K.H. Arehart, J.H.L. Hansen, S. Gallant, L. Kalstein, "Evaluation of an Auditory Masked Threshold Noise Suppression Algorithm in Normal-Hearing and Hearing-Impaired Listeners," Speech Communication, vol. 40, pp. 575-592, May 2003.

Evaluation of an Auditory Masked Threshold Noise Suppression Algorithm in Normal-Hearing and Hearing-Impaired Listeners

Kathryn H. Arehart†, John H.L. Hansen, Stephen Gallant, and Laura Kalstein†

Center for Robust Speech Systems
Erik Jonsson School of Engineering & Computer Science
Department of Electrical Engineering
The University of Texas at Dallas
P.O. Box 830688, EC33
Richardson, TX 75083-0688
972 – 883 – 2910 (Phone) 972 – 883 - 2710 (Fax)
John.Hansen@utdallas.edu (email)

†**Dept. Speech, Language & Hearing Sciences**,
University of Colorado Boulder,
Boulder, Colorado 80309-0594
303-492-3036 (Phone)
Kathryn.Arehart@colorado.edu (email)

Speech Communication,
vol. 40, pp. 575-592, May 2003.

Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

ELSEVIER

Speech Communication 40 (2003) 575–592

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Evaluation of an auditory masked threshold noise suppression algorithm in normal-hearing and hearing-impaired listeners

Kathryn Hoberg Arehart [a,*], John H.L. Hansen [a,b], Stephen Gallant [b], Laura Kalstein [a]

[a] Department of Speech, Language and Hearing Sciences, Box 409 UCB, University of Colorado at Boulder, Boulder, CO 80309, USA

[b] Robust Speech Processing Laboratory, Center for Spoken Language Research, Box 594 UCB, University of Colorado at Boulder, Boulder, CO 80309, USA

Received 6 February 2002; accepted 22 October 2002

Abstract

While there have been numerous studies in the field of speech enhancement, the majority of these studies have focused on noise reduction for normal-hearing (NH) individuals. In addition, no speech enhancement algorithms reported in the signal processing community have reported an improvement in intelligibility, with the exception of a recent study by Tsoukalas et al. [IEEE Transactions of Speech and Audio Processing 5 (6) (1997) 497]. This study addresses the problem of speech enhancement for both NH and hearing-impaired (HI) subjects. A noise suppression algorithm based on auditory masked thresholds was implemented and evaluated for NH and HI subjects. Two different tests for intelligibility were used in the evaluation including the nonsense syllable test and the diagnostic rhyme test. Speech quality was evaluated using sentences from the hearing-in-noise test. Tests were performed using two types of noise (voice communications channel and automobile highway noise) at two different signal-to-noise ratios. Ten NH and 11 HI listeners were used to evaluate the enhancement algorithm. Results indicate that the enhancement algorithm yielded significantly better quality ratings and significantly better intelligibility scores in both NH and HI listeners in some but not all of the test conditions. The algorithm resulted in the greatest intelligibility improvements in the communications channels noise and for the nonsense syllable stimuli.

Keywords: Noise suppression; Hearing loss; Hearing aids; Auditory masked threshold; Speech intelligibility; Speech quality

1. Introduction

Hearing aids are the common prescription for individuals with hearing loss due to cochlear damage. Recent studies document the benefit hearing aids provide listeners. In their large multicenter clinical trial, Larson et al. (2000) demonstrated that three common hearing aid circuits all provided significant benefit to hearing-impaired (HI) listeners when compared to unaided listening conditions. However, this same study cited statistics from recent surveys indicating that only one in

* Corresponding author.
E-mail address: kathryn.arehart@colorado.edu (K.H. Arehart).

0167-6393/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0167-6393(02)00183-8

five individuals with hearing loss who might benefit from hearing aids actually wears hearing aids and only 65% of individuals with hearing aids report that they are satisfied with their hearing instruments. While beneficial, hearing aids do not restore normal audition. The varied successes of our strategies for hearing aid design reflect our incomplete understanding of specific auditory deficits that cause decreased speech recognition in individual listeners with hearing loss.

Listeners with cochlear hearing loss have much more difficulty understanding speech than do normal-hearing (NH) listeners. This increased difficulty is especially pronounced in noisy environments. Speech recognition thresholds refer to the speech-to-noise ratio required to achieve a particular level of intelligibility (usually 50%). Speech recognition thresholds for one voice in the presence of a competing voice are up to 12 dB worse in listeners with hearing loss than in listeners with NH (Duquesnoy, 1983; Festen and Plomp, 1990; Hygge et al., 1992).

A primary factor contributing to the increased difficulty in understanding speech in noise is the reduced audibility of speech sounds in listeners with elevated auditory thresholds (Zurek and Delhourne, 1987; Humes et al., 1987). Other perceptual factors also contribute to human speech recognition deficits. These factors have been described generally as "distortion" (Plomp, 1986) and as deficits in "suprathreshold discrimination abilities" (Moore, 1998). Suprathreshold discrimination deficits that may affect speech understanding in noise include reduced dynamic range/loudness recruitment, degraded frequency selectivity and reduced temporal resolution (e.g., Hou and Pavlovic, 1994; Baer and Moore, 1994; Moore et al., 1995; Glasberg and Moore, 1989; Festen and Plomp, 1983; Van Rooj and Plomp, 1990).

Hearing aids incorporate a number of different strategies to compensate for reduced audibility and suprathreshold processing deficits. These strategies include frequency-dependent amplification, compression, and directional microphones. Digital signal processing (DSP) hearing aids may also incorporate algorithms for feedback cancellation and for noise reduction/spectral enhancement. For example, all new multichannel hearing aids are digital (Mueller, 2000) and use DSP to set within and across channel gain and compression functions. Recently, several new digital instruments have appeared where the algorithm for channel processing for gain and compression is based on models of cochlear function.

DSP hearing aids also incorporate active noise reduction algorithms. The goal of these systems is to reduce the deleterious effects of unwanted noise in the input speech signal. Some hearing aids employ a noise cancellation scheme based on estimation of the speech-to-noise ratio. When the speech-to-noise ratio degrades to a determined criterion level, the noise cancellation system is activated and causes a decrease in gain in one or more frequency channels.

Spectral subtraction has also been considered as a noise reduction strategy for hearing aid applications (e.g., Elberling et al., 1993; Levitt et al., 1993; Jamieson et al., 1995). One of the first algorithm formulations of spectral subtraction was developed by Boll (1979) and reflects a range of techniques that remove stationary background noise by subtracting the noise power spectrum estimated during silent frames or from a reference channel from the noisy speech signal. While the mathematical formulation of spectral subtraction was originally based on subtraction in the power spectral domain, numerous variations have been proposed in the Fourier transform domain, cepstrum domain, autocorrelation domain and using a wavelet transform. These domains have been considered with the goal of improved noise suppression with less processing artifacts. For many commercially available digital hearing aids, the Fourier transform domain is used with the noise being estimated based on its steady state characteristics and speech being estimated based on its modulation characteristics.

Generally, noise reduction circuits using spectral subtraction in both telecommunications and hearing aid applications use mathematical criteria based on the estimated speech-to-noise ratio. These types of algorithms generally result in quality improvement, but not in intelligibility improvement (e.g., Elberling et al., 1993; Levitt et al., 1993; Jamieson et al., 1995). For example, Levitt et al. (1993) compared four noise reduction methods, including a two-microphone adaptive noise canceller, short-term Wiener filtering, a transformed spectrum subtraction technique, and sinusoidal modeling. Significant improvements in speech intelligibility were observed for the two-microphone noise canceller and for some HI listeners with the single-microphone short-term Wiener filtering system. The stimuli processed with the single-microphone spectral subtraction technique did not yield significant improvements in intelligibility. Similarly, Elberling et al. (1993) studied three different spectral subtraction algorithms with a four-alternative forced-choice speech intelligibility test. Results indicated that the noise reduction algorithms decreased the level of noise, but did not result in speech intelligibility improvements in either a group of NH listeners or in a group of HI listeners. Jamieson et al. (1995) evaluated an adaptive Wiener filtering approach to speech enhancement in three different noise conditions (wideband noise, narrow-band noise and multitalker babble). Using a paired comparison procedure, Jamieson et al. (1995) showed that listeners preferred speech processed with the Wiener filtering strategy at positive signal-to-noise ratios (SNRs) (+10, +20 dB). They also reported significant gains in intelligibility as measured by a spondee speech recognition threshold test in narrowband noise but not in wideband noise or in multitalker babble. Finally, listeners showed no change or a decrease in speech intelligibility with the processed speech for a distinctive feature nonsense syllable speech test.

One possible reason for the lack of intelligibility improvement relates to musical noise artifacts caused by these processing techniques (Cappa, 1994). A primary issue is the balance between pure noise suppression within a signal-plus-noise model using a mathematical based criterion such as SNR and the exact components that contribute to changes in speech intelligibility in noise. Another reason has to do with the difficulty and repeatability of assessing intelligibility in noisy speech. A number of studies have considered various measures of objective speech quality (Hansen, 1999), but in the speech enhancement community, intelligibility is less well understood because of the differences in performance across various types of distortion.

One recent study reported a significant gain in speech intelligibility using a spectral subtraction technique based on aspects of the auditory process (Tsoukalas et al., 1997). The method of Tsoukalas et al. (1997) considers an enhancement approach that uses the auditory masked threshold (AMT) in conjunction with a version of spectral subtraction to adjust the parameters used in the subtraction process based on the masked threshold of the noise across the frequency spectrum. One of the important aspects of the study by Tsoukalas et al. (1997) is the claim that their method improves intelligibility by as much as 40% as determined by listening tests using a standard English version and their Greek version of the diagnostic rhyme test (DRT). The listeners in Tsoukalas et al. were a mixture of native English speakers and native Greek speakers with English experience.

In this study, our goals were as follows: (1) evaluate the effectiveness of the auditory masked threshold noise suppression (AMT-NS) technique in improving the quality and intelligibility of speech in native English-speaking listeners, (2) determine whether the effectiveness of the algorithm in NH and HI listeners varies as a function of noise type, speech type and SNR and (3) determine potential hearing aid applications of the algorithm by determining if the algorithm is equally effective for both NH and HI listeners. To address these three goals, the intelligibility and quality of speech in noise for both degraded and enhanced conditions was measured in 10 listeners with NH and in 11 listeners with cochlear hearing loss.

2. Algorithm

One limitation of many speech enhancement techniques, such as traditional spectral subtraction, is that they are based on a mathematical criterion (e.g., the estimated speech-to-noise ratio) that quite often does not correspond well to auditory perception. Recently, there have been several methods proposed by researchers in telecommunications to improve speech quality based on auditory processing or auditory masking properties. These include studies by Tsoukalas et al.

(1997), Virag (1999), Nandkumar and Hansen (1995) and Hansen and Nandkumar (1995). Nandkumar and Hansen used a sequential maximum A posteriori (MAP) estimation procedure of the speech modeling parameters and noise-free speech. The primary novel aspect in these two schemes is the incorporation of auditory processing constraints between MAP estimation steps (algorithms termed Auditory Constrained Enhancement I and II: ACE-I, ACE-II). While these methods employed auditory-based filters with a neural lateral inhibition response, there was no direct processing that focused on the noise-masking threshold. Tsoukalas et al. used an enhancement approach based on the definition of the psychoacoustically derived quantity of audible noise spectrum and its subsequent suppression using nonlinear filtering in the short-time spectral domain. The method proposed by Virag is similar, but the manner in which the noise-masking threshold is obtained is different. Sarikaya and Hansen have also formulated an improved approach to the AMT estimation by including a clean and noisy AMT codebook for improved threshold estimation.

The AMT approach to speech enhancement assumes that some components of the noise will be masked (below threshold) by the speech. Any noise components that are below this masked threshold will not be detectable by the human listener and so perceptually are not important components to suppress. The goal, then, is to minimize only the audible portion of the noise spectrum. This notion has been used extensively in the audio coding field for data compression (Johnston, 1988) and has only recently been applied to speech enhancement. Enhancement schemes based on the AMT typically use spectral subtraction and adjust the parameters used in the subtraction process based on the masked threshold of the noise across the frequency spectrum.

In a single-microphone system, the noisy speech signal $y(n)$ consists of clean speech $x(n)$ and noise component $d(n)$:

$$y(n) = x(n) + d(n) \tag{1}$$

Here, the assumptions are that (i) the speech and noise signals are uncorrelated at least over a short-time basis, (ii) the noise is either stationary or slowly varying over several frames of speech and (iii) that the noise signal can be represented as zero mean random process. Let the spectrum of the noisy speech be represented as $\Gamma_y(f)$. With this notation, the spectrum of the noisy speech can be represented as the spectrum of the clean speech and noise as follows:

$$\Gamma_y(f) = \Gamma_x(f) + \Gamma_d(f) \tag{2}$$

since $d(n)$ is an uncorrelated random process. Given $\Gamma_y(f)$, it is possible to estimate the spectrum of the uncorrelated speech as:

$$\widehat{\Gamma}_x(f) = \Gamma_y(f) - \widehat{\Gamma}_d(f) \tag{3}$$

While this formulation is of interest theoretically, the signals $y(n)$ and $x(n)$ are generally actual waveforms. Recognizing that, at best, $x(n)$ will be "locally stationary" over short-time ranges, we therefore select a frame of $y(n)$, using a window of length N ending at time m, $x(n;m) = x(n)w(m-n)$. Therefore, the selected frame can be expressed in terms of the underlying speech and noise frames as follows:

$$y(n;m) = x(n;m) + d(n;m) \tag{4}$$

Therefore, by analogy to Eq. (3), we can write the clean speech short-term power spectrum estimate as:

$$\widehat{\Gamma}_x(f;m) = \Gamma_y(f;m) - \widehat{\Gamma}_d(f;m) \tag{5}$$

As is well known, the short-term power density spectrum can be related to the short-term discrete-time Fourier transform in a simple way as shown below,

$$\Gamma_y(f;m) = \frac{S_y(f;m)S_y^*(f;m)}{N^2} = \frac{|S_y(f;m)|^2}{N^2} \tag{6}$$

With no loss of generality, we can drop the $1/N^2$ term, so the density is written as $\Gamma_y(f;m) = |S_y(f;m)|^2$. Using the short-term Fourier transform, it may appear that an estimate of the noisy phase is also necessary. However, Wang and Lim (1982) have determined that for all practical purposes, it is sufficient to use the *noisy* phase spectrum, $\theta_y(f;m)$ as an estimate of the clean speech phase spectrum $\theta_s(f;m)$. Therefore, estimation of the frame of speech resulting from spectral subtraction is recovered from the short-term Fourier transform estimate as follows:

$$\begin{aligned}\widehat{S}_s(f;m) &= |\widehat{S}_s(f;m)|\mathrm{e}^{j\widehat{\Theta_s(f;m)}}\\ \widehat{S}_s(f;m) &= [\Gamma_y(f;m) - \widehat{\Gamma}_d(f;m)t]^{1/2}\mathrm{e}^{j\Theta_y(f;m)}\end{aligned} \tag{7}$$

where $\Gamma_y(f;m)$ and $\theta_y(f;m)$ are both obtained from the short-term Fourier transform of the present noisy speech frame,

$$S_y(f;m) = |S_y(f;m)|\mathrm{e}^{j\Theta_y(f;m)} = \Gamma_y^{1/2}(f;m)\mathrm{e}^{j\Theta_y(f;m)} \tag{8}$$

and $\Gamma_d(f;m)$ can be estimated using any frame of the signal in which speech is not present, or from a reference channel with noise only. This method is referred to as traditional spectral subtraction.

There have been numerous versions of spectral subtraction formulated over the past 20 years, in particular because of simplicity in computation, but also to address the typical musical tone artifacts that can arise because of the use of a average noise spectrum when the frame-to-frame noise spectrum normally varies (i.e., this results in frames which have components that are either over-subtracted or under-subtracted). We now turn to our discussion of the AMT, which was proposed for use in audio coding by Johnston (1988), and later integrated into a spectral subtraction method in the study by Tsoukalas et al. (1997). AMT defines a spectral amplitude threshold below which all frequency components are masked in the presence of the masker signal. A detailed derivation of AMT can be found in Johnston. Here, we summarize the main derivation steps in the calculation of AMT:

Step 1. Obtain energies in speech critical band (CB) analysis.
Step 2. Convolve the spreading function (Schroeder et al., 1979) with the CB spectrum to obtain a spread masking threshold.
Step 3. Compute offset term for spread masking thresholds to take into account signal tonality.
Step 4. Normalize/compare and account for the absolute auditory thresholds.

So, the steps needed to estimate the AMT are described in detail as follows:

Step 1: The energy $B_{P_{x,b}}(i)$, in CB b for frame i is obtained from the power spectrum $P_{x,b}(k,i)$, of the speech, where k_{lb} and k_{hb} are the lower and upper limits of CB b and B' is the total number of CBs and is dependent on the sampling frequency of the signal. The equation is shown below:

$$B_{P_{x,b}}(i) = \sum_{k=k_{lb}}^{k_{hb}} P_x(k,i) \tag{9}$$

Step 2: Next, CB energies are convolved with the basilar membrane spreading function $\mathrm{SPR}_{(.)}$ to obtain the spread CB spectra, $C_{P_{x,b}}(i)$

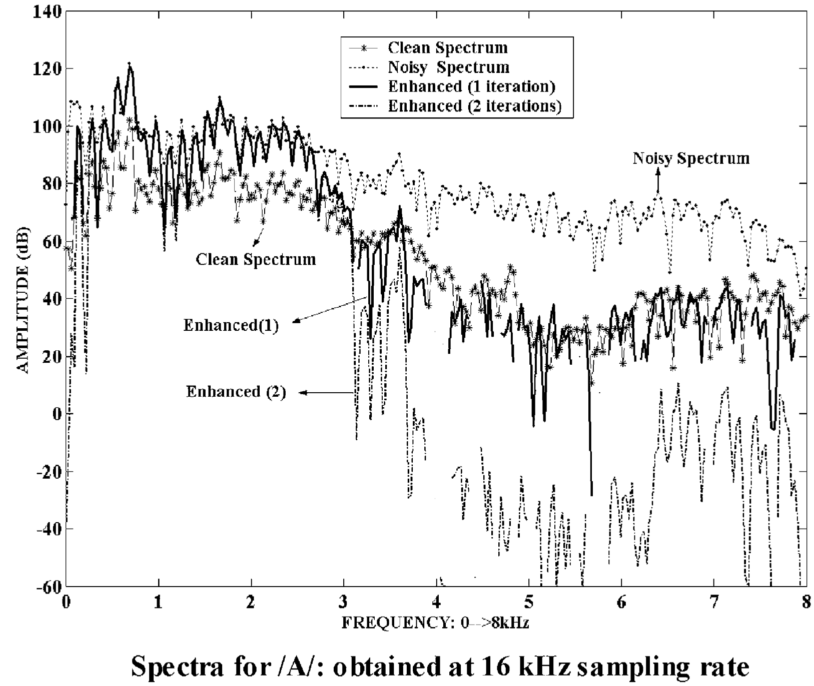
$$C_{P_{x,b}}(i) = \sum_{j=1}^{B'} \mathrm{SPR}_{b-j+B'} B_{P_{x,j}}(i) \tag{10}$$

Step 3: In order to determine the tone-like and noise-like natures of the spectrum, a spectral flatness measure (SFM) is used. In the first branch of Eq. (11), SFM is defined as the ratio of the geometric mean of the power spectrum $G_{P_x}(k,i)$ to the arithmetic mean $A_{P_x}(k,i)$ of the power spectrum. In the second branch, $\mathrm{SFM}_{P_x}(k,i)$ is used to generate a measure of tonality with $\mathrm{SFM}_{\max}$ which is equal to -60 dB for a sine wave. For white noise only, the SFM is equal to 0 dB and hence $\mathrm{ton}_{P_x}(k,i) = 0$. An offset, $O_{P_{x,b}}(i)$ is then estimated by which the threshold is adjusted to take signal tonality into account.

$$\begin{aligned}\mathrm{SFM}_{P_x}(k,i) &= \frac{G_{P_x}(k,i)}{A_{P_x}(k,i)}\\ \mathrm{ton}_{p_x,(k,i)} &= \min\left[\frac{10\log_{10}(\mathrm{SFM}_{p_x,}(k,i))}{\mathrm{SFM}_{\max}}, 1\right]\\ O_{P_{x,b}}(i) &= \mathrm{ton}_{p_x,(k,i)}(14.5+b)\\ &\quad + (1-\mathrm{ton}_{p_x,(k,i)})5.5 \quad 1 \leqslant b \leqslant B'\end{aligned} \tag{11}$$

Step 4: The final AMT is given in Eq. (12) after normalization and incorporating the absolute hearing threshold, $T_{\mathrm{abs}}(b)$, for each CB b.

$$T_{P_{x,b}}(i) = \max\left[T_{\mathrm{abs}}(b), \frac{T_{P_{x,b}}(i)}{\sum_{j=1}^{B'}\mathrm{SPR}_{b-j+B'}}\right] \tag{12}$$



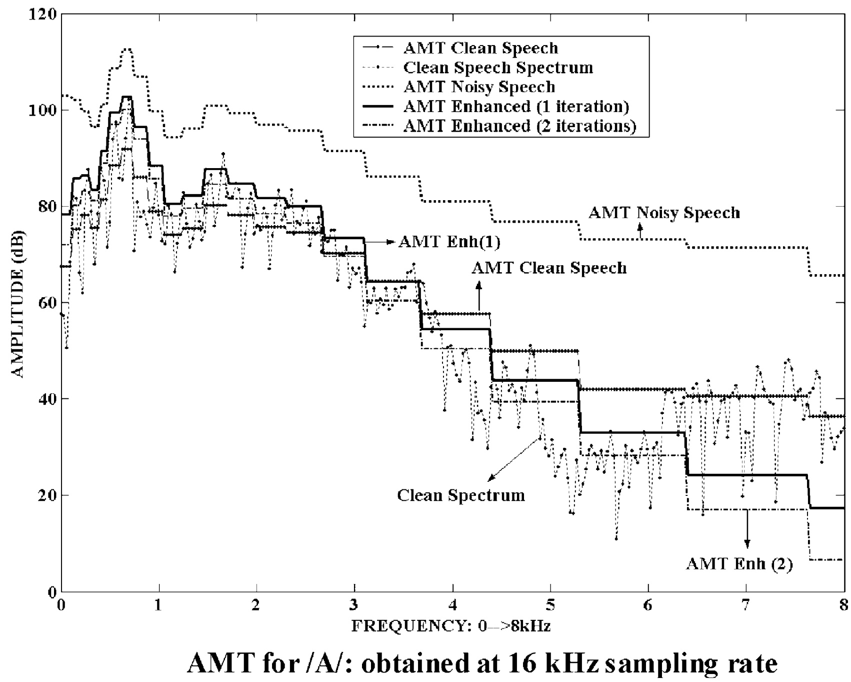


Fig. 1. Noisy, clean and enhanced power spectra of the vowel /a/ degraded with flat communication channels noise. The left panel shows the resulting power spectra and the right panel shows the clean speech spectrum with corresponding AMTs from the original noisy speech, clean speech, and first two iterations of the estimation scheme.

Having formed the estimate of the AMT, we now turn to the noise suppression formulation. The enhancement framework used is based on the scheme by Tsoukalas et al. (1997), which uses an AMT as developed in Johnston (1988) with a parametric nonlinear gain term developed by Clarkson and Bahgat (1989). The noise suppression scheme, AMT-NS, first estimates the portion of the audible portion of the noisy speech $A_y(k,i)$ that reflects the actual portion of the distortion that is audible $A_d(k,i)$. That is,

$$A_d(k,i) = A_y(k,i) - A_x(k,i). \quad 0 \leqslant k \leqslant K-1 \tag{13}$$

The goal is therefore to suppress $A_d(k,i)$ as much as possible across the CBs $0 \leqslant k \leqslant K-1$,

$$A_d(k,i) \leqslant 0 \tag{14}$$

Again, following Tsoukalas et al. (1997), we briefly summarize the solution that satisfies this criterion,

$$T_b(i) = \mathrm{AMT}(\widehat{P}_x^{(j-1)}(k,i)) \tag{15}$$

$$a_b(i) = [D_b + T_b(i)]\left[\frac{D_b}{T_b(i)}\right]^{1/v_b(i)} \tag{16}$$

$$\widehat{P}_X^j(k,i) = \frac{(\widehat{P}_X^{j-1})^{v_b}(k,i)}{a_b^{v_b}(i) + (\widehat{P}_X^{j-1})^{v_b}(k,i)}\widehat{P}_x^{j-1}(k,i) \tag{17}$$

where $T_b(i)$ is the masking threshold for CB b and speech frame i, D_b is the mean power spectrum of the noise in CB b, which is updated at each iteration of the algorithm. $\widehat{P}_X^{(j)}(k,i)$ is the estimate of the clean power speech spectrum at iteration (j) (note that for the first iteration $j=0$ the input speech frame represents the original degraded speech, $\widehat{P}_X^{(j-1)}(k,i) = P_Y(k;i)$). Here $a_b(i)$ defines a threshold below which frequency components of the noisy speech are highly attenuated, whereas $v_b(i)$ controls the rate of suppression. Therefore, Eq. (17) is a parametric nonlinear function that is used to obtain an estimate of the speech spectrum.

In order to illustrate the performance of the auditory masking threshold based noise suppression scheme, we consider a clean and degraded vowel section sampled at 16 kHz. Fig. 1 (left) shows the noisy, clean and enhanced power spectra of the vowel /a/. The vowel was degraded with flat communications channel noise. The left figure shows the resulting power spectra, and the right figure shows the clean speech spectrum and corresponding AMTs from the original noisy speech, clean speech and first two iterations of the estimation scheme. We see that AMT estimation after the first iteration comes close to what is obtained from the clean speech AMT. Noise suppression is also seen with one iteration of the algorithm. There is additional suppression after the second iteration of AMT-NS, but it appears that this may be due to differences in where the resulting AMT threshold is being estimated. We note that for the evaluations presented in this paper, we consider speech sampled at an 8 kHz rate (unlike that reported in Tsoukalas et al. (1997) which used a 16 kHz rate), and therefore we are only interested in performance in the 0–4 kHz frequency range. The results from Fig. 1 reflect the more desirable performance in the 0–4 kHz range versus the 4–8 kHz range based on mismatch in performance between clean and noisy estimated AMT thresholds. This point will be addressed further in the discussion section.

3. Evaluation

In this section, we describe the procedures used to evaluate the effectiveness of the AMT-NS scheme in both normal and HI listeners. Our current evaluation is focused on the 0–4 kHz frequency range, which was motivated by our earlier studies on speech enhancement for telephone/telecommunication applications (Hansen and Arslan, 1995), as well as computational restrictions for hearing aid systems.

We point out that in applying the AMT enhancement method, an iterative process was employed to estimate the AMT response. In the Tsoukalas et al. (1997) approach, after estimating the AMT, the original degraded speech is always submitted to the enhancement process, while the AMT response is updated (typically done for two iterations). For our implementation, we perform enhancement, not on the original degraded speech, but on the output of the previous enhancement filtering operation. In our previous work on constrained iterative Wiener filtering (Hansen and

Clements, 1991), we found similar levels of quality improvement for both schemes, but a slight reduction in noticeable processing artifacts if we re-filter the previous iteration's output in a subsequent enhancement iteration.

3.1. Stimuli

3.1.1. Speech materials

Three different sets of speech stimuli were used in this study. Speech quality was assessed using 256 sentences from the hearing-in-noise test (HINT) (Nilsson et al., 1994). Speech intelligibility was assessed using 102 syllables from the CUNY nonsense syllable test (NST) (Resnick et al., 1975) and 232 words from the DRT (Voiers, 1983). The speech stimuli were digitized at an 8 kHz sampling rate and stored on a Pentium III personal computer.

3.1.2. Noise conditions

All three sets of speech stimuli were degraded with two types of noise at two different SNRs. Speech stimuli were degraded with automobile highway noise at SNRs $= -5$ and 0 dB and with telephone communications channel noise at SNRs $= 0$ and 5 dB SNR. The time–frequency spectra of these noises are shown in Fig. 2. The communications channel noise has a relatively flat frequency response across the 4 kHz bandwidth and is stationary. The automobile highway noise is primarily low frequency (i.e., below 800 Hz) and is slowly varying. These noise sources were selected to establish performance for low versus high frequency distortion and stationary versus slowly varying distortion.

3.1.3. Signal processing

Digitized speech was degraded with sample noise files with appropriate scaling to generate each SNR. This set of "degraded" signals was then processed by the AMT-NS technique to generate the set of "enhanced" speech signals. In all enhancement processing, the noise spectrum was estimated during an initial portion of silence/noise prior to speech activity. Thus, there were both degraded and enhanced speech stimuli in each of the four noise conditions, making for a total of eight different sets of stimulus conditions.

3.2. Listeners

Ten listeners with NH and 11 listeners with hearing loss participated in this study. Listeners with NH had thresholds of 20 dB HL (ANSI, 1989) or better at octave frequencies from 250 to 8000 Hz, inclusive. Listeners with hearing loss demonstrated test results consistent with cochlear pathology: normal tympanometry; absence of otoacoustic emissions in regions of threshold loss and absence of an air-bone gap exceeding 10 dB at any frequency. Listeners with hearing loss had a

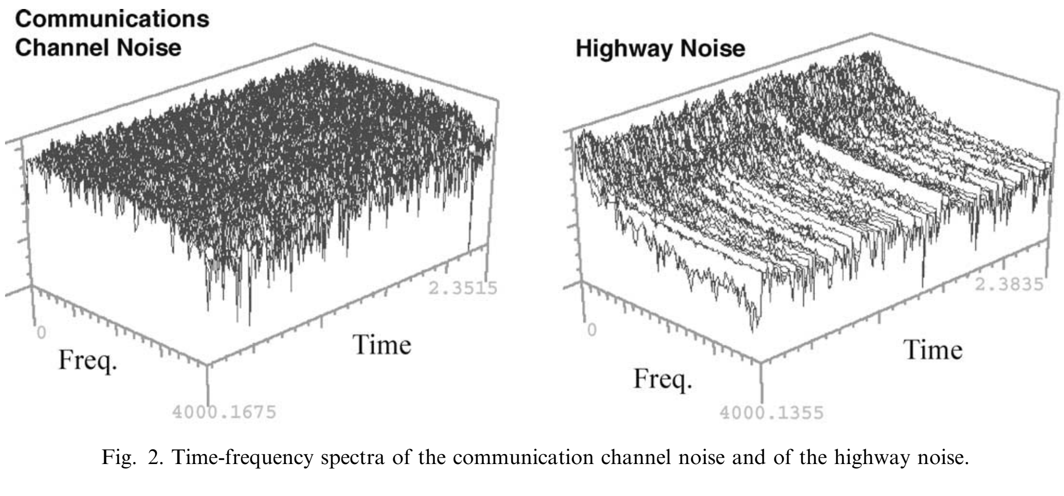


Fig. 2. Time-frequency spectra of the communication channel noise and of the highway noise.

Table 1
Age, test ear, and audiometric thresholds (in dB HL) of listeners with hearing loss

Listener	Age	Gender	Ear	Frequency (Hz)					
				250	500	1000	2000	4000	8000
HI1	72	M	R	45	35	45	75	70	75
HI2	63	M	R	30	55	60	60	55	75
HI3	38	F	L	25	20	50	70	65	50
HI4	24	F	R	40	50	60	70	70	80
HI5	72	M	R	10	20	20	40	65	70
HI6	43	F	R	30	25	30	55	95	75
HI7	25	F	R	30	40	45	50	55	55
HI8	43	F	L	35	50	60	70	70	55
HI9	60	M	R	35	25	40	80	75	60
HI10	68	F	L	70	70	65	65	50	50
HI11	78	M	L	20	15	15	65	70	60

mild to severe hearing loss. All listeners were tested monaurally. Table 1 provides a summary of the characteristics of the listeners with hearing loss, including the audiometric thresholds of the test ear. The test ear of the HI listeners was chosen based on the ear with a threshold configuration allowing the best digital filter design for linear amplification (see below). Listeners were tested individually in a double-walled sound booth. Daily test sessions typically lasted 1 h but did not extend beyond 2 h. Listeners were compensated \$8/h for their participation.

3.3. Equipment

For listener presentation, the digitally stored speech-in-noise stimuli went through a digital to analog converter (TDT AP2,DD1), a 4000 Hz antialiasing filter (TDT FT3), an attenuator (TDT PA4) and a headphone buffer (TDT HB6). Finally, the stimuli were presented monaurally to the test ear of each listener through a TDH-49 earphone. All stimuli were presented at an equalized RMS level of 78 dB SPL to NH listeners. This 78 dB SPL signal was amplified (through digital filtering) for each individual HI listener, approximating the linear gain prescribed by the NAL-R fitting procedure (Byrne and Dillon, 1986). The average decibel (dB) gains provided to the HI listeners were 11 dB at 500 Hz, 22 dB at 1000 Hz, 26 dB at 2000 Hz and 26 dB at 4000 Hz. The maximum gain was 24 dB at 500 Hz and ≈30 dB at 1000 Hz and above.

3.4. Test procedures

3.4.1. Speech quality ratings

The rating scales used for the quality ratings are the same as those used by Neuman et al. (1998) and are similar to those developed by Gabrielsson et al. (1988, 1990). A 10-point rating scale was used to obtain ratings on five different stimulus attributes: clarity, pleasantness, background noise, loudness and overall impression. Listeners used a written response form to record their ratings (Fig. 3). For each condition, participants listened to a block of 30 of the 256 HINT sentences and then used the 10-point scales to rate the quality of the speech for each of the five attributes. The starting sentence for each block of 30 sentences was randomly selected such that on one block of trials the subject would listen to sentences 45 through 75, on the next block sentences 125 through 155 and so forth. A set of quality ratings consisted of ratings on each of the five attributes in each of the eight conditions. The order of the conditions in each set was randomized. Four sets of quality ratings were obtained. Each set took about 40 min to complete.

3.4.2. Intelligibility: diagnostic rhyme test

On each trial of the DRT, a listener hears one word of a pair and then must choose which of the two words in the pair is heard. Each pair differs in

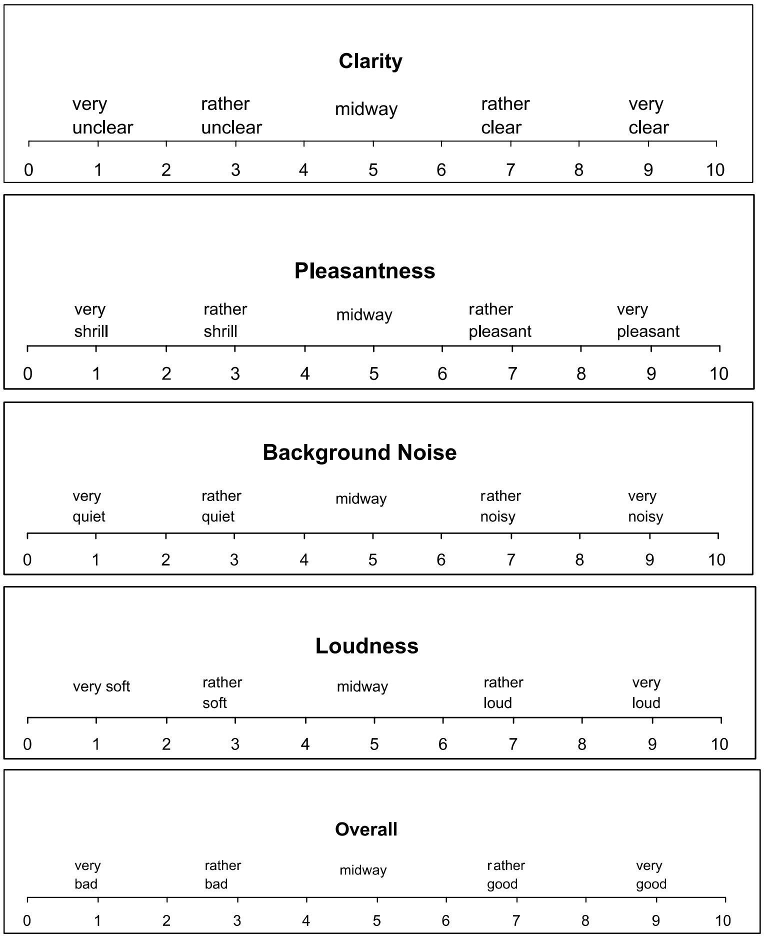


Fig. 3. The rating scales used by listeners in the evaluation of the quality of degraded and enhanced speech (after Neuman et al., 1998).

just one distinctive feature (e.g., veal–feel; dune–tune; pool–tool; bond–pond; pot–tot). Word pairs test distinctive features such as voicing, nasality, sustension, sibilation, graveness, and compactness. Each listener was tested with two full 232-item randomized lists in each of the eight conditions. However, preliminary testing indicated that testing full 232-item lists in each of the eight conditions was too much for one experimental session. Therefore, this testing was carried out in four separate sessions, with a 116-item half-list presented for each condition in each session. The order of the conditions was randomized in each session. Each of eight randomized 232-item full lists was divided up into two 116-item half-lists. The first half-list was presented in a particular condition in the first session; the corresponding second half-list was presented in the same condition in the second session. This was repeated for the third and fourth sessions. The data reported below are based on 464 DRT trials for each listener in each condition, with the overall measure of performance based on the percentage of correct responses.

3.4.3. Intelligibility: nonsense syllable test

The NST (Resnick et al., 1975; Dubno and Dirks, 1982) is a closed-set test in which a listener hears a nonsense syllable and then chooses a response from seven or nine response alternatives.

The test consists of 102 syllables contained in 11 subtests, each of which contains between seven and nine syllables. The subtests differ in terms of voicing and position of consonants as well as the vowel. The order of presentation of the 102 nonsense syllables was randomized on each block of trials. A set of NST blocks consisted of one 102-syllable list in each condition, with the order of the conditions randomized within the set. Listeners each completed three sets of NST (three 1.5 h sessions). The data reported below are based on 306 NST trials for each listener in each condition. The overall measure of performance is the percentage of correctly identified nonsense syllables.

4. Results

4.1. Speech quality ratings

Average ratings for the five attributes of quality for the NH listeners and HI listeners in the degraded and enhanced conditions are shown in Figs. 4 and 5, respectively.

The following steps were taken to insure the validity of the data analysis process used for the quality ratings. We first obtained the arithmetic average of each subject's ratings across the four sessions on each scale for each listening condition. When the data for each condition of each rating scale were examined for differences or trends in the ratings across sessions, no consistent patterns were observed. In addition, analysis of variance procedures using 2×4 mixed design (two groups, NH and HI, by the four sessions) revealed no significant differences in ratings across the sessions. Therefore, the ratings of the four sessions were averaged to serve as the dependent variable in the analyses of variance (ANOVAs) reported below.

To insure that the assumptions of normalcy and homogeneity of these data could be met, the distributions of these averaged session ratings were inspected visually and statistically. Across all conditions the individual mean ratings ranged from 0.25 to 9.75 within the clarity scale; from 0.75 to 9.75 within the pleasant scale; from 0.0 to 8.75 within the noisiness scale; from 3.5 to 7.25 within the loudness scale and from 0.25 to 9.5 within the overall scale. When the sampling distribution of each listening condition for each rating scale was examined, no significant departures from normalcy regarding their characteristics of skewness and kurtosis were observed.

Finally, the reliability of each of the five rating scales was examined by obtaining Pearson product–moment correlations of each individual's session rating with the mean session rating (i.e., four correlations based on 21 pairs of scores) for each

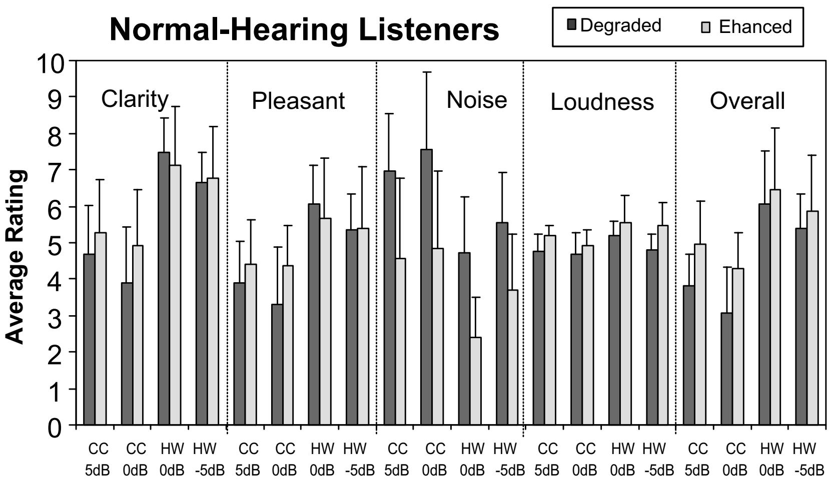


Fig. 4. Average ratings of the NH listeners for the five attributes of quality (clarity, pleasantness, background noise, loudness, and overall impression) for degraded and enhanced speech conditions.

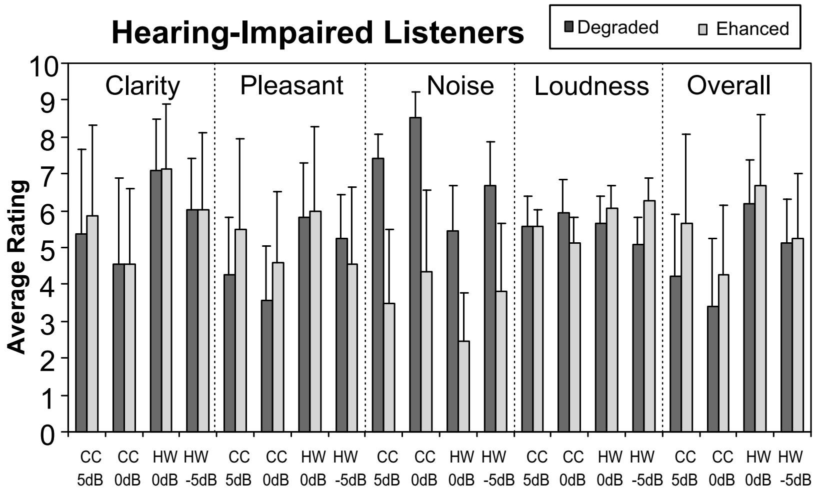


Fig. 5. Average ratings of the HI listeners for the five attributes of quality (clarity, pleasantness, background noise, loudness, and overall impression) for degraded and enhanced speech conditions.

of the eight listening conditions. Theses reliability coefficients ranged from 0.71 to 0.94 with an average of 0.85 for clarity; from 0.66 to 0.96 with an average of 0.85 for pleasant; from 0.32 to 0.96 with an average of 0.81 for noisiness; from 0.42 to 0.91 with an average of 0.66 for loudness and from 0.63 to 0.94 with an average of 0.84 for overall.

A separate repeated measures analysis of variance (ANOVA) was done for each quality attribute for each of the two noise types. (Because the SNRs were not the same for the two noise conditions, separate ANOVAs were carried out for highway noise and for communication channel noise.) A summary of these analyses is shown in Table 2.

Enhancement with the AMT-NS technique resulted in significant benefit in quality ratings on several attributes in both listener groups. On the background noise scale, enhancement resulted in significantly less noisy ratings in both communication channel noise and in highway noise. Enhancement also resulted in significant improvements in

Table 2
Summary of the main effects (group, enhancement, SNR) from the analysis of variance carried out for the five attributes of quality using HINT sentences for each noise type

Source	Loud Comm $F(1,19)$	Loud Hwy $F(1,19)$	Clarity Comm $F(1,19)$	Clarity Hwy $F(1,19)$	Overall Comm $F(1,19)$	Overall Hwy $F(1,19)$	Pleasant Comm $F(1,19)$	Pleasant Hwy $F(1,19)$	Bck-Noise Comm $F(1,19)$	Bck-Noise Hwy $F(1,19)$
Group	14.1***	5.5*	0.3	1.2	0.5	0.2	0.5	0.133	0.0426	2.157
Enh	0.1	25.0***	4.0*	0.03	10.5*	0.9	7.5*	0.2148	58.5***	67***
SNR	2.6	2.6	31.2***	15.25***	61.2***	23.9***	21.3***	20.43***	66.6***	28.1***
Enh × Group	9.7**									
SNR × Group									5.38*	
Enh × SNR	5.6*	8.3**								

F-values are also reported for significant interactions.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

clarity, pleasantness and overall quality for the communication channel noise but not for the highway noise.

No significant differences between groups were noted for ratings on four of the five attributes. In contrast, the loudness ratings for the HI listeners were significantly higher than the loudness ratings for the NH listeners across noise conditions. In highway noise only, both subject groups rated the enhanced speech to be slightly louder than the unenhanced speech. A significant enhancement by group interaction was also evident for the loudness ratings in the communication channel noise condition. Whereas HI listeners on average rated enhanced speech in communication channel noise to be the same or less loud than degraded speech, NH listeners rated enhanced speech to be slightly louder than the unenhanced speech. Finally, increasing SNR had a significant effect on ratings for clarity, pleasantness, background noise and overall quality.

4.2. Intelligibility: DRT

Fig. 6 shows DRT scores (in percent correct) for degraded and enhanced conditions for both NH listeners (left) and HI listeners (right). The DRT percent correct scores were first subjected to an arcsine transform and then submitted to repeated measures ANOVAs. The ANOVA results are shown in Table 3. For the communication channel noise, the factors of group, enhancement and SNR were all statistically significant. For the highway noise, the factors of group and SNR were statistically significant, as was the enhancement by group interaction.

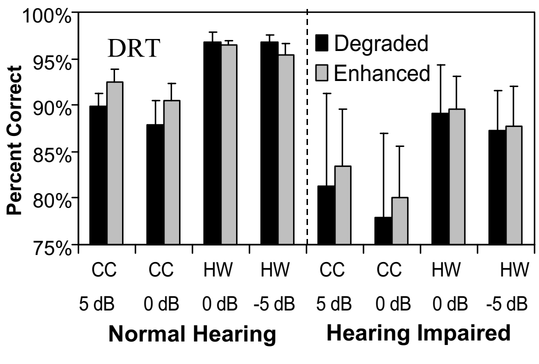


Fig. 6. Intelligibility percent correct scores on the DRT for degraded and enhanced speech conditions in NH listeners (left) and HI listeners (right).

Table 3
Summary of main effects for ANOVA for DRT scores for each noise type (Comm, Hwy), with factors of group, enhancement and SNR

Source	df	*F*(Comm)	*F*(Hwy)
GROUP	1,19	20.7***	49.2***
ENH	1,19	20.4***	4
SNR	1,19	81.7***	23.9***
ENH × GROUP	1,19	2.5	10.2**

Significant interaction effects are also listed.
** $p < 0.01$.
*** $p < 0.001$.

DRT scores were better (10–20% on average) and less variable in the NH listeners than in the HI listeners. Enhancement resulted in small, albeit statistically significant, effects on DRT intelligibility. In the communication channel noise, average DRT scores improved 3% for the NH listeners and 2% for the HI listeners. Consistent with the significant enhancement by group interaction for the highway noise conditions, average DRT scores improved by ≈2% in the HI listeners but decreased by 1–2% in the NH listeners. Finally, DRT intelligibility was also significantly affected by SNR: listeners had lower intelligibility scores in lower SNR conditions in both noise types.

4.3. Intelligibility: NST

Fig. 7 shows the NST scores (in percent correct) for degraded and enhanced conditions for both NH listeners (left) and HI listeners (right). Percent correct scores were first subjected to an arcsine transform and then submitted to a repeated measures ANOVA for each noise type. The ANOVA results are shown in Table 4. For both the communication channel noise and the highway noise, the factors of group, enhancement and SNR were all statistically significant.

Enhancement had a significant positive effect on NST intelligibility in the communication channel noise conditions. NH listeners showed a 7% improvement due to enhancement in the 5 dB SNR

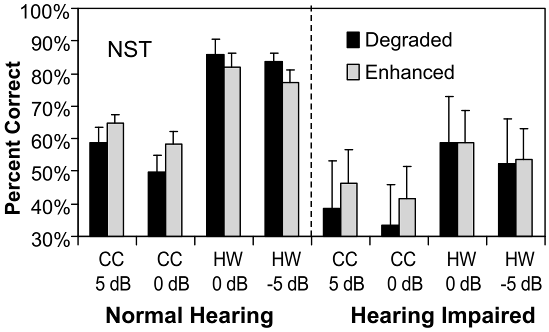


Fig. 7. Intelligibility percent correct scores on the NST scores for degraded and enhanced speech conditions for NH listeners (left) and for HI listeners (right).

Table 4
Summary of main effects for ANOVA for NST scores for each noise type (Comm, Hwy), with factors of group, enhancement and SNR

Source	df	F(Comm)	F(Hwy)
Group	1,19	22.7***	52.8***
ENH	1,19	36.7***	11.8***
SNR	1,19	170.3***	64.5***
ENH × Group	1,19	0.2	14.8***
SNR × Group	1,19	5.5*	0.9

Significant interaction effects are also listed.
* $p < 0.05$.
*** $p < 0.001$.

condition and a 9% improvement due to enhancement in the 0 dB SNR condition. Similarly, HI listeners improved 8% due to enhancement in both the 5 and 0 dB SNR conditions. Both listener groups also had significantly better NST intelligibility scores in the +5 dB SNR condition relative to the 0 dB SNR condition. In fact, the improvement obtained by enhancement in the communication channel 0 dB condition (8–9%) was comparable to the 8% improvement obtained by increasing the SNR by 5 dB (from 0 dB SNR to +5 dB SNR).

As indicated by the significant group by enhancement interaction effect, enhancement in the highway noise condition differed between the two subject groups. Enhancement in the highway noise conditions resulted in little or no improvement in the average NST scores of the HI listeners. In contrast, enhancement resulted in a significant decrease in NST intelligibility in NH listeners in the highway noise conditions: average performance by the NH listeners decreased nearly 4% in the 0 dB SNR condition and 6.5% in the −5 dB SNR condition. Similar to the SNR improvement in communication channel noise, both listener groups had significantly better intelligibility in the 0 dB SNR highway noise condition than in the −5 dB SNR highway condition. NST intelligibility also varied significantly as a function of SNR. Finally, the overall NST scores obtained from the NH listeners were better and less variable than the overall NST scores obtained from the HI listeners.

5. Discussion

A primary goal of this study was to evaluate the effectiveness of the AMT-NS technique in improving the quality and intelligibility of speech in native English-speaking listeners. In both NH and HI listeners, the AMT-NS algorithm resulted in significantly better ratings of noisiness ("less noisy") in both communication channel noise and highway noise. The algorithm also resulted in significantly better ratings of pleasantness, clarity and overall quality in communication channel noise. These improved quality ratings are consistent with other extended spectral subtraction techniques showing improved quality ratings of the processed speech (e.g., Eberling et al., 1993; Levitt et al., 1993; Jamieson et al., 1995).

Due in part to the musical noise artifacts caused by spectral subtraction techniques, previous studies have shown that spectral subtraction approaches to enhancement may result in improved quality but not improved intelligibility. The AMT-NS technique evaluated in this study yielded improvements in intelligibility in some but not all of the conditions tested. The statistically significant improvements due to enhancement observed in our NH listeners in the communication channel noise conditions were 9% for the NST and 3% for the DRT. However, these same NH listeners showed a significant loss of intelligibility due to enhancement for the NST (4–7%) and for the DRT (1–2%) in the highway noise conditions.

If we compare the results of our NH listeners to the results reported by Tsoukalas et al. (1997), we see similar trends, but some interesting differences. The improvement in intelligibility scores for NH listeners on their English-based DRT was 30%, 5%, 5%, respectively, at SNRs of −5, 0, +5 dB, using speech-spectrum shaped broadband noise. The improvement in intelligibility for their listeners on their Greek-version of the DRT was 14%, 3.6%, 10%, respectively, at the −5, 0, +5 dB SNRs. Generally, these improvements are larger than we observed in our study.

The smaller effects observed in our study may be due in part to possible ceiling effects in the DRT intelligibility tests. The DRT intelligibility scores for our NH listeners were already very high (as high as 97%) in the degraded noise conditions, such that there was very limited room for observing any improvements provided by the AMT-NS enhancement.

Another important difference between the DRT results reported by Tsoukalas et al. (1997) and those presented here is the frequency range. They used an 8 kHz bandwidth while our bandwidth was 4 kHz. This bandwidth, we feel, was important to consider for both HI individuals as well as for issues in voice telephony applications for NH and HI subjects. Allowing listeners to evaluate the algorithm with the additional 4000–8000 Hz frequency band would perhaps provide the listeners with additional information content that generally would not be available in voice communication systems, or accessible to HI listeners with high frequency sloping loss.

Another goal of this study was to determine whether the effectiveness of the algorithm in NH and HI listeners varies as a function of noise type, speech type and SNR. Results indicate that the AMT-NS algorithm is effective in the communication channel noise but not in the highway noise. The communication channel noise has a relatively power spectrum across the 4 kHz bandwidth and is stationary. The automobile highway noise is primarily low frequency (i.e., below 800 Hz), and is slowly varying. The broadband communication channel noise will have a more profound impact on higher frequency speech sounds than will the highway noise. Therefore, in the communication channel noise conditions, the enhancement algorithm has a greater potential to suppress the noise content. The loss of intelligibility observed in the highway noise condition could be attributed to processing artifacts that could occur in the high frequency portion if the AMT is not perfectly estimated there, since little noise is present at the high frequencies.

To explore further the effectiveness of the algorithm in the communication channel noise, we examined the effectiveness of the algorithm in terms of specific features of the DRT and NST tests. Fig. 8 shows the DRT scores for both listeners groups in communication channel noise analyzed in terms of percent scores on six distinctive features (Voiers, 1983): (1) nasality, (2) voicing, (3) graveness (a place feature which contrasts front-mouth constrictions with other constrictions), (4) compactness (a place feature which contrasts back-mouth constrictions with other constrictions), (5) sustention (a manner feature

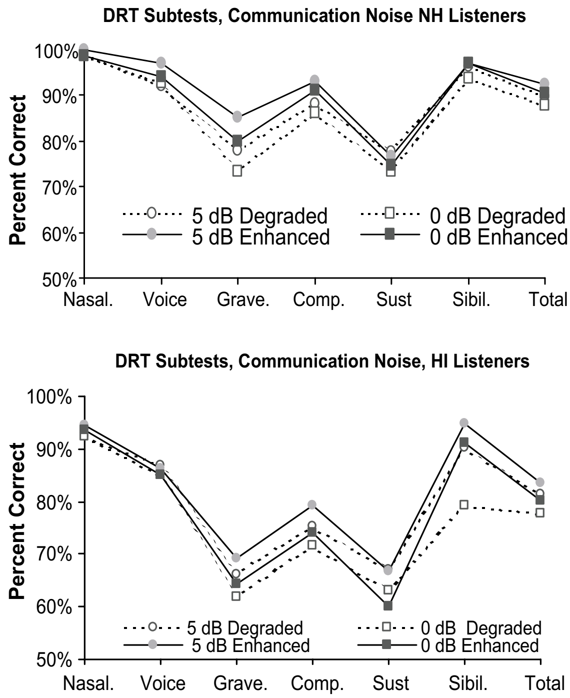


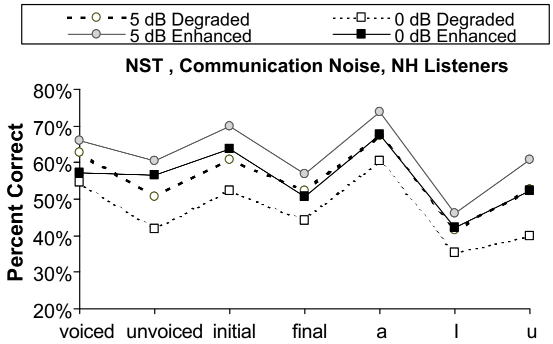
Fig. 8. Percent correct scores on the DRT in the communication channel noise conditions analyzed in terms of six distinctive features (Voiers, 1983) for NH and HI listeners.

which contrasts continuous and interrupted) and (6) sibilation (a manner feature which contrasts fricatives and stops). Frequency importance functions (Duggirala et al., 1988) show the following crossover (peak) frequencies for each of the six features: nasality 472 Hz; voicing 758 Hz; graveness 1290 Hz; compactness 1618 Hz; sustention 1800 Hz; sibilation 2521 Hz.

The largest differences between scores for enhanced and degraded conditions are observed for features of voicing (NH group), graveness (NH and HI group), compactness (NH and HI groups) and sibilance (HI group). Post hoc ANOVA comparisons (group, enhancement, SNR) on individual features indicate significant improvements in enhanced versus unenhanced conditions for the features of voicing, graveness and compactness ($p < 0.01$). (However, we exercise caution in interpreting the significance of a large number of post hoc multiple comparisons due to the increased probability of a Type I statistical errors.)

Fig. 9 shows the NST scores for the communication channel noise conditions analyzed in terms of percent correct scores for stimulus items in which the listener is asked to distinguish among voiced consonants, unvoiced consonants, initial consonants, final consonants and the three different vowels contained in the NST test. A large improvement due to enhancement (13–14%) is evident for unvoiced tokens but not for voiced tokens (1%). This differential benefit may be due in part to the lower effective SNR of the unvoiced consonants and may be especially important for HI listeners who have increased difficulty perceiving unvoiced sounds.

A final goal of this study was to determine potential hearing aid applications of the algorithm by studying the relative effectiveness of the algorithm in both NH and HI listeners. Statistically significant between-group differences in the intelligibility scores are explained by the poorer overall performance and greater variability on the DRT and NST shown by the HI listeners. This reduced overall performance may be due in part to the reduced frequency resolution characteristic of cochlear hearing loss (Moore, 1998). Reduced frequency resolution is also observed at higher presentation levels. Therefore, the higher absolute



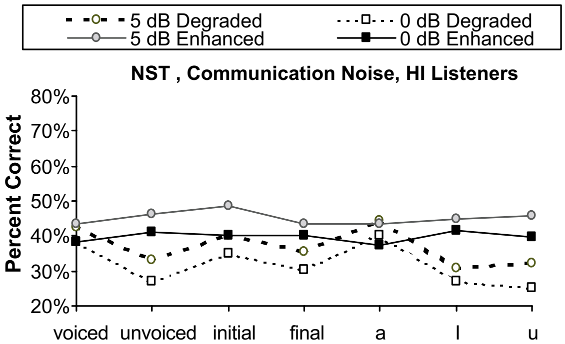


Fig. 9. Percent correct scores on the NST in the communication channel noise conditions for the NH and HI listeners, analyzed in terms of the following distinctions: voiced consonants, unvoiced consonants, initial consonants, final consonants and three different vowel contexts.

Sound Pressure Levels (due to NAL-R gain of up to ≈30 dB, depending on the severity of the hearing loss), may also contribute to the reduced intelligibility scores of the HI listeners.

The algorithm effectiveness for HI listeners might be improved by incorporating alternative processing strategies. For example, AMT parameter estimation might be modified in order to increase the effectiveness of the AMT algorithm in the impaired listeners by considering the degraded frequency selectivity and increased susceptibility to masking that characterizes cochlear hearing loss (e.g., Moore, 1998).

6. Summary

Enhancement with the AMT-NS technique resulted in significant improvement in quality ratings on several attributes in both listener groups. On the background noise scale, enhancement resulted in significantly lower ratings of noisiness in both communication channel noise and in highway noise. Enhancement also resulted in significant improvements in clarity, pleasantness, and overall quality for the communication channel noise but not for the highway noise. In NH listeners, enhancement significantly improved intelligibility in the communication channel noise condition but resulted in a loss in intelligibility in the highway noise condition. Enhancement resulted in significant benefit to speech intelligibility by listeners with hearing loss for communication channel noise but not for highway noise. This differential effect of the algorithm based on the noise type suggests that the AMT-NS processing that occurs in the mid-higher frequency region is beneficial when noise is actually present in this region but is deleterious when noise is not present in this frequency region. Similar to NH listeners, HI listeners showed the most pronounced benefits in the communication channel noise condition with the NST listening tests. Alternative processing strategies that specifically consider the effects of cochlear hearing loss may optimize the AMT-NS approach to speech enhancement for HI listeners.

Acknowledgements

This research was supported by the Whitaker Foundation, National Organization of Hearing Research, the University of Colorado Council on Research and Creative Work and the University of Colorado Undergraduate Research Opportunity Program. This work was also supported in part by a research grant from the US Navy—SPAWAR Systems division. Portions of these results were reported at the Acoustical Society of America, Chicago, Illinois, May 2001. The authors extend thanks to Jennifer Schryer for assistance in data collection.

References

American National Standards Institute (ANSI), 1989. Specification for audiometers. ANSI S3.6. New York: ANSI.

Baer, T., Moore, B.C.J., 1994. Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. J. Acoust. Soc. Amer. 95, 2270–2280.

Boll, S.F., 1979. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Process. 39, 795–805.

Byrne, D., Dillon, H., 1986. The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. Ear Hear. 7, 257–265.

Cappa, O., 1994. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor. IEEE Trans. Speech Audio Process. 2, 345–349.

Clarkson, P., Bahgat, S., 1989. Envelope expansion methods for speech enhancement. J. Acoust. Soc. Amer. 89, 1378–1382.

Dubno, J.R., Dirks, D.D., 1982. Evaluation of hearing-impaired listeners using a nonsense-syllable test. I. Test reliability. J. Speech Hear. Res. 25, 135–141.

Duggirala, V., Studebaker, G.A., Pavlovic, C.V., Sherbecoe, R.L., 1988. Frequency importance functions for a feature recognition test material. J. Acoust. Soc. Amer. 83, 2372–2382.

Duquesnoy, A.J., 1983. Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. J. Acoust. Soc. Amer. 74, 739–743.

Elberling, C., Ludvigsen, C., Keidser, G., 1993. The design and testing of a noise reduction algorithm based on spectral subtraction. Scandinavian Audiol. (Suppl. 38), 39–49.

Festen, J.M., Plomp, R., 1983. Relations between auditory functions in impaired hearing. J. Acoust. Soc. Amer. 73, 652–662.

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. Amer. 88, 1725–1736.

Gabrielsson, A., Hagerman, B., Bech-Kristensen, T., Lundberg, G., 1990. Perceived sound quality of reproductions with different frequency and sound levels. J. Acoust. Soc. Amer. 88, 1359–1366.

Gabrielsson, A., Schenkman, B.N., Hagerman, B., 1988. The effects of different frequency responses and sound levels. J. Speech Hear. Res. 31, 166–177.

Glasberg, B.R., Moore, B.C.J., 1989. Psychoacoustic abilities of subjects with unilateral and bilateral cochlear impairments and their relationship to the ability to understand speech. Scandinavian Audiol. (Suppl. 32), 1–25.

Hansen, J.H.L, Arslan, L., 1995. Robust feature estimation and objective quality assessment for noisy speech recognition using the credit card corpus. IEEE Trans. Speech Audio Process. 3, 169–184.

Hansen, J.H.L., Clements, M.A., 1991. Constrained iterative speech enhancement with application to speech recognition. IEEE Trans. Signal Process. 39, 795–805.

Hansen, J.H.L., Nandkumar, S., 1995. Robust estimation of speech in noisy backgrounds based on aspects of the auditory process. J. Acoust. Soc. Amer. 97, 3833–3849.

Hansen, J.H.L., 1999. Speech enhancement. Chapter in Encyclopedia of Electrical and Electronics Engineering, Vol. 20. John Wiley & Sons. pp. 159–175.

Hou, Z., Pavlovic, C.V., 1994. Effects of temporal smearing on temporal resolution, frequency selectivity, and speech intelligibility. J. Acoust. Soc. Amer. 96, 1325–1340.

Humes, L.E., Dirks, D.D., Kincaid, G.E., 1987. Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal listeners. J. Acoust. Soc. Amer. 81, 765–773.

Hygge, S., Ronnberg, J., Larsby, B., Arlinger, S., 1992. Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech and noise backgrounds. J. Speech Hear. Res. 35, 208–215.

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, 274–286.

Johnston, J.D., 1988. Transform coding of audio signals using perceptual noise criteria. IEEE J. Select. Areas Comm. 6, 314–323.

Larson, V.D., Williams, D.W., Henderson, W.G., Luethke, L.E., Beck, L.B., Noffsinger, D., Wilson, R.H., Dobie, R.A., Haskell, G.B., Bratt, G.W., Shanks, J.E., Stelmachowicz, P., Studebaker, G.A., Boysen, A.E., Donahue, A., Canalis, R., Fausti, S.A., Rappaport, B.Z., 2000. Efficacy of 3 commonly used hearing aid circuits—a crossover trial. J. Amer. Med. Assoc. 284, 1806–1813.

Levitt, H., Bakke, M., Kates, J., Neuman, A., Schwander, T., Weiss, M., 1993. Signal processing for hearing impairment. Scandinavian Audiol. (Suppl. 93), 7–19.

Moore, B.C.J., 1998. Cochlear Hearing Loss. Whurr Publishers, London.

Moore, B.C.J., Glasberg, B.R., Vickers, D.A., 1995. Simulation of the effects of loudness recruitment on the intelligibility of speech in noise. Brit. J. Audiol. 29, 131–143.

Mueller, G., 2000. Patient benefit: What's the digital difference? Hear. J. 53, 27–32.

Nandkumar, S., Hansen, J.H.L., 1995. Dual-channel iterative speech enhancement with constraints based on an auditory spectrum. IEEE Trans. Speech Audio Process. 3, 22–34.

Neuman, A.C., Bakke, M.H., Mackersie, C., Hellman, S., Levitt, H., 1998. The effect of compression ratio and release time on the categorical rating of sound quality. J. Acoust. Soc. Amer. 103, 2273–2281.

Nilsson, M., Soli, S.D., Sullivan, J.A., 1994. Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. J. Acoust. Soc. Amer. 95, 1085–1099.

Plomp, R., 1986. A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. J. Speech Hear. Res. 29, 146–154.

Resnick, S.B., Dubno, J.R., Hoffnung, S., Levitt, H., 1975. Phoneme errors on a nonsense syllable test. J. Acoust. Soc. Amer. 58 (Suppl. 1), 114.

Schroeder, M.R., Atal, B.S., Hall, J.L., 1979. Optimizing digital speech coders by exploiting masking properties of the human ear. J. Acoust. Soc. Amer. 66, 1647–1652.

Tsoukalas, D.E., Mourjopoulos, J.N., Kokkinakis, G., 1997. Speech enhancement based on audible noise suppression. IEEE Trans. Speech Audio Process. 5 (6), 497–513.

Van Rooj, J.C., Plomp, R., 1990. Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses. J. Acoust. Soc. Am. 88, 2611–2624.

Virag, N., 1999. Single channel speech enhancement based on masking properties of the human auditory system. IEEE Trans. Speech Audio Process. 7, 1–382.

Voiers, W.D., 1983. Evaluating processed speech using the diagnostic rhyme test. Speech Technol. 1, 30–39.

Wang, D.L., Lim, J.S., 1982. The unimportance of phase in speech enhancement. IEEE Trans. Acoust. Speech Signal Process. 30 (4), 679–681.

Zurek, P.M., Delhourne, L.A., 1987. Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment. J. Acoust. Soc. Amer. 82, 1548–1559.