is not fast or accurate enough to satisfy these requirements so gain reductions from DNR are applied to both speech and noise. This limits the likelihood that DNR can selectively reduce background noise in a way that would be expected to improve speech recognition in noise (Anzalone et al. 2006). However, DNR algorithms do have the potential to affect the SNR of a signal over a longer period of time in ways that may be quantified through acoustic analysis. For example, DNR may reduce the noise level during pauses in speech. Therefore, if the overall level of the speech is maintained and DNR reduces the noise level during pauses in speech, an overall improvement of SNR may occur.

Inversion Method. First described by Hagerman and Olofson (2004), one method proposed to quantify SNR changes due to signal processing is the inversion method. This method was used by Souza et al. (2006) to determine the effect of compression amplification on the SNR for speech in noise. The inversion method uses a noise signal with an inverted phase which, when combined with a signal comprised of speech and noise, creates separate waveforms of processed speech and noise. These waveforms are processed by the HA and compared to the original signal. An increase in the output SNR with DNR would indicate that the processing maintains the intensity level of the speech while decreasing the relative level of the noise. Because improvements in the SNR of a speech signal have been shown to improve ease of listening (Mackersie et al. 1999), this same improvement may be expected for HA users in noisy environments if the overall SNR of the signal is improved with DNR processing. Furthermore, the reduction in overall gain following the detection of noise provided by DNR indicates that there is a potential for DNR to reduce the level of the speech signal along with the noise, especially if the spectral characteristics of the speech and noise are similar. If DNR processing degrades the speech signal by altering the envelope or does not significantly reduce the level of the noise relative to the speech signal, it is unlikely that ease of listening will be improved. However, DNR has been shown to improve a listener's acceptable noise level (Mueller et al. 2006; Wu and Stangl 2013), which may also influence ease of listening.

The inversion method has been to quantify the SNR change with DNR processing (Pittman 2011a,b; Wu and Stangl 2013). Pittman used a dual-task paradigm with thirty children with HL (ages 8-12 years) to examine the effects of DNR on complex listening performance in demanding listening conditions. Results showed no effect of DNR with a 2 dB SNR improvement. Wu and Stangl measured an improvement in acceptable noise levels in adult listeners with HL that corresponded to the change in output SNR measured by the inversion method. Discrepancies in predictive performance of the inversion method to derive SNR improvement could be related to the fact that the inversion method assumes a linear, time-invariant system. With few studies reporting on this issue, the relation between changes in the output SNR measured by the inversion method and changes in performance with DNR remains unclear.

Magnitude Squared Coherence. While the inversion method can provide information relative to the entire frequency spectrum of a signal, processing changes using current HA technology occur in narrower frequency bands. Another method that can be used to quantify the changes caused by signal processing is magnitude squared coherence which has been used to quantify internal HA noise (Lewis et al. 2010) and the impact of HA signal processing on signal integrity, including amplitude compression (Kates and Arehart 2005) and DNR (Ma and Loizou

2011). Coherence represents the comparison of the signal power at the input of the HA to the total power at the output of the HA as a function of frequency. Coherence varies between 0, representing no relation between signal input and output, and 1, indicating that the output power spectrum is identical to the input signal. With this method, the spectrum of the input signal can be compared to the spectrum of the HA output signal to determine the magnitude of distortion as a function of frequency. Changes in coherence of the combined speech and noise between conditions with and without DNR would reflect distortion introduced by the DNR processing. Furthermore, coherence measures represent relative differences in spectral distortion between the DNR on and DNR off and may not reflect changes in the amplitude of the signal that occur across DNR conditions (Kates and Aerhart 2005).

The primary goal of this study was to evaluate how the DNR algorithms from two different manufacturers impact listening effort and subjective judgments of sound clarity in children with NH. Because the SNR improvement measured by the inversion method has not fully been evaluated perceptually, a secondary goal of this study was to characterize the relationship between SNR improvement with DNR processing measured by the inversion method and the impact of this relationship on listening effort. It was hypothesized that, when the two DNR algorithms are compared, the algorithm which provides the largest improvement in overall output SNR as measured by the inversion method will show a greater reduction in listening effort and result in a better clarity ratings from child listeners.

Materials and Methods

Participants

Twenty-four children (ages 7-12 years) with NH were recruited from the Boys Town National Research Hospital Human Research Subject's Core database. Criterion for NH was thresholds ≤15 dB HL at octave frequencies from 0.25 through 8 kHz. The test ear chosen alternated between right and left ears across children. All children passed the Bankson-Bernthal Quick Screen of Phonology (Bankson and Bernthal 1990) within six months prior to participation. This test was used to confirm that the children had no significant speech production errors that would influence scoring. All participants spoke English as their native language. Children were paid $15 per hour for participation. Testing required no more than a single, two-hour session. Informed consent and assent were obtained for all participants according to the procedures required by the Institutional Review Board at Boys Town National Research Hospital.

Stimuli Development

Test stimuli. Test stimuli were 225 consonant-vowel-consonant (CVC) nonwords. Nonwords were chosen to represent the open-set possibilities of a novel language such as might be experienced by a child learning English and to reduce the effects of vocabulary knowledge. These stimuli were used in a previous study of speech recognition with children (McCreery and Stelmachowicz 2011). All CVCs were spoken by a female talker and digitally recorded in a sound-treated booth using a condenser microphone with a flat response (±2 dB) from 0.2 to 20 kHz. Speech tokens were amplified and sampled at a rate of 22,050 Hz with a quantization of 16 bits. Stimulus files were scaled to 65 dB SPL at a calibrated position in the sound booth. Of the nine lists containing 25 CVCs, each had an average phonotactic probability biphone sum of 0.005 based on an online calculator for children (Storkel and Hoover 2010). Each list also had at least one instance of each consonant (initial or final position), and an overall similar number of occurrences for all manners of articulation. See Appendix I for word lists. Samples of unmodulated white noise with a flat spectrum from 0 – 11,025 Hz scaled to both 60 and 65 dB

SPL at the calibrated position were used to create two conditions of +5 and 0 dB input SNR, respectively. SNRs of typical classrooms have been reported to range from -7 to +5 dB (Arnold and Canning, 1999; Crandell and Smaldino 1995, 2000). In the current study, SNRs of 0 and +5 dB were selected to represent realistic listening conditions. White noise was used in this study because previous studies have shown that DNR may not be activated when speech-shaped noise is used due to the spectral similarity between signal and noise (Alcantara et al. 2003; Natarajan et al. 2005; Bentler et al. 2008).

Stimulus recording procedures. Using a custom computer program, the white noise and CVCs were routed simultaneously through two channels of a Lynx TWO B sound card. Stimuli were presented via a single loudspeaker placed at zero-degree azimuth in the sound field of a sound-treated room. The output of the HA was recorded through an IEC 711 coupler in a G.R.A.S. Knowles Electronics Manikin for Acoustic Research (KEMAR). The HA was coupled to the left ear of the manikin using a G.R.A.S. KB0110 canal extension adaptor. A 30-sec white noise precursor was included at the beginning of each list to properly engage the DNR algorithm. Following this noise, two presentations of the 17.5-second concatenated word lists were played. Because the onset time of DNR can vary for noise-only and speech-in-noise signals, the recording of the 30-seconds of noise and the first list were discarded prior to saving the sample, leaving one full set of CVCs per list for each condition. These steps were taken to ensure that the magnitude of DNR was constant for all experimental stimuli. Each list was recorded in the eight experimental conditions and stimuli were RMS equated, for a total of 72 lists.

Hearing aid selection. Five commercially-available behind-the-ear hearing instruments from different manufacturers were considered for the study. The HAs were programmed to match DSL v5.0 (Scollie et al. 2005) quiet targets (+/- 2 dB) at 55, 65, and 75 dB SPL for a 10 year-old female with a flat, 50 dB HL using the Audioscan Verifit (Audioscan, Ontario, Canada) and simulated real-ear measures. Microphones in all HAs were set to the omni-directional mode. All additional processing features available for adjustment in the programming software were deactivated. A copy of Program 1 was used to create the second experimental program, however, DNR was set to the maximum noise reduction setting available for each device.

The inversion method (Hagerman and Olofson 2004) was used to quantify changes in the SNR between DNR-on and DNR-off settings. For this method, three measurements were made in a $2cm^3$ coupler: (1) speech and noise in its original form (S_oN_o), (2) the same speech signal - phase inverted with the noise in its original form (S_iN_o), and (3) the same noise sample - phase inverted with the speech sample in its original form (S_oN_i). Each of these digital wave files was processed through the HAs at the desired SNRs (both with and without DNR engaged). The recorded output files were then digitally mixed to produce a speech- and noise-only waveform. For example, to create the speech-only waveform, the S_oN_o waveform was mixed with the S_oN_i waveform. Using Adobe Audition, the average rms voltages (dB) of the speech- and noise-only digital waveforms were compared to obtain an output SNR. The speech files used in this process were concatenated test stimuli (nonwords) with 10 ms between each token and continuous white noise. Each recording was 65 seconds in length, comprised of 30 seconds of noise followed by two repetitions of the 17.5 second concatenated nonword list. To ensure that activation times (which vary across manufacturer) did not impact the experimental recordings, the preceding noise and first repetition of each list were discarded, leaving one processed list of nonwords to be used for all analyses. This also helped ensure that the DNR was fully engaged during the recording of the experimental lists.

Using the inversion method, the two hearing aids representing the least and greatest SNR improvement with DNR engaged were chosen for this study: the Oticon Agil Pro (Oticon, Smørum, Denmark) and Phonak Naida V SP (Phonak AG, Zurich, Switzerland). Hereafter, these devices will be referred to as HA1 and HA2, respectively. In each system, DNR technology utilized the manufacturers' proprietary method of amplitude modulation detection to determine the appropriate gain adjustments needed to minimize the effects of noise. These adjustments were made in individual frequency bands for HA2 and in independent channels for HA1. HA1 also incorporated a synchrony detection algorithm to identify speech in the poorest SNRs. Synchrony detection monitors the spectral envelope of the input signal. When a fixed level of envelope-synchrony is detected (suggesting that speech is present), the DNR is reduced, especially in the 1-4 kHz range [see Chung 2004 for more detailed review of synchrony detection]. Both manufacturers call for an attenuation amount that is inversely proportional to the SNR estimated in each band. For both manufacturers, precise time constants of advanced signal processing are proprietary, therefore, activation times for both devices were measured using the experimental noise to determine the time at which the DNR began and when maximum gain reduction was reached. On average, HA1 reached complete activation at 20.6 seconds while DNR in HA2 was fully active after 13.7 seconds. A recording window of 130 seconds was examined to confirm that neither device provided additional gain reduction outside of the 65 second experimental recording period. The inversion method calculations were completed with stimuli processed at the two input SNRs (0 and +5 dB) by both hearing aids with DNR on and DNR off.

Processed Experimental Stimuli

Programming in both devices was identical to settings used in the device selection described above. Features that were deactivated for both experimental programs included: My Voice in

HA1, and Bass Boost, Occlusion Control, Whistle Block, and Wind Block in HA2. Specific features active in Program 2 included TriState Noise Management in HA1 and Comfort in Noise with Noise Block set to "strong" in HA2.

A measure of magnitude squared coherence (Lewis et al. 2010) was used to compare the spectral distortion of stimuli at the input and output of the HA for the DNR-on and DNR-off conditions. This comparison was made in attempts to quantify any spectral distortion resulting from DNR processing or from the inversion method. The magnitude squared coherence of the wavefiles for each device and SNR condition were calculated in MATLAB, using a rectangular window with a length of 1024 ms.

Experimental Procedures

Recorded speech tokens were presented to listeners monaurally via Sennheiser HD-25 earphones at 65 dB SPL. Each subject listened to an unprocessed list for familiarization with the procedure followed by the eight experimental conditions. Variables in each condition included: device (HA1 and HA2), input SNR (0 and +5) and DNR setting (On and Off) for a total of eight conditions with differing output SNRs. The presentation order of CVC nonwords within each list and the presentation order of each condition was randomized for each child. Additionally, the nine lists were randomly assigned across each of the eight experimental conditions and an initial practice condition.

Children were instructed to repeat each CVC as it was presented. An examiner in the control room initiated all stimulus presentations and a customized computer program produced an audio-only recording of the children's responses. To help maintain interest in the task, visual reinforcement was given immediately after each response and was not contingent upon correct responses. Responses were recorded using a Shure Beta 53 omnidirectional head-worn microphone positioned approximately 2 cm from the corner of the subject's mouth. At the conclusion of each condition, children were shown six photos of a tiger that were progressively degraded in terms of visual clarity. A rating of one corresponded to the clearest photo and a rating of six represented the most degraded photo. Children were asked to use numbers below the six photos to indicate the clarity of the nonwords. Two independent judges phonemically scored children's responses to CVC stimuli following data acquisition. Phoneme recognition scores from the two judges were compared for agreement, and a third judge evaluated responses if the original judges did not agree on the response. Cohen's Kappa (K) for phoneme scoring was 0.9877, reflecting excellent agreement between judges. A custom computer program was developed to measure VRT using the zero crossings from the onset of the stimulus to the onset of the subject's response. This measurement was confirmed by one of the two judges and verified for accuracy with a second judge for the first three participants. Due to high inter-rater reliability (Cohen's K = .9253) for the VRT judgments of the first three children, a single judge was used to verify VRT for the remaining data.

Results

Figure 1 displays the long-term average spectra in relative dB for speech and noise stimuli separated using the inversion method for both devices and DNR settings in the +5 dB input SNR condition. The spectra in Figure 1 are presented to demonstrate the relative effects of DNR on the individual speech and noise spectra from the inversion method and do not reflect the stimuli as presented to the listeners in the current study. Although the frequency responses of these devices were matched using a calibrated speech stimulus during programming, a discrepancy in the responses obtained with experimental stimuli is evident. The variability between processing and device is consistent with the results of Hu and Loizou (2007) that showed differences both between and within manufacturers depending upon stimulus parameters.

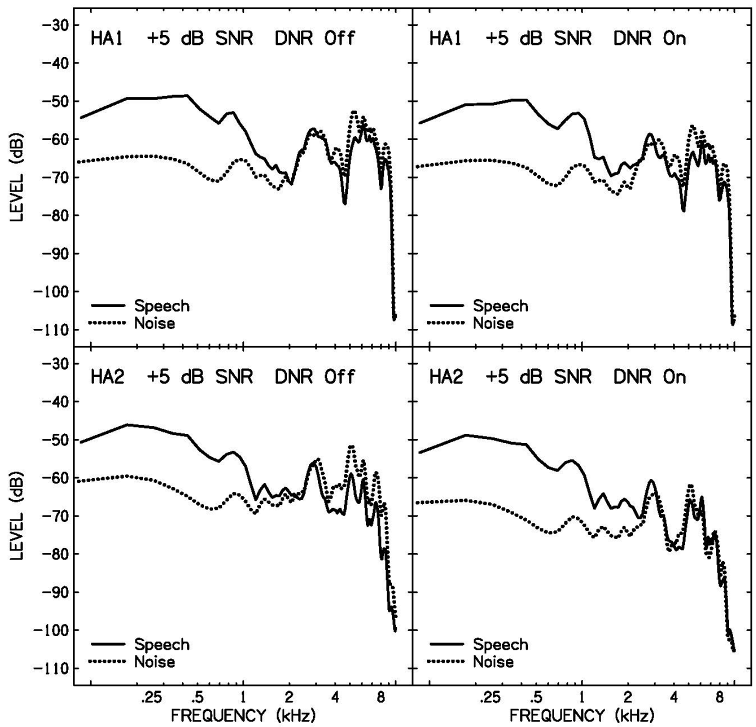

Figure 1. Long-term average spectra for both devices in relative dB using speech and noise stimuli separated using the inversion method at +5 dB input signal-to-noise ratio (SNR) in both digital noise reduction (DNR) –off and –on conditions.

Inversion Method

Table 1 shows the output SNR, as derived with the inversion method, for each experimental HA in each of the two input SNR conditions. The +5 dB SNR calculations in this table represent the stimuli depicted in Figure 1. Results of these measurements suggest that neither of the hearing aids faithfully reproduced the input SNR whether DNR was activated or not activated. In fact, the inversion method suggests that the HA processing alone decreased the SNR of the signal in both input conditions for both devices (e.g. in the DNR off condition, measured output SNR for both devices was lower that the input SNR). This pattern is consistent with previous research showing the detrimental effects of wide dynamic range compression processing on HA output SNR (Naylor and Johannesson 2009). When DNR was activated, the two devices yielded different amounts of measured change in SNR. While both HAs showed a similar output SNR with DNR-off in both 0 and +5 dB input SNR conditions, HA2 appears to have provided a measured output SNR that was higher (better) with DNR activated, compared to HA1. These specific differences between the two experimental devices for the DNR-on condition can be seen as relative level differences between speech and noise above 1000 Hz in Figure 1.

To demonstrate the repeatability of the inversion method for stimuli used in this study with a controlled SNR relationship, the change in SNR between five different stimulus lists was calculated using the inversion method and compared at 0 dB and +5 dB SNR without hearing aid processing. The inversion method calculation comparing 0 dB and 5 dB SNRs was 4.96 dB on average with a range of 4.7 – 5.2 dB.

Table 1. Output signal-to-noise (SNR) measurements of experimental hearing aids (HA) using the Inversion Method.

0 dB	HA1	HA2
	Output SNR	Output SNR
DNR-off	-4.41	-3.04
DNR-on	-0.62	3.96
SNR Difference	3.79	7.00
+5 dB		
DNR-off	-0.07	0.29
DNR-on	1.39	6.97
SNR Difference	1.46	6.68

Magnitude Squared Coherence

Figure 2 shows coherence for each device as a function of frequency for the speech in noise files used for the speech recognition task for both input SNR conditions. A value of 1 on the abscissa indicates that the spectra for the DNR-on and DNR-off signals are identical and that no portion of the spectra was altered by DNR processing. Note that spectral coherence is high for both devices. The reduction in coherence for HA2 across the entire frequency range corresponds well with the differences from the inversion method. Namely, HA2 showed a greater change in spectral power difference with DNR processing and a measured improvement of 6.68-7.00 dB SNR with the inversion method, whereas HA1 showed less change in coherence and only 1.46-3.79 dB of measured SNR improvement using the inversion method.

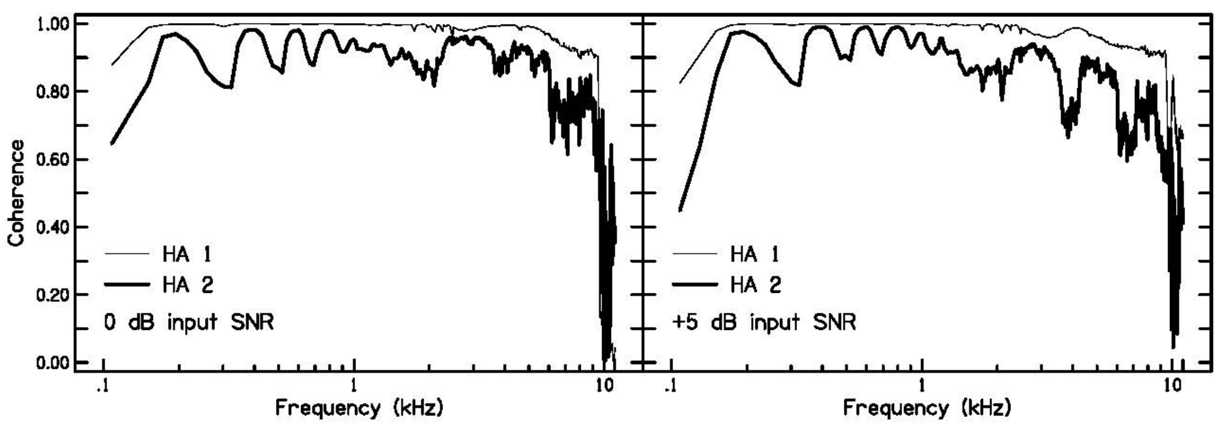

Figure 2. Magnitude squared coherence for each device as a function of frequency for the speech in noise recordings for the two input SNR conditions as measured using the inversion method.

Preliminary Analysis

Prior to analysis, unintelligible recordings and missing data were excluded. Outliers (i.e.,±3 SD) from each subject's VRT mean Z-score (explained below) in each condition were also excluded as in previous studies of response time in children (Ratliff 1993). This process resulted in elimination of 263 of 2000 total possible response times for all subjects and conditions (13.2%). A Kolmogorov-Smirnov test of normality comparing the raw VRT data distribution to a normal distribution was significant ($p < 0.001$) suggesting that the raw VRT data were not normally distributed. Additionally, results from previous studies have cautioned the analysis of raw response time data due to the variability of shift, scale, and shape of the distributions across subjects (Ratliff 1979; Heathcote, Popiel and Mewhort, 1991; Rouder et al. 2003; Hervey, Epstein, Curry, et al. 2006; Whelan 2008). For example, some children were able to respond consistently fast in all conditions, whereas others showed inconsistent response times within conditions regardless of condition difficulty. Using central tendency (mean and SD) to compare these two types of responders ignores the variable distribution of data and may obscure significant findings (Whelan 2008). To account for this variability, averaged raw scores in each experimental condition for each subject were converted to Z-scores using a method similar to Slocomb and Spencer (2009)[1]. This conversion allowed for the comparison of relative changes in VRT across children in each condition and for the explanation of overall VRT effects due to experimental variables, accounting for the individual children's variability. The mean and SD of children's transformed VRT in each condition are shown in Figure 3. These Z-scores show that, when using DNR, children listening with HA2 in the 0 dB SNR condition showed a VRT that was, on average, 0.029 standard deviations shorter than their average response time whereas they

---

[1] While Z-scores in Slocumb and Spencer (2009) were calculated for individual tokens to control for within-item variability, calculations in the present study were made using individual subject mean scores to control for within-subjects variability. For instance, the difference between mean VRT using HA1 with DNR on in +5 dB SNR and mean VRT in all conditions for one child divided by the standard deviation of VRT in all conditions for that same child resulted in the child's Z-score for the HA1 with DNR on in +5 dB SNR condition.

showed a VRT that was 0.233 standard deviations longer than their average response time without the use of DNR. Phoneme recognition data were converted to rationalized arcsine units for purposes of statistical analysis to normalize the error variance of proportional scores (Studebaker 1985).

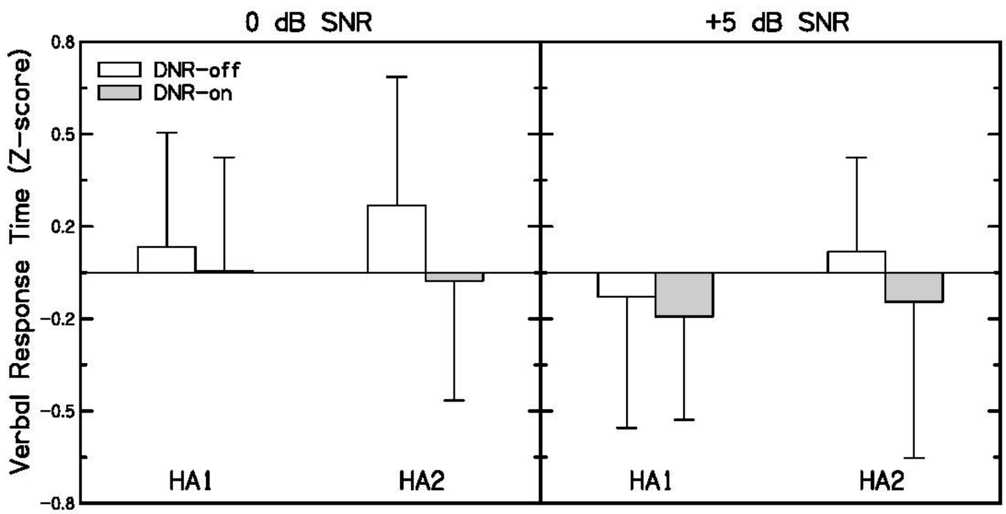

Figure 3. Transformed verbal response time (VRT) Z-scores for each device is plotted in the DNR-off and DNR-on conditions at 0 dB (left panel) and +5 dB (right panel) input SNR.

Correlations between outcome variables were examined to determine the direction and extent of relationships between transformed VRT Z-scores, clarity ratings, and phoneme recognition. The correlations between the three outcome variables were calculated based on the mean data for each subject in every condition. Significant correlations ($p < 0.05$) were observed for the same outcome across conditions. However, no significant correlations were observed between VRT, clarity, and phoneme recognition for the same listening condition. The overall correlations between VRT and clarity ($r = 0.090$, $p = .675$), VRT and phoneme recognition ($r = -.120$, $p = .575$) and clarity and phoneme recognition ($r = .044$, $p = .839$) were not significant, suggesting that the relationship between outcome variables did not follow a consistent pattern when collapsed across conditions.

To analyze the combined multivariate effects of SNR, HA, and DNR on the outcomes of transformed VRT Z-scores, clarity ratings, and phoneme recognition, a repeated-measures multivariate analysis of variance (MANOVA) was completed. The multivariate two way-interaction between HA and DNR was significant [Wilks' $\lambda = 0.406$, $F(3,21) = 10.26$, $p < .001$, $\eta^2_p=0.594$], suggesting that the pattern of results for the outcome variables was significantly different for each HA as a function of DNR. The multivariate effects of SNR [Wilks' $\lambda = 0.056$, $F(3,21) = 118.281$, $p < .001$, $\eta^2_p=0.944$] and DNR [Wilks' $\lambda = 0.459$, $F(3,21) = 8.26$, $p = .001$, $\eta^2_p=0.541$] were also significant, but the effect of DNR should only be interpreted in the context of the significant multivariate two-way interaction between HA and DNR. The multivariate effect of HA and all other two- and three-way interactions were not significant. The univariate effects for each outcome variable from the MANOVA were then examined to determine how each of the experimental factors influenced each outcome, while controlling for the other outcomes.

Verbal Response Time (VRT)

Analysis of the univariate effects from the MANOVA of DNR on transformed VRT Z-scores indicated that VRT varied significantly as a function of DNR [$F\ (1,23) = 5.105$, $p=0.034$, $\eta^2_p=0.182$]. Specifically, VRT was significantly faster for DNR on compared to DNR off. The univariate effect of SNR [$F\ (1,23) = 7.026$, $p=0.014$, $\eta^2_p=0.234$] was significant due to faster response times at 5 dB SNR than 0 dB SNR. The univariate effect of HA on VRT [$F\ (1,23) = 1.731$, $p=0.201$, $\eta^2_p=0.070$] was not significant, indicating that VRT did not vary across devices. Higher-order interactions involving VRT (HA x DNR [$F\ (1,23) = 1.23$, $p=0.279$, $\eta^2_p=0.051$], HA x SNR [F (1,23) = 0.091, p=0.765, η^2_p=0.004], DNR x SNR [F (1,23) = 0.268, p=0.609, η^2_p=0.012], three-way interaction between HA, DNR, and SNR [F (1,23) =0 .058, p=0.812, η^2_p=0.003]), were not significant.

The MANOVA and related analysis of univariate effects of HA, SNR, and DNR on VRT using the untransformed VRT (in ms) showed the same pattern of significant multivariate effects and interactions compared to the above reported analysis conducted using transformed VRT Z-scores with the exception of the univariate effect of SNR on VRT. Table 2 shows the raw VRT mean and SD for both devices in each condition. The untransformed VRT followed the same pattern as the Z-score transformed VRT with faster response times on average for 5 dB (1316 ms) than for 0 dB SNR (1353 ms) when collapsed across device and DNR condition. However, the univariate effect of SNR on untransformed VRT did not reach significance [F (1,23) = 3.471, p=0.075, η^2_p=0.131], likely due to the increased variability across subjects in the untransformed VRT data.

Table 2. Raw verbal response times (VRT) in milliseconds (*1SD*) for both devices in each digital noise reduction (DNR) and input signal-to-noise ratio (SNR) condition.

	DNR-off		DNR-on	
	Input SNR		Input SNR	
	0 dB	+5 dB	0 dB	+5 dB
HA1	1365 *(511)*	1308 *(414)*	1336 *(501)*	1277 *(350)*
HA2	1383 *(427)*	1360 *(505)*	1331 *(458)*	1319 *(459)*

Subjective Ratings of Sound Clarity

Subjective clarity ratings from listeners in each condition were compared based on the univariate effects from the MANOVA. Figure 4 shows the clarity rating for each condition.

Higher clarity ratings indicate a clearer signal. No effect of HA was found [F (1, 23) = 1.938, p=0.117, η^2_p=0.078], suggesting that both devices provide comparable listener ratings of overall clarity. Similar to VRT, significant main effects for SNR [F (1, 23) = 18.2, p<0.001, η^2_p=0.442] and DNR [F (1, 23) = 8.282, p=0.008, η^2_p=0.265] were found, indicating that clarity ratings improved as input SNR increased and were better (higher) in the DNR-on condition. The HA x SNR [F (1, 23) = 1.841, p=0.188, η^2_p=0.074], HA x DNR [F (1, 23) = 0.596, p=0.448, η^2_p=0.025], DNR x SNR [F (1, 23) = 0.268, p=0.609, η^2_p=0.012] and three-way interaction [F (1, 23) = 0.218, p=0.645, η^2_p=0.009] were not statistically significant for listeners' ratings of clarity.

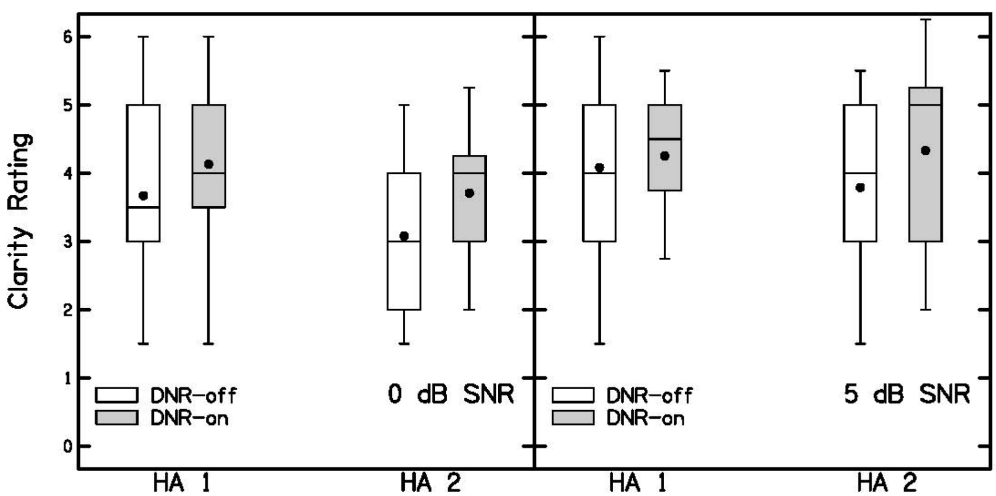

Figure 4. Clarity ratings for each device by DNR condition at 0 dB (left panel) and +5 dB (right panel) SNR. The boxes represent the interquartile range and the error bars represent the 5-95% confidence intervals of the mean. Filled circles represent the means, while the solid horizontal lines represent the medians for each condition.

Speech Recognition

Performance for CVC non-word recognition was calculated as the percentage of phonemes repeated correctly. Figure 5 shows phoneme recognition in percent correct for each device in the DNR-off and DNR-on conditions for each input SNR condition. The influence of experimental variables on phoneme recognition was examined using the univariate tests from the MANOVA. As expected, overall phoneme recognition improved as a function of SNR [$F(1, 23) = 309.306$, $p<0.001$, $\eta^2_p = 0.931$]. There was no overall mean difference in phoneme recognition between the two HAs [$F(1, 23) = 0.033$, $p=0.858$, $\eta^2_p = 0.001$]. The main effect for DNR was significant [$F(1, 23) = 11.644$, $p=0.002$, $\eta^2_p = 0.336$]. A significant device x DNR interaction indicated that the improvement in phoneme recognition with DNR varied by device [$F(1, 23) = 20.838$, $p<0.001$, $\eta^2_p = 0.475$]. A significant improvement in phoneme recognition occurred only with HA2. Specifically, the difference in phoneme recognition between DNR on and DNR off for HA1 was not significant (62% / 61.2%), but phoneme recognition for HA2 was significantly higher for DNR on (65%) than DNR off (58.6%). The device x SNR interaction [$F(1, 23) = 0.304$, $p=0.587$, $\eta^2_p = 0.013$], SNR x DNR interaction [$F(1, 23) = 1.177$, $p=0.289$, $\eta^2_p = 0.049$] and three-way interaction [$F(1, 23) = 0.002$, $p=0.963$, $\eta^2_p < 0.001$] were not statistically significant.

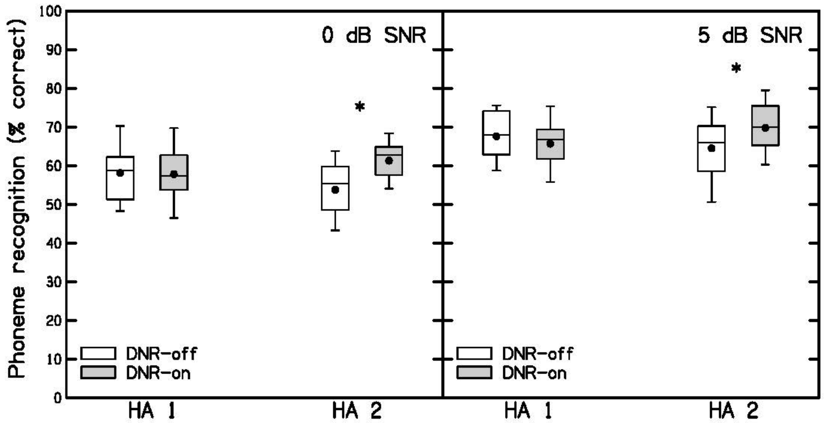

Figure 5. Phoneme recognition for each device is plotted in the DNR-off and DNR-on conditions at 0 dB (left panel) and +5 dB (right panel) input SNR. The asterisk indicates a significant improvement between DNR-off and DNR-on conditions. The boxes represent the interquartile range and the error bars represent the 5-95% confidence intervals of the mean. Filled circles represent the means, while the solid horizontal lines represent the median for each condition.

Discussion

The primary purpose of this study was to evaluate how DNR processing impacts listening effort and subjective judgment of sound clarity in children. A secondary purpose of this study was to evaluate the use of the inversion method to characterize listening effort using DNR processing. Overall, VRT decreased and children's ratings of clarity improved with DNR compared to conditions without DNR for both hearing aids used in the current study. Speech recognition was the same with and without DNR for HA1, while a small (6.4%), but statistically significant, improvement in phoneme recognition was observed for HA2 when DNR was active.

Based on the expectation that VRT reflects listening effort, VRT was expected to increase as the listening conditions became more difficult (e.g. conditions without DNR or with a 0 dB SNR). As hypothesized, an increase in VRT was observed as the listening condition became more difficult with the deactivation of DNR. However, this increase in VRT with the more difficult SNR was only observed once data were converted to Z-scores. This inconsistent finding is likely due to the large within- and between-subject variability in our raw VRT data. Recall that the raw VRT analysis showed a small but significant effect of DNR ($p=0.033$, $\eta^2_p=0.183$) and an even smaller, but still significant, effect once analysis accounted for within-subject variability by converting to Z-scores ($p=0.034$, $\eta^2_p=0.182$). Conversely, a small effect approaching significance ($p=0.075$, $\eta^2_p=0.131$) was seen with the main effect of SNR, which became a small but significant effect after accounting for subject variability ($p=0.014$, $\eta^2_p=0.234$). This suggests that the variability in response times may have been greater when comparing 0 dB to +5 dB SNR than when comparing DNR on to DNR off. In the case of SNR, the variability of response distribution in the raw VRT likely diluted this already small effect for SNR, resulting in the non-significant main effect.

Increases in VRT have been proposed to reflect efficiency of short-term memory access (Cowan et al. 2003) and ease of listening (Gatehouse and Gordon 1990). In more challenging listening conditions, auditory processing requires greater cognitive effort and may lead to a delayed repetition of a verbal response in tasks such as the one used in this study. VRT results from this study are consistent with results from dual-task studies that have shown increases in reaction time with increased auditory task difficulty, including studies of DNR in adults (Sarampalis et al. 2009). The similarity in findings between this study and previous studies supports the use of VRT as a measure of listening effort.

Although these findings support the use of VRT as a measure of listening effort, significant individual variability in VRT was observed across participants. Developmental factors such as age and attention may have influenced this variability. These factors were not evaluated as part of the present investigation. The current study attempted to minimize the influence of individual variability by converting each participant's VRT in ms to a Z-score based on the average and standard deviation from that participant across all conditions. While this transformation allows for a relative comparison of each child's VRT for each condition to their own average, comparisons of VRT using these transformed Z-scores across participants or descriptions of the size of differences across conditions in ms cannot be reliably made due to the use of participant averages across groups rather than group averages across conditions.

Few, if any, studies have examined sound quality ratings in children who use devices with DNR. Ratings of clarity were expected to follow a pattern similar to that observed in VRT studies (i.e., higher ratings of clarity found in the DNR conditions or at the more favorable +5 dB SNR). On average, children's clarity ratings followed this prediction across conditions, consistent with improvements in sound quality for adults HA users with DNR engaged (Ricketts and Hornsby 2005; Mueller et al. 2006). However, similar improvements in clarity ratings were observed for both DNR systems evaluated in the current study, regardless of the differences observed between devices in the output SNR using the inversion method. A wide range of clarity ratings was observed across subjects for the same condition. Many previous studies of sound quality with DNR have used a forced-choice or paired-comparison task wherein participants rate their listening preference on a trial-by-trial basis. In the current study, children gave a single overall clarity rating for each condition, which may have created greater variability than would have been observed using more traditional methods.

Phoneme recognition was measured across conditions to determine if DNR affected perception. Because DNR acts on the combined speech and noise signal to reduce gain, improvements in speech recognition were not expected and have not been observed in previous studies with adults (Boymans and Dreschler 2000; Alcantara et al. 2003; Ricketts and Hornsby 2005) or children (Stelmachowicz et al. 2010). Consistent with previous findings, no difference in speech recognition was observed for HA1. However, contrary to previous studies, the use of DNR in HA2 resulted in a small (6.4%), but statistically significant improvement in phoneme recognition. To ensure that the significant improvement measured in HA2 was not due to performance with DNR off being poorer with HA2 when compared to HA1, a Tukey's Honestly Significant Difference (HSD) post hoc test was completed to examine the pattern of phoneme recognition scores across HA and DNR condition while controlling for Type I error. The calculated minimum mean significant difference for the interaction was 4.6%. While phoneme recognition improved significantly when DNR was activated for HA2, significant differences in scores were not observed between HA2 in the DNR off condition and either DNR condition of HA1. This was likely because performance was slightly (but not significantly) poorer with DNR off for HA2 compared to HA1 but slightly (but again, not significantly) better with DNR on for HA2 compared to HA1. This resulted in equivalent performance between the two devices regardless of whether or not DNR was activated. Finally, while the difference between DNR on and DNR off conditions for HA2 was statistically significant, this effect is smaller than the increases in effect size for phoneme recognition observed between 0 dB and 5 dB SNRs (9%).

The underlying mechanisms for improvements in phoneme recognition observed with the activation of DNR in HA2 remain unclear. Analyses of signal quality are not conclusive. Recall that inversion method measurements made with HA2 showed a 6.84 dB SNR difference when DNR was activated and only a 2.63 dB SNR difference between DNR conditions in HA1 averaged across SNR conditions. However, the change in SNR measured using the inversion method reflects a long-term average of SNR over the entire signal and is not analogous to a change in the SNR that would be necessary to improve perception. It is difficult to compare the results from the present study with previous studies of speech recognition because performance in SNR conditions poorer than 0 dB SNR is rarely reported. However, based upon studies in NH children where the SNRs were more favorable (Bradley and Sato 2008; McCreery et al. 2010), it is likely that SNR improvement measured using the inversion method is not directly related to a true SNR improvement. Recall that phoneme recognition improved only 5.32% in the 0 dB input SNR HA2 condition (from -3.04 to 3.96 dB SNR resulting in a change of 7 dB). Results of McCreery and Stelmachowicz (2011) suggest that this amount of SNR improvement should have yielded a larger performance increase. Using the same CVC nonwords stimuli as used in the present study and NH children of the same age-range, McCreery and Stelmachowicz (2011) found that with an increase from 0 to +6 dB SNR, children's performance improved by 30-40%. The discrepancy between their data and those found in the present study's results suggests that the measured change in SNR using the inversion method should not be viewed as comparable to an actual change in SNR. Because the inversion method assumes a linear, time-invariant system, it not an appropriate representation of the compressive and time-varying nature of digital HA algorithms. Based on the results of this study, the inversion method is unlikely to accurately predict speech recognition performance with digital signal processing in the same way that would occur by independently altering the level of the speech or noise to change SNR at the input.

A comparison of the magnitude-squared coherence (Figure 2) for both devices with DNR on and DNR off suggests larger differences in spectral distortion for HA2 compared to HA1. However, these changes were observed for both the speech and noise signals as they are processed concurrently through the hearing aid, and are not sufficient to explain the enhancement of phoneme recognition observed with HA2. The coherence of the speech and noise signals in HA1 reflect minimal differences in spectral distortion between DNR conditions, which would support the observed effect of no difference in phoneme recognition with that device.

Improvements in clarity and reductions in listening effort observed with DNR could have facilitated the improvements in phoneme recognition observed with HA2. However, the magnitude of the effects for clarity and VRT observed in the current study were similar for both

HAs and thus would not be sufficient to explain an improvement in perception for only HA2. It is possible, however, that the measures of clarity and listening effort in this study may not be sensitive enough to differentiate the processes that supported improvements in phoneme perception with HA2 that were not observed with HA1. More sensitive measures of sound quality or listening effort may be needed to examine differences between signal processing algorithms that result in improved perception.

In the current experiment, all stimuli were scaled to 65 dB before being presented to the participants. Scaling the experimental stimuli to a common RMS of 65 dB may have obscured the anticipated effects of SNR and DNR. For example, if the output of HA1 were higher than HA2 prior to scaling to 65 dB SPL, it would require a greater reduction in gain to the output of HA1 in order to achieve the same 65 dB output as HA2. Because of this scaling, the effects of DNR and SNR in actual practice may have been reduced or enhanced for one of both devices during this experiment. Future studies should aim to preserve these relationships across conditions to prevent this confound.

Because output SNR as measured by the inversion method has not been reported in the majority of previous studies and the noise used in the present study is not typical of every-day background noise, results of previous studies of DNR effects (e.g. Sarampalis et al. 2009; Stelmachowicz et al. 2010) should be cautiously compared to those obtained here. DNR algorithms are designed to engage when the modulation of the input spectrum is steady-state, which is why the steady-state white noise masker was selected for the current study. The use of a broadband noise as a masker and activating stimuli for DNR may not reflect perceptual performance or the behavior of DNR algorithms in realistic listening environments. The use of this type of masker may have increased the likelihood of a significant DNR effect, but does not explain why an improvement in phoneme recognition was only observed for HA 2 or why improvements in clarity ratings and VRT were observed for both devices despite differences in the inversion method estimates of SNR. One potential explanation is that because both devices resulted in an improvement in the estimate of the SNR from the inversion method, the improvement was sufficient to improve clarity and VRT in both devices but only for HA 2 with regard to phoneme recognition. Further research is necessary to directly compare maskers with different spectral characteristics for the same algorithm before the influence of this factor on DNR can be quantified. The effect of DNR is likely to be diminished for maskers with spectral content similar to that of speech or with temporal modulations, because the degree to which DNR is engaged by these stimuli is likely to vary.

The CVC stimuli used in this study may have also contributed to the significant improvement in speech intelligibility. The speech recognition materials used in previous DNR studies were typically words and sentences, allowing the listener to utilize linguistic cues when the signal is degraded. In studies that use stimuli with redundant contextual cues such as sentences, performance differences with and without DNR may be more difficult to observe (Ricketts and Hornsby 2005). Stelmachowicz et al. (2010) attempted to reduce the number of linguistic cues by using vowel-consonant-vowel (VCV) stimuli where the initial and final vowels were identical. However, the resulting VCV stimuli were presented in a closed-set task and resulted in performance that was higher than performance for monosyllabic words. The use of CVC nonwords in the current study allowed for an open set of speech stimuli stripped of contextual cues. While recognition of CVC nonwords may not predict speech understanding for children in situations where semantic and syntactic context are readily available, such as a noisy classroom, improved phoneme recognition with DNR algorithms may benefit a young child by improving clarity of individual phonemes during language development.

Although the results of the present study suggest the potential for benefits of DNR for school-age children, this study was conducted with NH children and these results may not be representative of the effects of DNR for children with HL, particularly since current algorithms can vary widely across differing degrees and configurations of HL. Future research is needed to determine if the patterns of performance observed in this study are consistent across varying types of competing noise and varying degrees and configuration of HL.

Conclusions

The overall goal of this study was to examine how DNR impacts listening effort and subjective judgments of clarity in children, and to better define the relationship between SNR improvement as measured with the inversion method and predicted performance on these same measures. Results suggest that the DNR systems used in this study reduce listening effort and improve ratings of clarity in children; however, these improvements are not dependent upon a significant improvement in SNR as measured by the inversion method or a preserved speech spectra provided by the DNR. Contrary to previous findings, results suggest that phoneme perception improves with the use of one DNR algorithm evaluated in the present study, although performance between systems was similar. Specifically, the DNR algorithm used in HA2 provided less distortion of the speech signal when compared to the other DNR algorithm. Further studies are needed to assess the effects of this measured improvement in SNR with DNR processing for children with HL, and to determine how the amount of SNR improvement relates to speech recognition, sound quality and listening effort. Finally, conclusions drawn from this study are unique to the two HAs used, to the type of competing noise selected, and to NH children and thus, cannot be generalized to DNR in other devices or to children with HL.

Acknowledgements

The authors would like to thank Andrea Pittman for her review of the manuscript prior to submission, Kanae Nishi for developing scripts to measure verbal response time, Marc Brennan for his comments during revision, and Daniel Ashmead for his assistance with statistical analysis. Research was supported by NIH grants RO1 DC4300 10, R03 DC009334, T35 DC8757, F31 DC010505, and P30 DC4662.

References

Alcantara, J. L., Moore, B. C., Kuhnel, V., et al. (2003). Evaluation of the noise reduction system in a commercial digital hearing aid. *Int J Audiol*, 42, 34–42.

Anzalone, M. C., Calandruccio, L., Doherty, K. A., et al. (2006). Determination of the Potential Benefit of Time-Frequency Gain Manipulation. *Ear Hear*, 27(5), 480-492.

Arnold, P. & Canning, D. (1999). Does classroom amplification aid comprehension? *British Journal of Audiology,* 33, 171-178.

Bankson, N. W., & Bernthal, J. E. (1990). Quick Screen of Phonology. San Antonio, TX: Special Press, Inc.

Bentler, R., & Chiou, L. K. (2006). Digital noise reduction: An overview. *Trends Amplif, 10*(2), 67-82.

Bentler, R., Wu, Y. H., Kettel, J., et al. (2008). Digital noise reduction: Outcomes from laboratory and field studies. *Int J Audiol, 47*(8), 447-460.

Boymans, M., & Dreschler, W. A. (2000). Field trials using a digital hearing aid with active noise reduction and dual-microphone directionality. *Audiology, 39*(5), 260-268.

Bradley, J. S., & Sato, H. (2008). The intelligibility of speech in elementary school classrooms. *J Acoust Soc Am, 123*(4), 2078-2086.

Chung, K. (2004). Challenges and Recent Developments in Hearing Aids Part I. Speech Understanding in Noise, Microphone Technologies and Noise Reduction Algorithms. *Trends Amplif*, 8(3), 83-124.

Cowan, N., Towse, J. N., Hamilton, Z., et al. (2003). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology. General*, 132(1), 113-132.

Crandell, C. C., & Smaldino, J. J. (1995) Speech perception in the classroom. In C. Crandell, J. J. Smaldino & C. Flexer (Eds). Sound-field FM Amplification: Theory and Practical Applications (pp, 29-48). San Diego, California: Singular Publication Group.

Crandell, C. C., & Smaldino, J. J. (2000). Classroom acoustics for children with normal hearing and with hearing impairment. *Lang Sp Hear Serv Sch*, 31, 362-370.

Gatehouse, S., & Gordon, J. (1990). Response times to speech stimuli as measures of benefit from amplification. *British Journal of Audiology*, 24(1), 63-68.

Hagerman, B., & Olofson, A. (2004). A method to measure the effect of noise reduction algorithms using simultaneous speech and noise. *Acta Acustica*, 90, 356-361.

Hallgren, M., Larsby, B., Lyxell, B., et al. (2001). Evaluation of a cognitive test battery in young and elderly normal-hearing and hearing-impaired subjects. *J Am Acad Audiol*, 12, 357-370.

Hallgren, M., Larsby, B., Lyxell, B., et al. (2005). Speech understanding in quiet and noise, with and without hearing aids. *Int J Audiol*, 44, 574-583.

Heathcote, A., Popiel, S.J., Mewhort, D.J.K. (1991). Analysis of Response Time Distributions: An Example Using the Stroop Task. *Psychological Bulletin,* 109(2), 340-347.

Hervey, A.S., Epstein, J.N., Curry, J.F., Tonev, S., Arnold, L.E., Conners, C.K., Hinshaw, S.P., Swanson, J.M. (2006). Reaction Time Distribution Analysis of Neuropsychological Performance in an ADHD Sample. *Child Neuropsychology,* 12(2), 125-140.

Hicks, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *J Speech Lang Hear Res, 45*(3), 573-584.

Hoetink, A. E., Korossy, L., & Dreschler, W. A. (2009). Classification of steady state gain reduction produced by amplitude modulation based noise reduction in digital hearing aids. *Int J Audiol, 48*(7), 444-455.

Hu, Y., & Loizou, P. C. (2007). A comparative intelligibility study of single-microphone noise reduction algorithms. *J Acoust Soc Am, 122*(3), 1777.

Jamieson, D. G., Brennan, R. L., & Cornelisse, L. E. (1995). Evaluation of a speech enhancement strategy with normal-hearing and hearing-impaired listeners. *Ear Hear, 16*(3), 274-286.

Jerger, S. (2007). Current state of knowledge: perceptual processing by children with hearing impairment. *Ear Hear,* 28(6), 754-765.

Kahneman, D. (1973). Attention and Effort. Englewood Cliffs, NJ: Prentice Hall.

Kates, J. M., & Arehart, K. H. (2005). Coherence and the speech intelligibility index. *J Acoust Soc Am*, 117(4 Pt 1), 2224-2237.

Kates, J. M. (2008). Digital hearing aids. Plural: San Diego.

Larsby, B., Hallgren, M., Lyxells, B., et al. (2005). Cognitive performance and perceived effort in speech processing tasks: Effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *Int J Audiol*, 4,131-143.

Lewis, J. D., Goodman, S. S., Bentler, R. A. (2010). Measurement of hearing aid internal noise. *J Acoust Soc Am, 127*(4), 2521-2528.

Luts, H., Eneman, K., Wouters, J., et al. (2010). Multicenter evaluation of signal enhancement algorithms for hearing aids. *J Acoust Soc Am, 127*(3), 1491-1505.

Ma, J. & Loizou, P. C. (2011) SNR loss: A new objective measure for predicting the speech intelligibility of noise-suppressed speech. *Speech Commun*, 53(3), 340-354.

Mackersie, C, Neuman, A.C., Levitt, H. (1999). A Comparison of Response Time and Word Recognition Measures Using a Word-Monitoring and Closed-Set Identification Task. *Ear Hear*, 20(2), 140-148.

McCreery, R. W., Ito, R., Spratford, M., et al. (2010). Performance-Intensity Functions for Normal-Hearing Adults and Children Using Computer-Aided Speech Perception Assessment. *Ear Hear,* 31(1), 95–101.

McCreery, R. W. & Stelmachowicz, P. G. (2011). Audibility-based predictions of speech recognition for children and adults with normal hearing. *J Acoust Soc Am,* 130 (6), 4070-4081.

Mueller, H. G., Weber, J., Hornsby, B. W. (2006). The effects of digital noise reduction on the acceptance of background noise. *Trends Amplif, 10*(2), 83-93.

Natarajan, A., Hansen, J., Arehart, K., et al. (2005). An auditory-masking threshold-based noise suppression algorithm for listeners with sensorineural hearing loss, *EURASIP. J Appl Signal Proc*, *18*, 2938–2953.

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.

Pichora-Fuller, M. K., Schneider, B. A., Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J Acoust Soc Am*, 97, 593–608.

Pisoni, D. B., Kronenberger, W. G., Roman, A. S., et al. (2010). Measures of Digit Span and Verbal Rehearsal Speed in Deaf Children After More Than 10 Years of Cochlear Implantation. *Ear Hear,* 32(1), 60S-74S.

Pittman, A.L. (2011a). Age-Related Benefits of Digital Noise Reduction for Short-Term Word Learning in Children with Hearing Loss. *J Speech Lang Hear Res,* 54, 1448-1463.

Pittman, A. L (2011b). Children's performance in complex listening conditions: Effects of hearing loss and digital noise reduction. *J Speech Lang Hear Res,* 54, 1224–1239.

Rakerd, B., Seitz, P. F., Whearty, M. (1996). Assessing the cognitive demands of speech listening for people with hearing losses. *Ear Hear, 17*(2), 97-106.

Ratliff, R. (1979). Group Reaction Time Distributions and an Analysis of Distribution Statistics. *Psych. Bulletin, 86(3),* 446-461.

Ratliff, R. (1993). Methods for dealing with reaction time outliers. *Psych. Bulletin, 114(3),* 512-532.

Ricketts, T. A., & Hornsby, B. W. (2005). Sound quality measures for speech in noise through a commercial hearing aid implementing digital noise reduction. *J Am Acad Audiol, 16*(5), 270-277.

Rouder, J. N., Sun, D., Speckman, P. L., et al. (2003). A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika, 68(4),* 589-606.

Sarampalis, A., Kalluri, S., Edwards, B., et al. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *J Speech Lang Hear Res, 52*(5), 1230-1240.

Scollie, S., Seewald, R., Cornelisse, L., et al. (2005). The Desired Sensation Level multistage input/output algorithm. *Trends Amplif,* 9(4), 159-197.

Slocomb, D. & Spencer, K. A. (2009). The Effect of Feedback Schedule Manipulation on Speech Priming Patterns and Reaction Time. *J of Psycholinguistic Res,38(1)*, 43-64.

Souza, P., Jenstad, L. M., Boike, K. T. (2006). Measuring the acoustic effects of compression amplification on speech in noise. *J Acoust Soc Am*, Volume 119, Number 1, pp. 41-44

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *J Exp Psychol Gen*, 112, 1–36.

Stelmachowicz, P. G, Lewis, D. E., Hoover, B. M, et al. (2010). Effects of digital noise reduction on speech perception for children with hearing loss. *Ear Hear, 31*(3), 345-355.

Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., et al. (2004). Novel-word learning in children with normal hearing and hearing loss. *Ear Hear*, 25, 47–56.

Storkel, H. L. & Hoover, J. R. (2010). An on-line calculator to compute phonotactic probability and neighborhood density based on child corpora of spoken American English. *Behav Res Methods*, 42, 497-506.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J Speech Hear Res*, 28(3), 455-462.

Whelan, R. (2008). Effective Analysis of Reaction Time Data. *The Psychological Record*, 58, 475-482.

Wu, Y. H., & Stangl, E. (2013). The Effect of Hearing Aid Signal-Processing Schemes on Acceptable Noise Levels: Perception and Prediction. *Ear and hearing*.

Zekveld, A. A., Kramer, S. E., Festen, J. M. (2011) Cognitive Load during Speech Perception in Noise: The Influence of Age, Hearing Loss, and Cognition on the Pupil Response. *Ear Hear*, 32(4), 498-510.