Denoising Speech Signals for Digital Hearing Aids: A Wavelet Based Approach

Nathaniel Whitmal, Janet Rutledge and Jonathan Cohen

Abstract This study describes research developing a wavelet based, single microphone noise reduction algorithm for use in digital hearing aids. The approach reduces noise by expanding the observed speech in a series of implicitly filtered, shift-invariant wavelet packet basis vectors. The implicit filtering operation allows the method to reduce correlated noise while retaining low-level high-frequency spectral components that are necessary for intelligible speech. Recordings of speech in automobile road noise at signal to noise ratios of 0, 5, 10, 15 and 20 dB were used to evaluate the new method. Objective measurements indicate that the new method provides better noise reduction and lower signal distortion than previous wavelet-based methods, and produces output free from audible artifacts of conventional FFT methods. However, trials of the Revised Speech Perception in Noise test with the new algorithm showed no significant improvement in speech perception. Subsequent analysis has shown that the algorithm imposes physical attenuation on low-intensity components that mimics the perceptual effects of mild hearing loss.

1 Introduction

Unwanted acoustic noise is present in a variety of listening environments. Common examples include the background of conversational babble found in a cocktail

Nathaniel Whitmal
Department of Communication Disorders, University of Massachusetts, Amherst, Amherst, Massachusetts, e-mail: nwhitmal@comdis.umass.edu

Janet Rutledge
Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore, Maryland, e-mail: jrutledge@umbc.edu

Jonathan Cohen
Department of Mathematics, DePaul University, Chicago, Illinois,
e-mail: jcohen@math.depaul.edu

party or a crowded restaurant; the broad spectrum noise produced by a loud air-conditioner or a jet engine at an airport, and the road noise heard in a car during highway driving. These noises can impair speech communication by "masking" (reducing the audibility of) nearby speech signals. Masking is particularly troublesome for hearing-impaired listeners, who have greater difficulty understanding speech in noise than normal-hearing listeners. Our research applied the mathematical theory of wavelet denoising as developed by Coifman and Saito [15], [60] and Donoho and Johnston [18],[19], to reducing unwanted noise for users of digital hearing aids. Specifically, we evaluated the utility of wavelets and wavelet packets in algorithms designed to remove the noise from noisy speech signals.

The final version of our algorithm was the result of a series of successive experiments in which the algorithm was adjusted to overcome a number of problems, including; the selection of a basis that best discriminated between noise and speech; choosing an optimal number of coefficients from which to reconstruct the denoised signal; removing the artificial sounds introduced by the processing; creating an algorithm robust enough to be able to respond to different types of noise.

The main body of the paper begins with a description of the original experiments that were performed on a set of nine noisy speech files. The next section is a detailed description of the algorithm we used to remove background noise. The following two sections are performance evaluations of the algorithm. Section 3 is devoted to objective measures of the performance of the algorithm. Using spectrograms and signal to noise ratios (SNR), we show improvements in our algorithm as compared to other forms of processing. Section 4 reports the results of testing the algorithm on a small group of hearing impaired subjects. These results produced intelligibility scores comparable to the algorithms with which it was compared. Section 5 uses the results of subsequent experiments to show that the hard thresholding used in our algorithm produces distortions similar to those produced by algorithms that simulate recruitment of loudness for non-impaired listeners. This partially explains how the algorithm produced improvements in objective measures without a corresponding improvement in intelligibility. In the final section, we discuss some of the lessons learned from our research.

Before proceeding to the description of our work, it is useful to review some of the other algorithms that have been tried to improve intelligibility. We also point out some of the difficulties that arise that are specific to the denoising of speech.

Poor intelligibility in background noise has been a long-standing problem for hearing aid users. Until about twenty-five years ago, intelligibility problems were related to the power and fidelity limitations of hearing aids themselves. A review of these constraints is given by Killion [39]. Subsequent improvements in the size, power consumption, and capability of digital signal processors led to the adoption of digital signal processing (DSP) technology in hearing aids. The first commercial digitally programmable hearing aids were introduced in the late 1980s, and followed in 1996 by two all-digital hearing aids: the Widex Senso and the Oticon Digifocus (Levitt [43]). Both digital aids had an immediate effect on the hearing aid market. A survey taken in 1997 revealed that programmable / digital aids already accounted for 13% of the market; this share increased to 28% by 2000 and to 51%

by 2005 (Kochkin [42]). Nevertheless, surveys taken over the period of our research (Kochkin [40]; Kochkin [41]) indicated that only 25-30% of hearing aid users remained satisfied with the performance of their hearing aids in noisy environments.

In anticipation of all-digital hearing aids, several 1980s-era research groups attempted to use DSP algorithms to enhance noisy speech for hearing aid users. These algorithms may generally be divided into two groups: multiple-microphone approaches and single microphone approaches.

Multiple-microphone approaches (Brey et. al. [8]; Chabries et al. [10]; Schwander and Levitt [62]; Peterson et al. [54]; Greenberg and Zurek [33]), exploited the correlation between noise signals from spatially separated microphones to enhance noisy speech. While this approach was shown to improve intelligibility, the microphone spacings used in those studies rendered their algorithms impractical for commercial hearing aids.

The single microphone approaches of (Lim and Oppenheim [47]; Boll [4]; McAulay and Malpass [51]; Porter and Boll [55]; Ephraim and Malah [25]; Graupe et al.[30]; Ephraim et al. [26]; Kates [38]), relied on simple probabilistic models of speech and noise. These approaches were considered appropriate for digital hearing aids and other portable communication systems (e.g., cellular phones) for which spatially separated inputs were impractical. These noise reduction algorithms generally attempted to estimate temporal or spectral features of the original signal from observations of the noise-corrupted signal. These approaches met with mixed success, particularly at low SNRs, where the algorithms attenuated or distorted consonants which contribute to intelligibility. Numerous evaluations using both objective (Lim [46]; Makhoul and McAulay [49]; Boll [5]) and subjective (Tyler and Kuk [67]; Sammeth and Ochs [61]; Levitt et al.[45]) criteria indicated that these single-microphone algorithms failed to produce consistent improvements in intelligibility.

A promising alternative class of single-microphone noise reduction methods used the subspace approach, which projects segments of the noisy speech signal onto orthogonal "speech" and "noise" subspaces and discards the "noise" component. The speech subspace is constructed from high-energy vectors in the segment's principal-component (or Karhunen-Loeve) basis. One type of subspace algorithm was evaluated in a series of papers by Ephraim and Van Trees [27] and [28] and Ephraim et al [29]. By using long segments of speech, Ephraim and Van Trees were able to use a discrete Fourier transform (DFT) to approximate principal-component bases in a computationally efficient manner. Their objective test results [28] showed that their approach eliminates the "musical noise" artifacts produced by earlier methods like spectral subtraction. Unfortunately, their subjective test results [29] indicate that (like earlier approaches), the subspace approach improves perceived speech quality without improving speech intelligibility.

It's possible that their algorithm's performance may be affected by the poor time localization of DFT basis functions, which can limit the function's usefulness in low order models of transient events (e.g. plosive consonants). Another factor may be the spectral and temporal properties of the 44 English phonemes (i.e. speech sounds). Vowels have a predictable quasi-periodic spectrum and may compress easily in a Fourier basis. Fricatives on the other hand, such as /s/, /sh/ or /f/, do not localize

as well in frequency and are almost indistinguishable from noise; therefore, the algorithm tends to delete them along with the background noise. The balance of phonemes, that is the various families of consonants, are somewhere in between and are only partially removed by denoising. Vowels are higher in amplitude than consonants, so background noise tends to mask consonants to a greater extent. This situation is compounded by the further auditory degradation resulting from hearing loss.

The orthogonal basis functions produced by the discrete wavelet transform (Mallat [50]) overcome some limitations of the DFT basis, providing good spectral localization at low frequencies and good temporal resolution at high frequencies. Projection of a signal onto these basis functions may be accomplished efficiently by passing the signal through a tree-structured conjugate quadrature filter bank (Smith and Barnwell [65]). More flexible time-frequency localization may be obtained with the wavelet-packet transform (WPT) of Coifman and Wickerhauser [16], which allows the signal to be decomposed into subbands providing appropriate temporal and spectral resolution for a given application. Both the wavelet and wavelet-packet transforms demonstrate compression capabilities that rival those of principal component bases for many types of real world signals (Wornell [76]; Donoho [18]; Sinha and Tewfik [64]). This combination of properties motivated several researchers to implement subspace denoising with wavelet transforms. Evaluation of subspace denoising using shift-variant and shift-invariant versions of the discrete wavelet (Donoho and Johnstone [19]; Donoho et al. [20]; Coifman and Donoho [12]; Lang et al [44]) and wavelet packet transforms (Coifman and Majid [14]; Saito [60]; Berger et al [2];Pesquet et. al. [53]) demonstrated improvements in signal-to-noise ratio (SNR) ranging between 5 and 10 dB for signals with initial SNRs between 0 and 20 dB. Shift invariant algorithms (which implicitly average together several enhanced versions of the signal) provided additional improvements of as much as 7 dB at the cost of some computational efficiency. Of the methods described above, only one (Berger et al. [2]) was evaluated with speech; for this method, no objective or subjective data were reported.

The objective of our study was to evaluate wavelet-based subspace noise reduction for use in digital hearing aid applications. Results of a preliminary study are described; these are followed by specification and development of a new algorithm evaluated with both objective measures and listening results from trials of the Revised Speech Perception in Noise (R-SPIN) test (Bilger et al. [3]) with both normal acuity and impaired acuity listeners.

1.1 The first algorithm

The first experiments were conducted with a software package that used an entropy criteria to find the coefficients for a wavelet packet best basis. The entropy criteria was chosen to pick the basis that concentrated the largest amount of energy in the fewest number of coefficients. With the coefficients placed in decreasing order by

size, the best basis was the one that began with the steepest negative slope. Before applying the software to the signal, the experimenter was prompted for a percentage of coefficients that were to be retained in the reconstruction of the signal. The selection process then started with the largest coefficient and proceeded in descending order until the prescribed percentage of coefficients was reached.

Letting S denote the original signal, we let C_1, the "coherent" signal, denote the signal reconstructed from the selected largest coefficients. The "noise" signal N_1 from the remaining coefficients associated with this selection of basis elements is the residual; that is,

$$S = C_1 + N_1.$$

The algorithm was then applied to the residual N_1 which resulted in the decomposition of N_1 into coherent and noise parts

$$N_1 = C_2 + N_2$$

Repeating the process yielded a sequence of Cs and Ns that was terminated as soon as the noise part was essentially free of any speech signal. The speech signal is then represented as the sum

$$C_1 + \cdots + C_n$$

of the coherent parts, all of which have presumably been separated from the noise.

The idea is that the speech is coherent and is concentrated in a few coefficients and the noise spreads out evenly among many coefficients. By repeating the procedure, the primary and secondary coherent parts of the signal can be peeled off, leaving the sequence of $N_i s$ containing decreasing amounts of the speech part of the original signal. With this version of the software, the experimenter could adjust the percentage of the coefficients to keep at each stage of the algorithm.

1.2 The original 9 noisy sentences

The original experiments were conducted at a computer lab in the Northwestern University Department of Electrical Engineering and Computer Science and the Department of Communication Sciences and Disorders. There were attempts to improve the intelligibility of 9 sentences chosen from the Harvard / IEEE sentence corpus [35], that had been recorded and corrupted by the addition of white noise.

The 9 phonetically balanced sentences were designed to minimize the ability of the listener to figure out the words from the context of the sentences. The noisy speech files were created by adding a file of randomly generated samples from a normal distribution to the original speech file. The signal to noise ratios of the sentences were sufficiently low that it was difficult for listeners with normal hearing to recognize any of the words, particularly on the first time hearing them. On the other hand, the noise level was sufficiently low that the words were readily recognizable to someone who knew the sentences beforehand. The nine sentences were:

1. *The pipe began to rust while new*
2. *Thieves who rob friends deserve jail.*
3. *Add the sum to the product of these three.*
4. *Open the crate but don't break the glass.*
5. *Oak is strong and also gives shade.*
6. *Cats and dogs each hate the other.*
7. *That hose can wash her feet.*
8. *Go and sit on the bed.*
9. *I read the news today.*

As an indication of how well disguised the sentences were, out of a group of several listeners to the first noisy sentence, only one guessed that the second word was pipe. Not surprisingly, the one who correctly identified pipe was the only native English speaker.

The nine sentences were originally unknown to the person conducting the experiments. Of the nine sentences, it was only for the last one that the experimenter was able to identify the exact sentence after repeated applications of the algorithm. And even in this case it is difficult to say whether this success was due to the improved intelligibility or simply to the repeated hearing of the noisy sentence.

A more sophisticated version of the software was obtained from the group at Yale that enabled the user to specify the number of levels (iterations of the signal decomposition into coherent and noise parts) to run the algorithm and also gave some choice of wavelet to use. This provided additional flexibility and efficiency. But it also gave up some flexibility as it required the user to specify the same percent of coefficients to keep at each level of the algorithm.

There were several observations that stand out from the original experiments.

- All ways of using the algorithm in all cases reduced the amount of background noise.
- With too few coefficients, all that was heard were incoherent musical tones.
- As more coefficients were used, the sounds became more coherent but a growing amount of background noise crept in.
- Beyond a certain point, the addition of more coefficients, contributed more and more to the noise volume while adding less and less to the speech volume.
- The gain from the extra levels was minimal. From the subjective point of view of the listener, there was little change after the first level and no improvement was noticeable after the second level.
- Prior knowledge of the sentence made a huge difference in the perception of the listener.
- The vowels were easy to identify and the fricatives disappeared.

1.3 The Wavelet Packet for Windows

The Wavelet Packet for windows, developed at Yale, allowed for a great deal more flexibility. It allowed for frame sizes of up to 4096 samples; it offered up a wide collection of wavelets with some giving closer approximations of the original signal than others. It included software for the use of local cosines; it allowed for a number of ways to choose coefficients; the windows environment was a lot more user friendly.

With the increased flexibility came some downside. The 4096 sample limit meant that to reconstruct an entire sentence with as many as 20,000 samples, several smaller intervals of speech had to be concatenated. This added a great deal of time to the process of experimenting.

There were some useful observations from the WPWL.

- The wavelets that came from scaling functions with longer recurrence relationships gave better results.
- For a given processing of the signal, the WPWL gave a good way of finding the approximately optimal choice of the number of coefficients, suggesting that algorithms for doing this could be built into a denoising algorithm.
- It was necessary for us to write our own programs to have the flexibility to deal with all the complications that would arise in developing an algorithm to meet the challenges posed by the problem of noisy sound signals.

2 Theory

In this section, our wavelet-based approach to noise reduction is described. This approach assumes that the noise is autoregressive (AR) and Gaussian, and uses the WPT of the noise's whitening filter to construct a second set of orthogonal basis vectors which are better suited for noise reduction. In this sense, the approach generalizes previous wavelet-based methods which assume the presence of white Gaussian noise (Saito [60]; Coifman and Donoho [12]; Lang et al [44]; Pesquet et al. [53]).

2.1 The Wavelet Packet Transform

A typical implementation of the wavelet packet transform is shown in Figure 1. There, a signal $\mathbf{x} = [x_0, x_1, ...x_{N-1}]$ (assumed periodic with period N) is input to a tree-structured bank of filter/decimator operations which divide the signal (in an approximate sense) into progressively narrower frequencey sub-bands. The filters $\widetilde{H} = \{h_{-n}\}_{m=0}^{M-1}$ shown in Figure 1 are time reversed versions of the perfect-reconstruction conjugate quadrature filters derived by Daubechies [17]. The outputs of each sub-

band's filter/decimator correspond to expansion coefficients $c_{j,k,m}$ for specific basis functions $\phi_{j,k,m}$, where $\{j,k,m\}$ approximately specify the time/frequency localization, frequency sub-band location, and temporal origin of the function. Basis functions for larger j tend, in an approximate sense, to be more precisely localized in the frequency domain and less precisely localized in the time domain. The transform coefficients for particular values of j are given as

$$c_{j,k,m} = \begin{cases} \sum_{n=0}^{M-1} h_n c_{j-1,k/2,n+2m} & \text{k even} \\ \sum_{n=0}^{M-1} g_n c_{j-1,(k-1)/2,n+2m} & \text{k odd} \end{cases} \quad (1)$$

where $j \in \{0,1,...,\log_2 N\}, k \in \{0,1,...,2^j-1\}, m \in \{0,1,...,2^{-j}N-1\}$, and $c_{0,0,m} = x_m$. Since the perfect-reconstruction property of the filters forces equation 1 to be a unitary transformation, the basis functions at any scale $j = J \geq 0$ can be constructed from linear combinations of basis functions at scales $j > J$. Hence, the collection of basis functions at all scales is redundant, and any collection of N basis functions corresponding to a disjoint covering of nodes of the tree provides a complete orthonormal basis for the signal.

The decimators used in Figure 1 cause the WPT to be shift-invariant, such that applying a time-shift to $\mathbf{x}$ can produce substantial changes in coefficient values. This shortcoming has led to the use of a shift-invariant WPT (or SIWPT), which implements the structure of Figure 1 in a pre-selected basis without the decimators. The outputs of each filter then become the expansion coefficients in the pre-selected wavelet-packet basis for all circularly shifted versions of the input signal. The inverse SIWPT recombines these coefficients in a manner that effectively reshifts the shifted signals to their origins and averages them together. More detailed discussions of shift-invariant transforms and their use in noise reduction are provided by Coifman and Donoho [12] and Pesquet et al. [53].

2.2 Selection of wavelet packet bases

The speech estimates produced by subspace algorithms are, essentially, truncated series representations of the original noisy signal. It is well known that, of all orthonormal basis representations of signals in white noise, principal component bases produce the truncated series representations with the smallest mean-squared-error, and best compress the signal energy into the fewest number of coefficients. Watanabe [72] has shown that achieving these properties is equivalent to minimizing the "entropy" function

$$H(\{p_i\}) = -\sum_{i=0}^{N-1} p_i \log_2 p_i \quad (2)$$

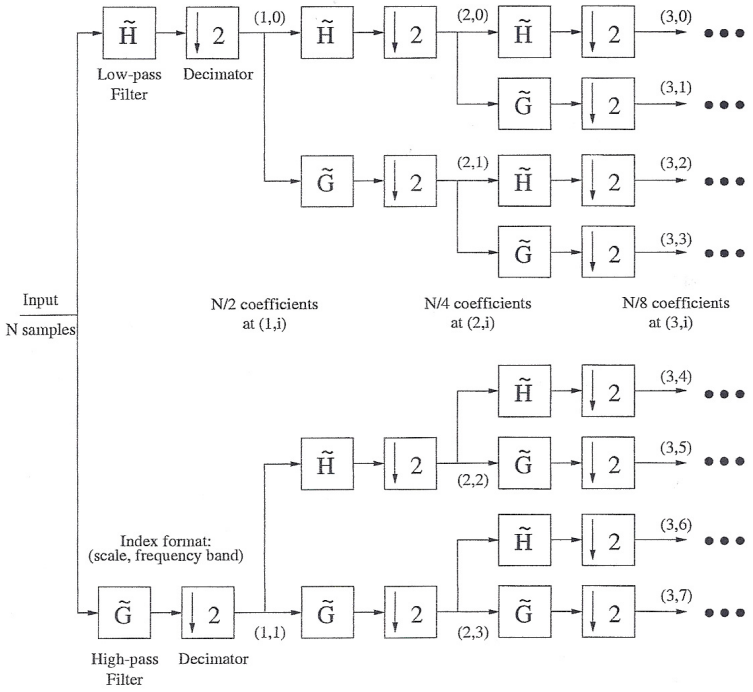

Fig. 1 Implementation of the wavelet packtet transform

where the "probability" p_i of the transform coefficient c_i is defined by

$$p_i = \frac{|c_i|^2}{\sum_{i=0}^{N-1} |c_i|^2}. \tag{3}$$

Wavelet packets, unlike principal-components, provide a multiplicity of bases from which to construct a "signal" subspace. Of these bases, the minimum-entropy basis provides the best signal compression. In previous studies assuming the presence of white noise, this basis was selected by means of the "best-basis" algorithm of Coifman and Wickerhauser [16], which used dynamic programming to select the wavelet-packet basis with minimum entropy and optimum signal compression. For the case of correlated noise, a generalized "best-basis" algorithm (Coifman and Saito [15]) is used which maximizes the "discrimination" function

$$H(\{p_i, q_i\}) = \sum_{i=0}^{N-1} p_i \log_2(p_i/q_i) \tag{4}$$

where $\{p_i\}$ and $\{q_i\}$ respectively represent the "probabilities" of the noise-corrupted speech and the noise alone. This new criterion (which reduces to the entropy criterion in the case of white noise) provides a "local discrimination basis" (LDB) which gives the best possible separation between the "signal"and "noise" components of **x**. The reader is referred to Coifman and Saito [15] for a more detailed explanation of this approach.

2.3 Separation of signal and noise subspaces

After the optimum basis is selected for denoising, the basis functions must be separated into "signal" and "noise" subspaces. Previous studies have compared transform coefficients with a threshold to determine in which subspace their respective basis vector resides. Here, as in two previous studies (Saito [60]; Pesquet et al. [53]), the Minimum Description Length (MDL) model selection criterion of Rissanen [58] is used to drive the thresholds.

In the MDL framework for noise reduction, a noisy observation $\mathbf{x} \in \mathbb{R}^N$ (consisting of a signal **s** in additive noise **n**) is described by a theoretical binary codeword with length (in bits)

$$L(\mathbf{x}, \boldsymbol{\lambda}^{(k)}) = L(\boldsymbol{\lambda}^{(k)}) + L(\mathbf{x}|\boldsymbol{\lambda}^{(k)}) \tag{5}$$

$L(\boldsymbol{\lambda}^{(k)})$ and $L(\mathbf{x}|\boldsymbol{\lambda}^{(k)})$ are the lengths of codewords respectively describing $\boldsymbol{\lambda}^{(k)}$, a k^{th} order parametric model of **x**, and the prediction error for the estimate $\widehat{\mathbf{x}}(\boldsymbol{\lambda}^{(k)})$ derived from the parametric model. $L(\boldsymbol{\lambda}^{(k)})$ and $L(\mathbf{x}|\boldsymbol{\lambda}^{(k)})$ are respectively determined by a universal prefix coding method proposed by Rissanen and by the Shannon coding method. Among admissable parametric models, the model which produces the minimum description length is selected as the model most representative of the signal. In Saito [60] and Pesquet et. al. [53], the parameter vector $\boldsymbol{\lambda}^{(k)}$ contained k transform coefficients and $N-k$ zeros, denoting respective "signal" and "noise" subspace contributions to **s**. For this representation, Saito showed that

$$\boldsymbol{\lambda}^{(k)} = \left\{ \boldsymbol{\lambda} : \max_{\boldsymbol{\lambda} \in \Lambda(k)} \log_2 p_n(\mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\lambda}) \right\} \tag{6}$$

$$L(\boldsymbol{\lambda}^{(k)}) = \frac{3k}{2} \log_2 N + C_1 \tag{7}$$

and

$$L(\mathbf{x}|\boldsymbol{\lambda}^{(k)}) = -\log_2 p_n(\widehat{\mathbf{n}}(k)), \tag{8}$$

where $p_n(\mathbf{n})$ was the probability density of the noise, $\boldsymbol{\Phi}$ was an orthogonal matrix whose columns $\{\phi_i\}_{i=1}^N$ were the minimum-entropy wavelet packet basis for **x**, $\Lambda(k)$ was the subset of vectors in $\mathbb{R}^N$ with $N-k$ zero elements, $\widehat{\mathbf{n}}(k) = \mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\lambda}^{(k)}$ was

the implicit noise estimate corresponding to the k^{th}-order signal model and C_1 was a constant independent of basis or model order.

In this study, the N elements $\{n_\ell\}_{\ell=0}^{N-1}$ of $\mathbf{n}$ are assumed to be generated by an autoregressive (AR) process of the form

$$n_\ell = -\sum_{m=1}^{p} \alpha_m n_{\ell-m} + u_\ell \tag{9}$$

where $p \ll N$, and $\{u_\ell\}$ is a Gaussian white noise process with mean zero and variance of σ_u^2. It has been shown (Gray [31], 1972) that, when $p \ll N$, the linear convolution in Equation 8 may be approximated by a circular convolution, such that $\mathbf{u} = [u_0, u_1, ..., u_n]^H$

$$\mathbf{A}^H \mathbf{n} = \mathbf{u} \tag{10}$$

where

$$\mathbf{A}^H = \begin{bmatrix} 1 & 0 & ... & \alpha_p & ... & \alpha_2 & \alpha_1 \\ \alpha_1 & 1 & 0 & ... & \alpha_p & ... & \alpha_2 \\ ... & ... & ... & ... & ... & ... & ... \\ \alpha_p & \alpha_{p-1} & ... & 1 & 0 & ... & 0 \\ 0 & \alpha_p & ... & \alpha_1 & 1 & ... & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & ... & \alpha_2 & \alpha_1 & 1 \end{bmatrix} \tag{11}$$

The probability density of $\mathbf{n}$ is then

$$p_n(\mathbf{n}) = \left((2\pi\sigma_u^2)^N \det(\mathbf{A}\mathbf{A}^H)^{-1}\right)^{-\frac{1}{2}} \exp\left[-\frac{\mathbf{n}^H \mathbf{A}\mathbf{A}^H \mathbf{n}}{2\sigma_u^2}\right] \tag{12}$$

The codeword length for the prediction error of the k^{th}-order estimate of the signal is then given by

$$\begin{aligned} L(\mathbf{x}|\boldsymbol{\lambda}^{(k)}) &= \frac{N\ln(2\pi\sigma_u^2) + \ln\det(\mathbf{A}\mathbf{A}^H)^{-1} + \mathbf{n}^H\mathbf{A}\mathbf{A}^H\mathbf{n}/\sigma_u^2}{2\ln 2} \\ &= \frac{1}{2\sigma_u^2 \ln 2}\|\mathbf{A}^H\mathbf{n}\|^2 + C_2, \end{aligned} \tag{13}$$

where C_2 is a constant independent of basis or model order. After substituting $\widehat{\mathbf{n}}(\boldsymbol{\lambda}^{(k)})$ for $\mathbf{n}$, $L(\mathbf{x}|\boldsymbol{\lambda}^{(k)})$ becomes

$$L(\mathbf{x}|\boldsymbol{\lambda}^{(k)}) = \frac{\|\mathbf{A}^H\boldsymbol{\Phi}_S(k)\boldsymbol{\lambda}_S^{(k)} + \mathbf{A}^H\boldsymbol{\Phi}_\eta(k)\boldsymbol{\lambda}_\eta^{(k)}\|^2}{2\sigma_u^2\ln 2} + C_2 \tag{14}$$

where

$$\boldsymbol{\Phi}_S(k) = \left[\phi_{z_1}, \phi_{z_2}, ..., \phi_{z_k}\right], \tag{15}$$

and

$$\boldsymbol{\Phi}_{\eta}(k) = \left[\phi_{z_{k+1}}, \phi_{z_{k+2}}, ..., \phi_{z_N}\right]. \tag{16}$$

The index set $\{z_1, z_2, ..., z_N\}$ is an ordering of the basis functions in which the first k basis functions are assumed to contain the signal and the last $N-k$ basis functions are assumed to contain only noise. The manner in which this partition is calculated is described below. The expressions

$$\boldsymbol{\lambda}_S^{(k)} = \boldsymbol{\Phi}_S(k)^H \widehat{\mathbf{n}}(\boldsymbol{\lambda}^{(k)}), \tag{17}$$

and

$$\boldsymbol{\lambda}_{\eta}^{(k)} = \boldsymbol{\Phi}_{\eta}(k)^H \widehat{\mathbf{n}}(\boldsymbol{\lambda}^{(k)}) \tag{18}$$

represent the projection of noise energy onto the "signal" and "noise" subspaces. Since the subspace spanned by $\{\phi_{z_i}\}_{i=k+1}^{N}$ is assumed to only contain noise, $\boldsymbol{\lambda}_{\eta}^{(k)} = \boldsymbol{\Phi}_{\eta}(k)^H \mathbf{x}$. The minimizing value of $\boldsymbol{\lambda}_S^{(k)}$ may then be found by solving a set of normal equations, leading minimization of $L(\mathbf{x}|\boldsymbol{\lambda}^{(k)})$ to become equivalent to maximization of $\|\mathbf{A}^H \boldsymbol{\Phi}_S(k) \boldsymbol{\Phi}_S^H(k)\mathbf{x}\|^2$. $\mathbf{A}^H$, which contains N circular shifts of the prediction error filter, can be written as

$$\mathbf{A}^H = \mathbf{F}\mathbf{Q}_A\mathbf{F}^H \tag{19}$$

where $[\mathbf{F}]_{mn} = \frac{e^{j2\pi mn/N}}{\sqrt{N}}$, and $\mathbf{Q}_A$ is a diagonal matrix containing the value of the FFT of the first row of $\mathbf{A}^H$ divided by $\sqrt{N}$. Sinha and Tewfik ([64], p.3476) have shown that, for the Daubechies filters having $M \geq 10$, $\boldsymbol{\Phi}^H \mathbf{F}\mathbf{Q}_A^H\mathbf{Q}_A\mathbf{F}^H\boldsymbol{\Phi}$ is nearly diagonal. Hence the columns of $\mathbf{A}^H\boldsymbol{\Phi}_S(k)$ are nearly orthogonal,

$$\|\mathbf{A}^H \boldsymbol{\Phi}_S(k) \boldsymbol{\Phi}_S^H(k)\mathbf{x}\|^2 \approx \sum_{i=1}^{k} (\phi_{z_i}^H \mathbf{x})^2 \|\mathbf{A}^H \phi_{z_i}\|^2, \tag{20}$$

and

$$L(\mathbf{x}|\boldsymbol{\lambda}^{(k)}) \approx \frac{\|\mathbf{A}^H\mathbf{x}\|^2 - \sum_{i=1}^{k} (\phi_{z_i}^H \mathbf{x})^2 \|\mathbf{A}^H \phi_{z_i}\|^2}{2\sigma_u^2} + C_2 \tag{21}$$

where $\{z_i\}_{i=1}^{k}$ corresponds to the set of filtered basis vectors with the largest coefficients, and the values $\left\{\mathbf{A}^H \phi_{z_i}\right\}$ are computed from the SIWPT of a time reversed version of the whitening filter $\{\alpha_i\}$. By taking the finite difference (with respect to k) of the sum of Equations (7) and (21), and looking for the value of k for which that difference changes from negative to positive, the MDL criterion may be shown to select the largest value of k for which

$$|\phi_{z_k}^H \mathbf{x}| > \frac{\sigma_u}{\|\mathbf{A}^H \phi_{z_k}\|} \sqrt{3 \ln N}. \tag{22}$$

Hence, for subspace denoising using a single shift-invariant WPT, evaluation of $L(\mathbf{x}|\boldsymbol{\lambda}^{(k)})$ for various values of k may then be replaced by simple threshold comparisons for each basis vector.

In this study, as in every other study of subspace methods, truncating the expansion of the noisy signal imposes some audible artifacts on the enhanced speech. These artifacts may be reduced by averaging together several denoised versions of the signal as described in Section (2.1) with the SIWPT. Note that each SIWPT coefficient for a given scale j and sub-band k corresponds to a shift-variant WPT coefficient in more than one circularly shifted version of the signal. Therefore, averaging N denoised, circularly shifted versions of the signal requires that each coefficient be subjected to multiple threshold rules.

For a more complete discussion of the MDL model of correlated noise, the reader is referred to Whitmal et. al. [74].

2.4 Summary of the algorithm

This section has described a wavelet based approach for enhancing signals in additive autoregressive noise. The algorithm uses the SIWPT of the noise's prediction-error filter to implement a form of transform-domain convolution with wavelet packet basis vectors. The sequence of operations in this algorithm is summarized as follows:

1. Use conventional methods to obtain an AR model of the noise.
2. Derive coefficient variance estimates for both the noise and the noise corrupted speech.
3. Use the variance estimates to select the LDB as described above and in Coifman and Saito [15].
4. Compute the SIWPT of the signal and the filter response in the LDB,
5. Use Equation (21) to select coefficients assumed to represent the signal, and attenuate noise coefficients assumed to represent noise.
6. Resynthesize the signal by means of the inverse SIWPT.

In the next section we will present experimental results which demonstrate the utility of the proposed algorithm.

3 Objective evaluation

This experiment compares the described approach with a conventional spectral subtraction algorithm (Schwarz et al.[63]).

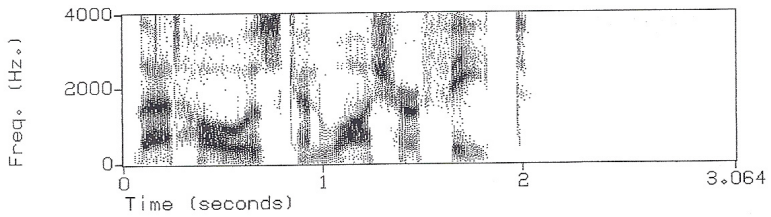

Fig. 2 "That hose can wash her feet"

3.1 Stimuli

A recording of the Harvard / IEEE ([35]) sentence "That hose can wash her feet" (followed by 800msec of silence) was used for all trials of the experiment. (A spectrogram of the signal is shown in figure 2).

The recording was sampled at 8 kHz and combined separately with two different noise maskers to produce two sets of waveforms with overall SNRs of 0, 5, 10, 15, and 20 dB. The first noise masker was derived from a 12 second segment of digitally recorded noise in an automobile traveling at highway speed. The second noise masker was a sequence of autoregressive gaussian noise, produced by a fourth order all pole filter of the form

$$n_k = 1.3733n_{k-1} - 0.7858n_{k-2} + 0.1138n_{k-3} + 0.0119n_{k-4} + u_k \qquad (23)$$

where $\{u_k\}$ was a sequence of white Gaussian noise, and the filter parameters were derived from a 2 second segment of digitally recorded car noise. The AR model produces a noise spectrum which is relatively flat below 1 kHz, and rolls off at about 12 dB/octave above 1 kHz, as shown in figure 3. A spectrogram of the noise corrupted sentence at 5 dB SNR is shown in figure 4.

Noise maskers of this type were selected for two reasons. First, the car noise provides an excellent example of a real-world masker that can be modeled suitably by autoregressive noise. Second, the car noise's inherent non-stationarity allows us to evaluate the algorithm's performance in settings which approximate practical applications.

The ten noise-corrupted speech signals were processed in 50% overlapped segments by a version of the SIWPT algorithm that uses a single-frame estimate of the the noise's whitening filter. In previous work (Whitmal et al.[74]), that algorithm improved the SNR of speech signals (taken in 256 sample frames) in simulated car noise at 0 dB by 11.68 dB, and reduced noise in pause regions by as much as 23.77 dB. The signals were also processed by a spectral subtraction algorithm similar to that of Schwartz et al. [63], which estimates the signal by subtracting a weighted version of an estimated noise power spectrum from the observed signal's power spectrum. Frame lengths of 64 msecs (512 samples at 8kHz) were used in both al-

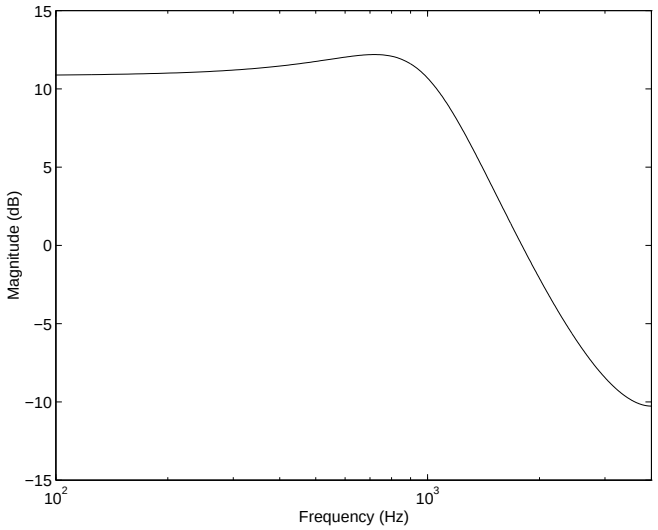

Fig. 3 Spectra of simulated car noise

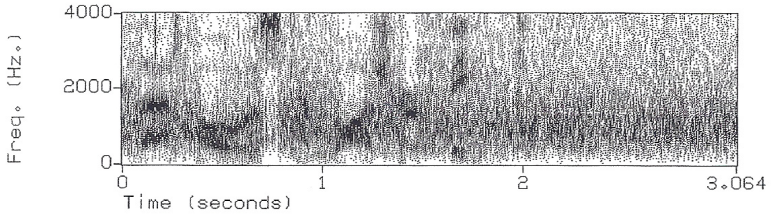

Fig. 4 Sentence of figure 2, in simulated car noise at 5 dB SNR

gorithms to produce enhanced signals with low residual noise output. For reference, signals were also processed by a SIWPT algorithm (similar to that of Pesquet et al. [53]) designed to reduce white noise.

Both Schwartz et al. [63] and Boll [4] note that the spectral subtraction algorithm's parameters must be carefully selected if the algorithm is to provide a good tradeoff between signal distortion and residual noise output. Preliminary listening indicated that a noise floor constant of .01 and oversubtraction by a factor of 5 would produce intelligible speech with acceptable amounts of residual noise. The long frame length contravenes the guidelines of Schwartz et al. [63]. In informal listening, however, spectral subtraction output for a 64 msec frame length sounded less noisy and more faithful than output using a 16 or 32 msec frame length. The reader is referred to Schwartz et al. [63] for a more detailed discussion of these parameters and their effects on noise reduction.

3.2 Methods

The performance of the algorithms was evaluated by means of two objective measures: segmental SNR and log spectral distance. These measures are respectively defined as

$$SNR = \frac{1}{L}\sum_{i=1}^{L} 20\log_{10}\frac{\|\mathbf{s}_i\|^2}{\|\mathbf{s}_i - \widehat{\mathbf{s}}_i\|^2} \tag{24}$$

and

$$LSD = \frac{1}{NL}\sum_{i=1}^{L}\left[\sum_{m=0}^{N-1}|20\log_{10}\frac{|S_i(m)|}{|\widehat{S}_i(m)|}|\right], \tag{25}$$

where $\mathbf{s}_i$ and $\widehat{\mathbf{s}}_i$ are the actual and estimated speech signals in the i-th of L $N-$sample frames of $\mathbf{s}$ and $S_i(\cdot)$ and $\widehat{S}_i(\cdot)$ are their respective N-sample DFTs. These measures were selected both because of their minimal computation requirements, and because of their high correlations with subjective speech quality measures (e.g., the Diagnostic Acceptability Measure (DAM) of Voiers [71]). The reader is referred to Quackenbush et al. [56] for a detailed discussion of the use of these measures.

Measurements of segmental SNR and log spectral distance were computed for non-overlapping 256-sample segments and averaged over speech regions to produce objective measures of quality. Data for both measures are shown in table 1 and table 2. To assess noise reduction in the silence regions, the RMS level of residual noise was computed and compared with the RMS level of the original noise masker. Comparisons of these noise levels are presented in table 3.

3.3 Results

The data of table 1 and figure 5 indicate that, for both types of noise, the SIWPT algorithm described above improves segmental SNR better than the standard spectral subtraction algorithm and the conventional SIWPT algorithm. These results are particularly encouraging for the case of actual car noise, where, as expected, the proposed algorithm's performance decreases. Spectrograms of sentences processed by the SIWPT algorithm and by the spectral subtraction algorithm in simulated noise are shown in figures 5 and 6. The presence of spectral subtraction's "musical noise" is indicated by the large number of short-duration, high-energy bursts visible in figure 6. For reference, a spectrogram of output from the conventional algorithm is shown in figure 7

In contrast, the data of table 3 indicate that the spectral subtraction algorithm improves log spectral distance better than the proposed algorithm, particularly in simulated car noise. The increase in distance is largely due to the algorithm's strong attenuation of low energy consonants which often resemble gaussian noise. This phenomenon is illustrated in figure 8, which compares log spectral distance measures for the proposed algorithm in both types of noise at 0 dB SNR. As indicated

Table 1 Segmental SNR measures (in dB) for spectral subtraction and SIWPT-based subspace denoising algorithms in two types of noise

	Control		SIWPT, prop.		SIWPT, conv.		Spec. sub.	
SNR	S	R	S	R	S	R	S	R
0	-10.21	-10.30	1.92	-0.20	-3.34	-4.76	-2.01	-1.88
5	-5.20	-5.30	4.53	2.98	0.39	-0.85	1.82	1.94
10	-0.20	-0.30	7.27	6.26	4.00	2.95	5.32	5.55
15	4.80	4.70	10.12	9.37	7.60	6.70	8.47	8.86
20	9.80	9.71	13.55	12.77	11.27	10.70	11.38	11.80

In each of the tables, measures taken in simulated and real car noise are respectively denoted by "S" and "R" where R = real automobile road noise and S = synthesized road noise. Parenthesized figures indicate number of frames (out of 34) correctly estimated to contain silence.

Table 2 Noise reduction measures (in dB) for spectral subtraction and SIWPT-based subspace denoising algorithms in two types of noise

	SIWPT, prop.		SIWPT, conv.		Spec. sub.	
SNR	S	R	S	R	S	R
0	42.55 (19)	18.49 (2)	11.34 (0)	8.05 (0)	12.96 (0)	12.77 (0)
5	42.72 (21)	18.49 (4)	11.35 (0)	8.06 (0)	12.96 (0)	12.77 (0)
10	42.89 (25)	18.50 (4)	11.37 (0)	8.06 (0)	12.96 (0)	12.77 (0)
15	43.15 (26)	18.51 (6)	11.39 (0)	8.07 (0)	12.96 (0)	12.77 (0)
20	43.54 (29)	18.56 (6)	11.38 (0)	8.10 (0)	12.96 (0)	12.77 (0)

R = real automobile road noise
S = synthesized road noise

Table 3 Log spectral distance measures (in dB) for spectral subtraction and SIWPT-based subspace denoising algorithms in two types of noise

	Control		SIWPT, prop.		SIWPT, conv.		Spec. sub.	
SNR	S	R	S	R	S	R	S	R
0	18.16	17.87	39.27	15.82	14.01	13.77	11.83	11.60
5	14.71	14.58	32.91	12.62	12.97	12.87	10.13	9.77
10	11.70	11.54	24.96	11.19	11.37	12.61	8.50	8.09
15	9.20	9.03	21.61	9.37	10.74	11.07	8.47	8.86
20	7.00	6.85	17.50	7.92	9.99	9.86	5.59	5.47

R = real automobile road noise

S = synthesized road noise

by the figure, distance measures for the two noise types are close in value in regions of moderate SNR. The conflicting trends in the two objective measures provide further illustration of the tradeoffs made between noise reduction and signal distortion. In understanding this conflict, it should be noted that the correlations between segmental SNR and DAM scores demonstrated in Quackenbush et al. [56] have only been confirmed for waveform coding methods. It is not clear how the type of dis-

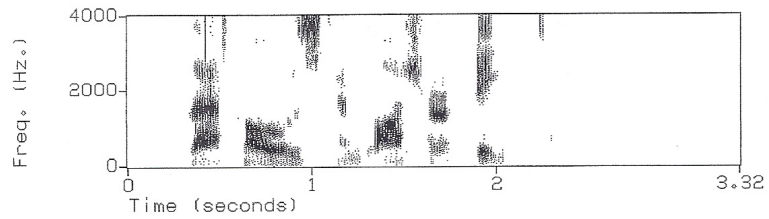

Fig. 5 Sentence of Figure 4 processed by proposed subspace algorithm

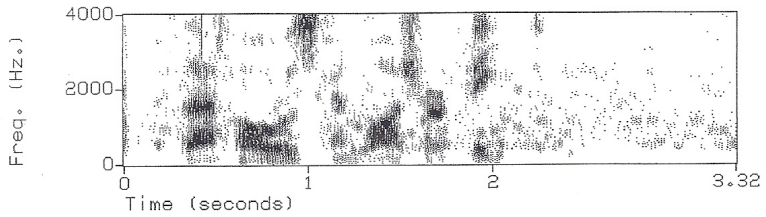

Fig. 6 Sentence of figure 4, processed by spectral subtraction

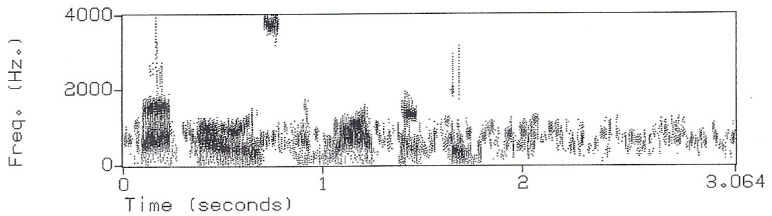

Fig. 7 Sentence of figure 4, processed by conventional subspace algorithm

tortion demonstrated in figure 8 will affect segmental SNR measurements, since the present approach bears a closer resemblance to transform coding.

4 Subject testing

The proposed noise reduction algorithm was tested on hearing impaired individuals at the Northwestern University department of Communication Disorders.

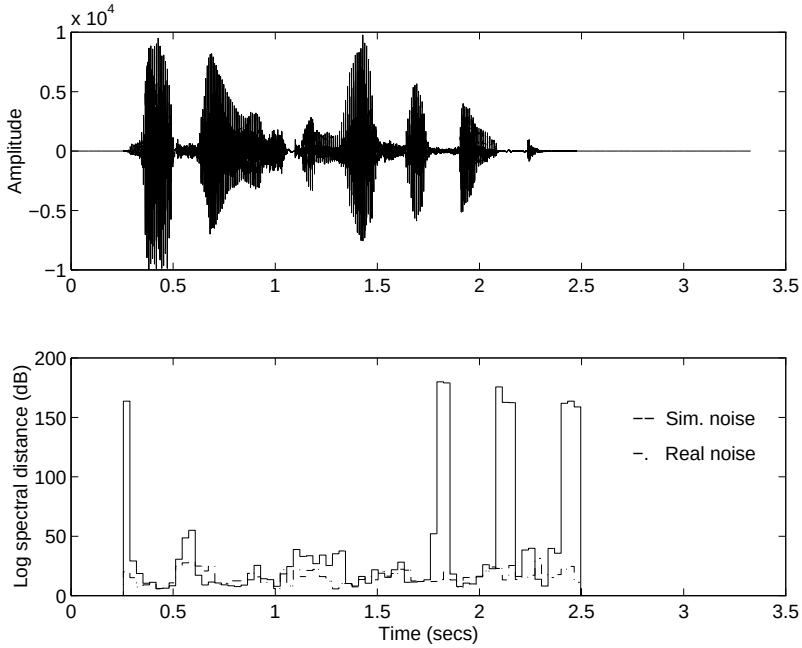

Fig. 8 Log spectral distance for the SIWPT method at 0 dB SNR

4.1 Subjects

Participants in this study consisted of 16 Northwestern University Hearing Clinic patients with impaired hearing acuity. Eligible patients were screened on the basis of their audiograms and assigned to one of two categories: "flat" (i.e., hearing loss slope above 500 Hz $\leq$ 10 dB/octave) and "sloped high-frequency" (i.e., hearing loss above 500 Hz $>$ 10 dB/octave). Each category consisted of eight subjects. Mean hearing losses for each group are shown in table 4.

4.2 Test protocol

The new method was evaluated in trials of the Revised Speech Perception in Noise (R-SPIN) test [3] at the facilities of the Northwestern University Hearing Clinic in Evanston, Illinois. The R-SPIN test protocol uses eight lists of 50 recorded sentences, each of which are presented to subjects in the presence of interfering noise. Subjects attempt to identify the last word of each sentence. In each list, the last word of 25 of the sentences is predictable from context (e.g., "Stir your coffee with a spoon"), while the last word of the other 25 sentences is difficult to predict from

Table 4 Profiles of impaired-acuity subjects

(Previously measured thresholds are listed in parentheses for reference. Standard deviations are rounded to nearest dB to maintain format.)

Flat-loss subjects								
Demographics				Hearing loss (dB HL)				
Subject	Age	Sex	Ear	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz
FC	77	F	R	20 (30)	30 (35)	45 (40)	55 (50)	55 (50)
BF	74	M	R	45 (50)	45 (50)	50 (55)	55 (60)	60 (60)
RF	61	M	R	60 (65)	55 (60)	70 (70)	65 (65)	80 (80)
IG	86	M	L	40 (30)	50 (50)	45 (45)	45 (40)	65 (50)
DK	18	F	R	45 (45)	45 (45)	50 (45)	55 (55)	50 (45)
LL	82	F	R	20 (25)	20 (30)	25 (30)	40 (40)	35 (45)
CM	77	M	L	25 (40)	30 (50)	45 (45)	50 (45)	60 (55)
MM	86	M	L	15 (30)	25 (35)	40 (45)	40 (45)	40 (45)
Means	70			34 (39)	40 (44)	46 (47)	49 (50)	56 (54)
SDs				15 (12)	12 (9)	12 (11)	8 (9)	13 (11)

Sloped-loss subjects								
Demographics				Hearing loss (dB HL)				
Subject	Age	Sex	Ear	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz
FA	56	M	L	35 (30)	30 (30)	40 (30)	50 (45)	55 (55)
LK	76	M	R	25 (30)	30 (35)	25 (35)	40 (45)	60 (60)
RL	66	M	L	25 (20)	25 (25)	30 (30)	55 (40)	65 (55)
DM	82	F	L	30 (25)	25 (30)	40 (45)	50 (60)	75 (70)
SM	88	F	R	15 (20)	20 (20)	20 (20)	50 (50)	65 (65)
HM	83	M	L	10 (20)	30 (35)	50 (50)	55 (55)	55 (60)
HN	83	M	R	20 (25)	25 (25)	30 (35)	50 (50)	65 (70)
HP	71	M	R	20 (20)	25 (30)	45 (40)	55 (55)	50 (55)
Means	76			23 (24)	28 (29)	35 (36)	51 (50)	60 (61)
SDs				8 (4)	3 (5)	10 (9)	5 (6)	7 (6)

context (e.g., “Bob could have known about the spoon”). These two types of sentences are denoted as “PH” and “PL” sentences. This approach allows researchers to simultaneously assess how well subjects are able to both hear and interpret noise-corrupted speech.

Recordings of R-SPIN sentences were sampled at 8 kHz and combined electronically with digital recordings of automobile road noise at signal-to-noise ratios (SNRs) of 5 dB and 10 dB. Individual tapes of the eight lists were then made for each subject participating in the test. For every tape, each of the eight lists was assigned one of eight possible combinations of the two SNR levels and one of four processing options:

1. Control (sentence in noise, no processing).
2. Linear filtering per the Australian National Acoustic Laboratories’ (NAL) hearing-aid gain-prescription rule [9].
3. The shift-invariant wavelet-packet transform (SIWPT) based noise reduction algorithm of [74].
4. a combination of NAL filtering and SIWPT-based processing.

The NAL gain-prescription rule is a formula (derived from clinical data) which assigns approximately .36 dB of gain for every 1.0 dB of hearing loss. NAL prescriptions for each patient were pre-calculated from previous audiological measurements, and used to design conventional FIR digital filters for convolution with R-SPIN sentence recordings. RMS levels for each sentence were then matched to that of a calibration tone before being transferred to digital audio tape. each tape, which contained a training list of R-SPIN sentences in quiet followed by the processed sentences, was then presented monaurally over headphones to the subject at that subject's most comfortable listening level.

4.3 Experimental results

Mean R-SPIN scores for both the flat and sloped subject groups are presented in table 5. Inspection of these scores reveals several interesting trends. In particular, for flat-loss subjects, NAL filtering appeared to have a negative effect on many R-SPIN scores. The proposed SIWPT method, in contrast, appeared to improve R-SPIN scores for flat-loss subjects.

Table 5 Mean R-SPIN scores for subjects with impaired acuity

	Processing for flat-loss subjects			
Context/SNR	Control	NAL	SIWPT	NAL+SIWPT
PH, 5 dB	18.63	16.00†	18.00	14.75†
PH, 10 dB	22.13	20.75	23.13	22.38
PL, 5 dB	7.25	7.38	8.13	8.88
PL, 10 dB	10.50	10.75	13.13†	12.50
	Processing for sloped-loss subjects			
Context/SNR	Control	NAL	SIWPT	NAL+SIWPT
PH, 5 dB	18.13	16.13	16.13	15.88
PH, 10 dB	21.75	22.25	21.25	21.75
PL, 5 dB	7.75	6.63	7.63	6.63
PL, 10 dB	10.13	10.38	10.00	9.75

† denotes significance per Dunnett T-test at $\sim$0.06 level

A repeated-measures analysis of variance (ANOVA) was conducted to evaluate the dependence of R-SPIN scores on impairment type, subject, sentence context, SNR, and processing type. The results of the ANOVA indicate that the factors of subject, context and SNR have statistically significant $(p < .001)$ effects on R-SPIN scores. Context effects on R-SPIN scores are consistent with effects seen in previous research [8]. However, the results of the ANOVA do not indicate the differences in processing have a significant effect on R-SPIN scores.

The significance of subjects as a factor led us to conduct an analysis of individual R-SPIN scores. The Thornton-Raffin binomial model of test scores [68] was used to construct 95% confidence intervals for each of the control scores. Scores for the

non-control process were then compared to the confidence interval limits of the corresponding control and subsequently classified as "better, "same" or "worse" than the control score. The results of these comparisons (presented in table 6) suggest that SIWPT processing tends to improve R-SPIN scores at 10 dB SNR, and that both types of processing tend to reduce R-SPIN scores at 5 dB SNR. These results appear to be strongly influenced by the flat loss group, which accounted for six of the nine subjects reporting at least one significant deviation.

Table 6 Distribution of relative R-SPIN score ratings

	5 dB SNR			10 dB SNR		
Rating	Worse	Same	Better	Worse	Same	Better
NAL	3	11	2	2	12	2
SIWPT	3	11	2	1	11	4
Both	3	12	1	2	12	2

5 Built-in distortions

Both the original and proposed subspace algorithms construct "signal" subspaces from the subset of basis vectors which produce the largest transform coefficients. The conventional algorithm uses a standard wavelet-packet basis (where all basis vectors contain the same amount of energy), and selects basis vectors solely on the basis of coefficient size. Since both the original speech and the car noise have most of their energy at low frequencies, the conventional algorithm tends to preserve basis vectors with substantial energy at low frequencies. In contrast, the filtering imposed on the new algorithm's basis vectors increases the energy in vectors whose spectrum is most severely compromised by the noise, thereby offsetting the low SNRs in these subspaces.

Some comments about the objective data are appropriate at this point. First, we note that the combination of coefficient attenuation and averaging can produce benign distortion artifacts which result in objective measures with overly pessimistic values. (This is particularly true at the 20 dB SNR). Of greater concern is the tendency of this algorithm to attenuate both speech and noise in regions of low SNR. This tradeoff is particularly noticeable in certain consonant regions, as indicated by figures 2, 4 and 5. This tendency is a direct consequence of the binary threshold measure used in wavelet-packet based denoising algorithms. Use of this approach is motivated by two factors. First, use of the entire coefficient value (rather than a scaled version) appears to provide smaller values of mean-squared error (Coifman and Donoho [12]). Second, in algorithms using wavelet-packet transforms with "best basis" approaches, the popular linearly-tapered thresholds used in many wavelet transform based studies (e.g., see Donoho [18]) appear to have little effect on the signal. This may be explained heuristically by noting that the "best-basis" algorithm

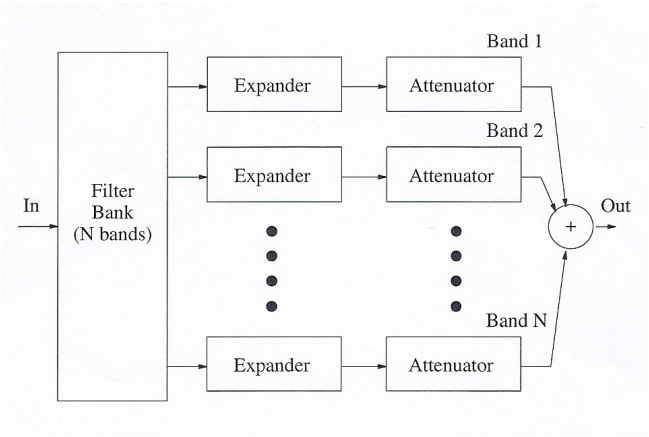

Fig. 9 Block diagram of ideal recruitment-of-loudness simulator

increases the size of "signal" and reduces the size of the "noise" coefficients to the point where thresholding effects on either type of coefficient are reduced considerably. It is possible that improved performance may be obtained by using several Bayesian threshold approaches (Vidakovic [69]; Chipman et al. [11]) which were developed in concurrence with the present work.

The results of section 4.3 demonstrated that the proposed method fails to provide significant improvements in intelligibility despite improvements in objective measures. One possible explanations is the level-dependent nature of the attenuation that the algorithms impose on the speech. In the following section, we show (both theoretically and empirically) that the proposed algorithm attenuates input speech in a manner that resembles the attenuation produced by recruitment-of-loudness simulators, such that their attenuation can be equated with an effective sensorineural hearing loss.

Figure 9 shows an idealized processor that simulates the effects of recruitment of loudness for normal-hearing listeners. The processor uses a filter bank to partition the signal into several frequency bands. Signals in each band are then input to a combination of amplitude expansion and attenuation modules. The expansion ratio of each expander is set equal to the ratio of the dynamic range in that band for a normal-hearing listener to the dynamic range in that band for the hearing-impaired listener. The attenuation factor is likewise determined such that the speech has the same relative intensity to the normal threshold as the corresponding element of the unprocessed speech has to the impaired threshold.

An example of recruitment simulation is provided by Duchnowski and Zurek [24], who used piecewise-linear expansion characteristics in each of 14 frequency bands. The input/output relationship in each band was given by the relation

$$y = \begin{cases} Kx + (1-K)T_c, & x < T_C \\ x, & x \geq T_C \end{cases}$$

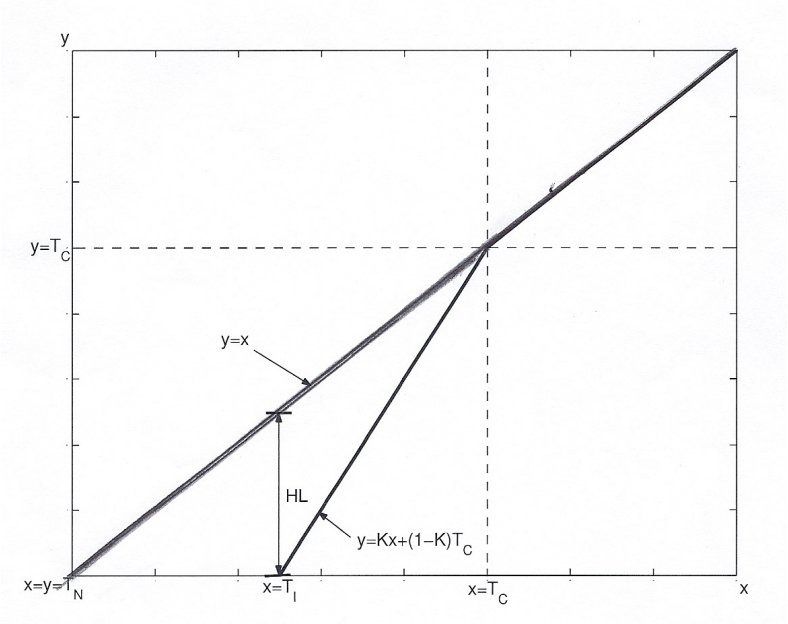

Fig. 10 Input-output characteristic of Duchnowski-Zurek recruitment simulator (Equation 26}

where x is the input level in dB SPL, (sound pressure level), y is the output level in dB SPL, T_c is the effective level of complete recruitment (i.e., the level at which loudness perception in the impaired ear equals that of the normal ear), and K, in turn, is given as

$$K = \frac{T_C - T_N}{T_C - T_I}$$

where T_N and T_I are (respectively) the normal and impaired thresholds of hearing, and $T_I - T_N$ is the hearing loss (HL) modeled by the simulator. An illustration of the expander characteristic is shown below in figure 10. Values of T_C were derived from the data of Hallpike and Hood [34], who expressed K as the tangent of a "recruitment angle" which was proportional to the degree of hearing loss. The data of duchnowski and Zurek [24] indicate good agreement between intelligibility scores of subjects with moderate sensorineural hearing losses and normal-hearing subjects listening to the output of simulators matched to the subjects' audiograms.

The attenuation of characteristics produced by these simulators are also produced by subspace algorithms like the proposed method. Subspace algorithms have typically employed two types of attenuation rules: "hard" threshold rules and "soft" threshold rules. Hard threshold rules apply no attenuation to "speech subspace" coefficients (i.e. all coefficients above the noise threshold), and infinite attenuation to "noise subspace" coefficients. An approximate expression for hard-thresholding effects can be derived for algorithms using orthonormal transforms, using common

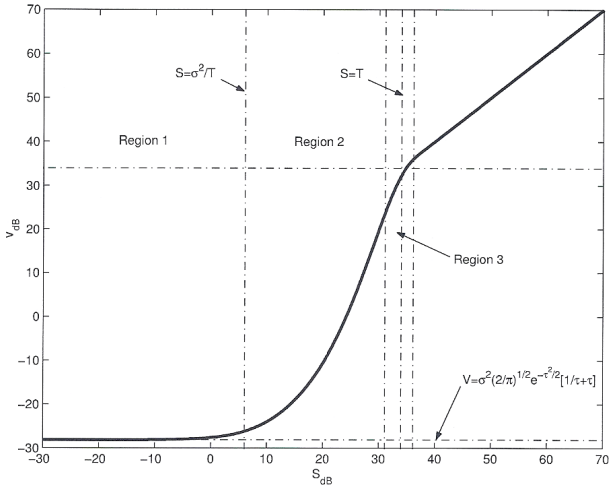

Fig. 11 A sample recruitment curve for subspace noise reduction ($\sigma_D = 10, \tau = 5$)

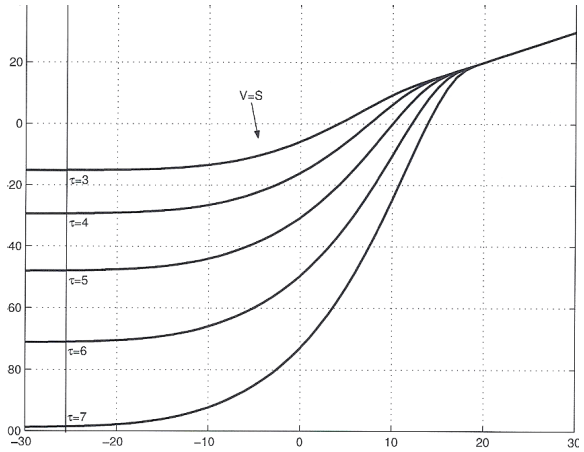

Fig. 12 Effects of τ on subspace noise reduction performance.

(Jansen [37]) assumptions of deterministic coefficients in additive white Gaussian noise. (Results derived for additive white noise can be extended to cases of colored noise through the use of whitening filters, as discussed in Ephraim and van Trees [28]). For a particular basis function, let the observed coefficient X (a sum of signal coefficient S and random noise coefficient D) be subjected to hard-thresholding. D is assumed to have a mean of zero and a variance of σ_D^2. The threshold coefficient is given as

$$V = \begin{cases} X, & |X| > T \\ 0, & |X| \leq T \end{cases}.$$

V has the probability distribution

$$f_V(v;S) = \left[\left(\frac{1}{\sqrt{2\pi\sigma_D^2}} e^{-(v-S)^2/2\sigma_D^2} \right) (u(v-T) + u(-v-T)) \right] + \Pr(S)\delta(v),$$

where $u(v)$ is the Heaviside step function, $\delta(v)$ is the Dirac δ function and $\Pr(S)$, the probability that the coefficient S is attenuated, is

$$\Pr(S) = \frac{1}{2}\left[\operatorname{erf}\left(\frac{T+S}{\sigma_D\sqrt{2}} \right) + \operatorname{erf}\left(\frac{T-S}{\sigma_D\sqrt{2}} \right) \right].$$

Using integration by parts (Stein [66]), we can show that the expected energy in the transform coefficient is

$$\begin{aligned} E\left\{v^2\right\} &= \int_{-\infty}^{\infty} v^2 f_V(v)dv \\ &= (S^2 + \sigma_D^2)(1 - \Pr(S)) \\ &+ \frac{\sigma_D^2}{\sqrt{2\pi}} \left(\frac{T+S}{\sigma_D} \right) e^{-(T-S)^2/2\sigma_D^2} + \frac{\sigma_D^2}{\sqrt{2\pi}} \left(\frac{T-S}{\sigma_D} \right) e^{-(T+S)^2/2\sigma_D^2} \end{aligned}$$

or

$$\begin{aligned} E\left\{v^2\right\} &= \sigma_D^2 \left(\left(\frac{S}{\sigma_D} \right)^2 + 1 \right) (1 - \Pr(S)) \\ &+ \frac{\sigma_D^2}{\sqrt{2\pi}} \left[\left(\tau + \frac{S}{\sigma_D} \right) \right] e^{-(\tau-S)^2/2\sigma_D^2} + \frac{\sigma_D^2}{\sqrt{2\pi}} \left[\left(\tau - \frac{S}{\sigma_D} \right) \right] e^{-(\tau+S)^2/2\sigma_D^2} \end{aligned} \tag{26}$$

where $\tau = \frac{T}{\sigma_D}$. (Typically $\tau > 2\sigma_D$.) Setting $S_{dB} = 20\log_{10}|S|$ and $V_{dB} = 10\log_{10}\left(E\left\{v^2\right\}\right)$ gives

$$\begin{aligned} V_{dB} &= 10\log_{10}\sigma_D^2 + 10\log_{10}\left[\left(\left(\frac{10^{S_{dB}}}{\sigma_D} \right)^2 + 1 \right) \left(1 - \Pr\left(10^{S_{dB}/20} \right) \right) \right] \\ &+ \left(\tau + \frac{10^{S_{dB}/20}}{\sigma_D} \right) \frac{e^{-\left(\tau - \frac{10^{S_{dB}}}{20} \right)^2/2\sigma_D^2}}{\sqrt{2\pi}} + \left(\tau - \frac{10^{S_{dB}/20}}{\sigma_D} \right) \frac{e^{-\left(\tau + \frac{10^{S_{dB}}}{20} \right)^2/2\sigma_D^2}}{\sqrt{2\pi}} \end{aligned} \tag{27}$$

At high input levels, where $S >> T$, $\Pr(S) \approx 0$ and $e^{-(\tau \pm S)^2/2\sigma_D^2} \approx 0$, it follows that $E\left\{v^2\right\} \approx 0$ and $V_{dB} = S_{dB}$. At input levels below T, equation (26) exhibits level-dependent attenuation (or equivalently, an accelerated rate of increase in output level) that resembles recruitment of loudness simulation. An example of this

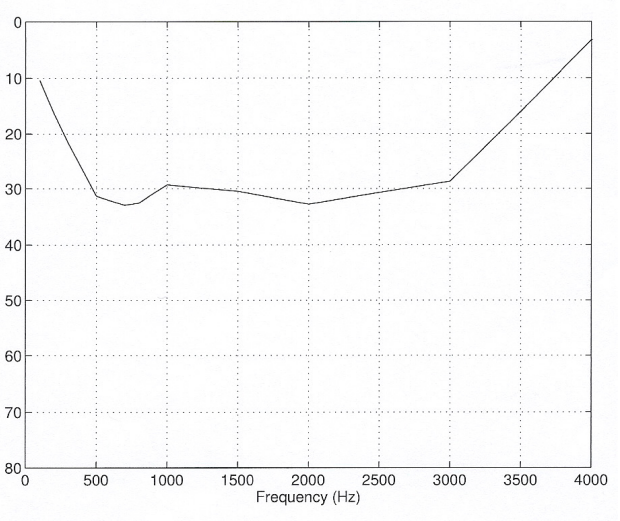

Fig. 13 Effective hearing loss due to hard-thresholding

behavior (with all regions marked) is shown in figure 11, with $\sigma_D = 10$ and $\tau = 5$. The effects of the choice of τ are illustrated in figure 12, which plots input and output levels (normalized by σ_D) for $3 \leq \tau \leq 7$. Figure 12 shows that large values of τ produce steep recruitment slopes, and small values of τ produce shallow slopes. Hence τ determines both the rate of attenuation and the level of residual noise. Note that the energy in each coefficient is typically spread over a band of frequencies, rather than being concentrated at a single frequency.

The idealized input-output relation of (27) is a smooth, differentiable function; as such, standard numerical algorithms can be used to find effective hearing losses for any given output level. However, for the present algorithm, the signal and noise subspaces are time-invariant, and selected on the basis of instantaneous values of signal and noise. The time-varying nature of these algorithms makes it difficult to derive an expression for their expected coefficient values. One way to address this problem is to fit long-term measurements of input and output level to a parametric model that characterizes the algorithm's level-dependent attenuation, and can be used to determine the effective hearing loss imposed by the algorithm.

This task was performed by Whitmal and Vosoughi [77], who measured speech and noise levels for recordings of the syllables "of", "ook", "od", "pa", "ja", and "ma", taken (respectively) from word lists from the Nonsense Syllable Test (Levitt and Resnick, [48]). The syllables were specifically selected to give an appropriately diverse representation of the syllables used in the test. Models for the processed signals were derived from RMS level measurements taken in critical bands. Clean speech at 74dB SPL was filtered into twelve bands centered at 100, 250, 500, 750, 1000, 1500, 2000, 3000, and 4000 Hz. Since the non-linear expansion of the

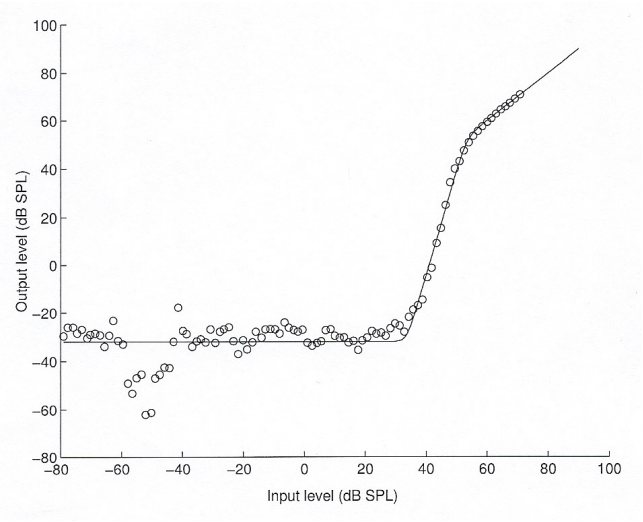

Fig. 14 A recruitment curve for wavelet-packet subspace noise reduction measured at 500 Hz

wavelet-based algorithm imposes smearing on the signal spectrum, prefiltering was used to isolate processed components in one band from leakage of processed components in other bands. Simulated automobile noise at 56 dB SPL was added electronically to each of the nine signals which were then processed by the wavelet-based noise reduction algorithm. (The use of broadband noise insures that the algorithm uses appropriate thresholds for each of the narrow-band signals.)

Following processing, the output signals were upsampled to 24000 Hz and input to a bank of critical filters with center frequencies 100, 250, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000, 8000 and 10000Hz. The output of each filter was then passed through an RMS level detector (time constant: 21.3 msecs) and converted to decibels. The RMS level measurements for the processed speech were combined in time-aligned ordered pairs with concurrent RMS level measurements for clean speech, plotted on scattergrams (see figure 14) and fitted to piecewise-linear functions from which values of T_N and T_I could be extracted. The resulting effective hearing loss is plotted in figure 13 according to audiological conventions, where normal hearing (0 dB loss) is at the top of the y-axis and profound loss (120 dB) at the bottom. The present algorithm produces an effective hearing loss for speech of about 30 dB between 500 Hz and 3000 Hz. In combination with a patient's hearing loss, the present algorithm would be expected to result in poorer intelligibility.

6 Conclusions

The presence of ambient noise presents a major problem for hearing impaired listeners. When amplified by a hearing aid, it creates an annoying sound that degrades the intelligibility of speech. The results of our experiments with noise reduction were mixed. As described in section 3, our algorithm was quite successful in reducing noise; however, as reported in section 4, it produced no comparable improvement in intelligibility when tested on hearing impaired subjects. The reduction of noise, in and of itself, is helpful because it improves the quality of sound. But people use hearing aids because they have trouble understanding conversations, particularly in noisy environments. Even people with normal hearing can have difficulty understanding a conversation in a crowded room with several conversations occurring simultaneously. Ideally, a hearing aid would be able to both reduce noise and improve intelligibilty, a goal that has eluded hearing aid producers. In Section 5, we give a partial explanation of what makes this problem so challenging. Whitmal and Vosoughi found [77] that the output of our algorithm is similar to a process that simulates recruitment of loudness (a hearing problem) for listeners with normal hearing.

Recruitment of loudness is a hearing disorder that is characterized by compression of a person's dynamic range. The term dynamic range refers to the level of loudness between the barely audible and the painfully loud. It can vary a great deal from one individual to another, and more importantly, from one frequency to another. With recruitment of loudness, the threshold of audibility is raised, usually more heavily in the higher frequencies, without a comparable rise in the threshold of pain. The resulting compression of the dynamic range causes distortions in the processing of speech signals leading to a loss of intelligibility by the impaired listener. Older people, for example, often find it difficult to understand the high pitched voices of children. What Whitmal and Vosoughi [77] showed was that hard-thresholding, a feature of our algorithm, produced effects that resembled recruitment of loudness.

Our experiments suggest a more fundamental problem. Algorithms that reconstruct a signal from a small percentage of the transform coefficients lose information. Data compression algorithms, for example, frequently result in some signal distortion. The television reception coming from uncompressed digital signals via an antenna compares favorably with the reception obtained from the compressed digital signals transmitted via cable. The full signal clearly produces a better image. Nevertheless, the amount of distortion is not very noticeable and most users find that the advantages of cable outweigh its disadvantages.

Improving the intelligibility of noisy speech is far more challenging. Speech signals contain some redundancy in the sense that, for example, if the high frequency components of recorded speech are filtered out, it may still be recognizable to a person with normal hearing. However, some of these redundancies are essential for hearing impaired listeners who may no longer be able to understand the filtered speech. Comprehending speech involves more than just the ear's processing of the signal. It involves the way the mind organizes and processes information. The loss

of some of the speech signal that occurs with the denoising algorithm reduces the ability of the impaired listener to do so.

When we started our experiments with denoising algorithms, people who worked on hearing aids told us that the key to whether wavelet packets would help was in how well they could match the shape of speech signals. The linear discriminant basis procedure produced a "signal" subspace that is only a "most likely to be signal" subspace. In the end the algorithm did not project onto a basis of phonemes but projected onto a basis of wavelet packets that may have collectively resembled human speech, but was was not close enough to render the noisy sentences more intelligible. The wavelet packets were built from a small library of wavelet functions. Though the number of possible wavelet packet bases generated by a single wavelet was quite numerous, the geometry of the different wavelet packet basis functions was similar. The underlying library of wavelets was simply not robust enough to match the components of speech.

Recent work in the area of the geometry of large data sets has given rise to a whole new generation of orthonormal bases. These bases are adapted to the data they are describing and so offer the promise of overcoming some of the intelligibility challenges in the construction of hearing aids. A special issue of the online journal *Applied and Computational Harmonic Analysis* devoted to diffusion maps [1] introduced novel ways of applying the tools of harmonic analysis to the study of the geometry of large data sets. The paper of Coifman and Maggioni, *Diffusion wavelets* [13], and the companion paper of Bremer, Coifman, Maggioni and Szlam, *Diffusion wavelet packets* [7], provide a framework in which diffusion wavelet packets could be used to denoise speech signals in a manner similar to the research described in this paper. The geometry of diffusion wavelet packets, matching more closely the geometry of human speech, might be able to reduce the noise without losing as much of the signal.

Partha Niyogi and Aren Jansen have applied these types of ideas to the geometry of voiced speech. Their paper *Intrinsic Fourier analysis on the manifold of speech sounds* [36] shows an efficient algorithm for creating a Fourier basis on the manifold of possible speech signals whose elements correspond to the phonemic content of speech. Their paper gives an example of using this type of basis to create a spectrogram for the word "advantageous" that more efficiently identifies its phonemes than a conventional Fourier based spectrogram.

So far as we know, nobody has applied these mathematical developments to research aimed at improving hearing aids. Past results suggest that such methods are very likely to improve noise reduction. It is less certain, however, that such methods will improve intelligibility, but because these wavelets are closely matched to the structure of speech at the phonemic level, they offer some hope of success.

We have found in our research that even small distortions of the speech component of a noisy signal can render the denoised speech no more understandable than the original noisy signal. The only way to determine if the method improves intelligibility is to test it with hearing impaired listeners. In the end, it is only the individual user who can tell you if the hearing aid is helpful. A mathematical theory can tell you that according to certain mathematical assumptions, a particular

threshold is optimal, or that according to a set of objective measures one algorithm compares favorably with alternative methods; but only through subject testing can you find out if the method improves intelligibility. We hope that our research will prove helpful in providing some insight into some of the challenges facing the use of wavelet type bases in hearing aid research as well as offering some guidelines about conducting such research and testing the validity of its findings.

References

1. *Special Issue on diffusion maps and wavelets*, Appl. Comput. Harmon. Anal. **21**, (2006).
2. J. Berger, R. Coifman, M. Goldberg, *Removing noise from music using local trigonometric bases and wavelet packets*, Jour. Audio Eng. Soc., 808-818, (1994).
3. R. Bilger, J. Nuetzel, W. Rabinowitz and C. Rzeczowski, *Standardization of a test of speech perception in noise*, J. Speech Hear. Res., **27**, 32-48 (1984).
4. S. Boll, *Suppression of acoustic noise in speech using spectral subtraction*, IEEE Trans. Accoustics, Speech, and Sig. Proc. **27**, 113-120, (1979).
5. S. Boll, *Speech enhancement in the 1980's: noise suppression with patern matching*. In S. Furui and M. Sondhi, editors, Advances in Speech Signal Processing, Marcel Dekker, 309-326.
6. G. Box and G. Jenkins, *Time Series Analysis-Forecasting and Control*, Holden Day, San Francisco, Ca., (1970).
7. J.C. Bremer, R.R. Coifman, M. Maggioni, A.D. Szlam, *Diffusion wavele packets*, Appl. Comput. Harmon. Anal. **21**, 95-112, (2006).
8. R. Brey, M. Robinette, D. Chabries and R.W. Christiansen, *Improvement in speech intelligibility in noise employing an adaptive filter with normal and hearing-impaired subjects*, Jour. Rehab. Res. and Dev. **24**, 75-86, (1987).
9. D. Byrne and H. Dillon, *The National Acoustics Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid, Ear and Hearing*, vol. **7**, 257265, 1986.
10. D. Chabries, R.W. Christiansen, R. Brey, M. Robinette, R. Harris, *Application of adaptive digital signal processing to speech inhancement for the hearing impaired*, Jour. Rehab. Res. and Dev., **24** 65-74, (1987).
11. H. Chipman, E. Kolaczyk and R. McCulloch, *Signal de-noising using adaptive Bayesian wavelet shrinkage*, Proc. IEEE-SP Intl. Symp. Time-Freq, Time Scale Anal., 225-228, (1996).
12. R. Coifman and D. Donoho, *Translation-invariant de-noising*. In A. Antoniadis, editor, *Wavelets and Statistics*, Springer-Verlag.
13. R. Coifman, M. Maggioni, *Diffusion wavelets*, Appl. Comput. Harmon. Anal. **21**, 53-94, (2006).
14. R. Coifman and F. Majid, *Adaptive waveform analysis and denoising*. In Y. Meyer and S. Roques, editors, *Progress in Wavelet Analysis and Applications*, 63-76, (1993).
15. R. Coifman and N. Saito, *Local discriminant bases and their applications*, J. Math. Imag. Vision **5** (1995) 337-358.
16. R. Coifman and V. Wickerhauser, *Entropy-based algorithms for best basis selection*, IEEE Trans. Inf. Theory **38**, (1992), 713-738.
17. I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Comm. on Pure and Appl. Math. **4**, (1988) 909-996.
18. D. Donoho, *Unconditional bases are optimal bases for data compression and for statistical estimation*, Appl. Comput. Harmonic Analysis **1**, (1993) 100-115.
19. D. Donoho and I. Johnstone, *Ideal spatial adaptation via wavelet shrinkage*, Biometrika **81**, 425-455, (1994).

20. D. Donoho, I. Johnstone and G. Kerkyacharian and D. Picard *Wavelet Shrinkage: Asymptopia?* J. Royal Stat. Soc. Ser. B. **2** 301-337, (1995)..
21. A.J. Duquesnoy and R. Plomp, *The effect of a hearing aid on the speech-reception threshold of hearing-impaired listeners in quiet and in noise*, Jour. Acoust. Soc. Amer., vol. **83**, pp. 2166-2173, 1983.
22. L.A. Drake, J.C. Rutledge and J. Cohen, *Wavelet Analysis in Recruitment of Loudness Compensation*, IEEE Transactions on Signal Processing, Dec.1993.
23. C.W. Dunnett, *A multiple comparison procedure for comparing several treatments to a control*, Jour. Amer. Stat. Assoc., vol. **50**, 1096-1121. 1955.
24. P. Duchnowski and P.M. Zurek, *Villchur revisited: another look at automatic gain control of simulation of recuiting hearing loss*, J. acoust. Soc. Amer. **98**, 6, pp. 3170-3181 (1995).
25. Y. Ephraim and D. Malah, *Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator*, IEEE ATrans. Accoust. Speech Sig. Proc. 32, (1984) 1109-1122
26. Y. Ephraim, D. Malah and B. Juang, *On the applications of hidden Markov models for enhancing noisy speech*, IEEE ATrans. Accoust. Speech Sig. Proc. **32**, (1984) 1846-1856.
27. Y. Ephraim and H.L. Van Trees, *A signal subspace approach for speech enhancement*, Proc. ICASSP 1993, volume **2**, (1993) 355-358.
28. Y. Ephraim and H.L. Van Trees, *A signal subspace approach for speech enhancement*, IEEE Trans. Speech Audio Proc. **3**, (1995) 251-266.
29. Y. Ephraim and H.L. Van Trees, M. Nilsson and S. Soli, Enhancement of noisy speech for for the hearing impaired using the signal subspace approach. In Proc. NIH Interdisciplinary Forum on Hearing Aid Research and Development (1996).
30. D. Graupe, J. Grosspietsch, and S. Basseas, *A single-microphone-based self-adaptive filter of noise from speech and its performance evaluation*. Jour. Rehab. Res. Dev. **24**, (1987) 119-126.
31. R. Gray, *On the asymptotic eigenvalue distribution of Toeplitz matrices*, IEEE Trans. Inf. Theory **18**, (1972) 725-730.
32. J.P. Gagne, *Excess masking among listeners with a sensorineural hearing loss*, Jour. Acoust. Soc. Amer., vol. **81**, pp. 2311-2321, 1988.
33. J. Greenberg and P. Zurek, *Evaluation of an adaptive beamforming method for hearing aids*, J. Acoust. Soc. Amer. **91**, 1662-1676, (1992).
34. C.S. Hallpike and J.D. Hood, *Observation upon the neurological mechanism of the loudness recruitment phenomenon*, Acta Oto-Laryng. **50**, pp. 472-486, (1959).
35. IEEE, *IEEE recommended practice for speech quality measurements*, IEEE Trans. Audio Electroacoustics, pp. 227-246, (1969).
36. A. Jansen and P. Niyogi, *Intrinsic Fourier analysis on the manifold of speech sounds*, Proc. International Conferene on Accoustics, Speech, and Signal Processing, Toulose, France, 2006.
37. M. Jansen, *Noise reduction by wavelet thresholding*, Springer Verlag, (2001).
38. J. Kates, *Speech enhancement based on a sinusoidal model*, J. Speech Hear. Res. **37**, 449-464, (1994).
39. M.C. Killion, *The K-Amp hearing aid: an attempt to present high fidelity for persons with impaired hearing*, Amer. J. Audiology, pp. 52-74, (1993)
40. S. Kochkin, *MarkeTrak III: Why 20 million in US don't use hearing aids for their hearing loss*, Hearing J. **46**, pp. 1-8, (1993).
41. S. Kochkin, *MarkeTrak III: 10-year Customer satisfaction trends in the US Hearing Instrument Market*, Hearing Rev. **9**, pp. 1-8, (2002).
42. S. Kochkin, *MarkeTrak VII: Customer satisfaction with hearing instruments in the digital age*, Hearing J. **58**, pp. 30-39, (2005).
43. H. Levitt, *A historical perspective on digital hearing aids: how digital technology has changed modern hearing aids*, Trends in Amplification **11**, pp. 7-24, (2007).
44. M. Lang, H. Guo, J. Odegard, J. Burrus and R. Wells, *Non-linear processing of a shift invariant DWT for noise reduction*, IEEE Sig. Proc. Letters **3**, 10-12, (1995).

45. H. Levitt, M. Bakke, J. kates, A. Neuman, T. Schwander and M. Weiss, *Signal processing for hearing impairment*, Scand. Audiol. **Supplement 38**, 7-19, (1993).
46. J. Lim, editor, *Speech Enhancement*, Prentice Hall, aenglewood Cliffs, N.J. (1983).
47. J. Lim and A. Oppenheim, *All-pole modeling of degraded speech*, IEEE Trans. Acoust. Speech Sig. Proc. **26**, 197-209, (1978).
48. H. Levitt and S.B. Resnick, *Speech reception by the hearing impaired: methods of testing and the development of new tests*, Scand. Audiol. Supp. **6**, pp. 107-130, (1978).
49. J. Makhoul and R. McAulay, *Removal of noise from noise-degraded speech*, Technical report, National Academy of Sciences, National Academy Press, Washington, D.C. (1989).
50. S. Mallat, *A theory for multiresolution signal decomposition: the wavelet representation*, IEEE Trans. Pattern Anal. and Machine Intell. **11**, 674-693, (1989).
51. R. McAulay and M. Malpass, *Speech enhancement using a soft-decision noise suppression filter*, IEEE Trans. Acoust. Speech Sig. Proc. **28**, 137-145, (1980).
52. E.S. Martin and J.M. Pickett, *Sensironeural hearing loss and upward spread of masking*, Jour. Speech Hear. Res., vol. **13**, pp. 426-437, 1970.
53. J.C. Pesquet, H. Krim and H. Carfantan, *Time-invariant orthonormal wavelet representations*, IEEE Trans. Sig. Proc. **44**, (1996) 1964-1970.
54. P. Peterson, N. Durlach, W. Rabinowitz and P. Zurek, *Multimicrophone adaptive beamforming for interference reduction in hearing aids*, J. Rehab. Res. Dev. *24*, 102-110, (1987).
55. J. Porter and S. Boll, *Optimal estimators for spectral resoration of noisy speech*. In Proc. ICASSP 1984, volume **1**, page 18A.2.
56. S. Quackenbush, T. Barnwell and M. Clements, *Objective measures of Speech Quality*, Prentice Hall, Englewood Cliffs, N.J. (1993).
57. J.C. Rutledge, *Speech Enhancement for Hearing Aids*, in Time-Frequency and Wavelet Transforms in Biomedicine, Metin Akay, Editor, IEEE Press, 1997.
58. J. Rissanen, *Modeling by shortest data description*, Automatica **14**, 465-471, (1978).
59. T. Roos, P. Myllymaki and J. Rissanen, *MDL Denoising Revisited*, IEEE Transactions on Signal Processing, vol. **57**, No. 9, 3347-3360, (2009).
60. N. Saito, *Simultaneous noise suppression and signal compression using a library of orthonormal and signal compression using a library of orthonormal bases and the minimum description length criterion*. In E. Foufoula-Georgiou and P. Kumar, editors, Wavelets in Geophysics, academic Press (1994).
61. C. Sammeth and M. Ochs, *A review of current "noise reduction" hearing aids: rationale, assumptions and efficacy*, Ear and Hearing **12**, 116S-124S (1991).
62. T. Schwander and H. Levitt, *Effect of two-microphone noise reduction on speech recognition by normal-hearing listeners*, Jour. rehab. res. and Dev. **24**, (1987) 87-92.
63. R. Schwartz, M. Berouti and J. Makhoul, *Enhancement of speech corrupted by acoustic noise*, ICASSP-79 Proceedings, page 208 (1979).
64. D. Sinha and A. Tewfik, *Low bit rate transparent audio compression using adapted wavelets*, IEEE Trans. Sig. Proc. **41**, 3463-3479, (1993).
65. M. Smith and T. Barnwell, *Exact reconstruction techniques for tree-structured subband coders*, IEEE Trans. Acoust.Sig. Proc. **34**, 434-441, (1986).
66. C. Stein, Estimation of the mean of a *multivariate normal distribution*, Ann. Stat. **9**, 6, pp.1135-1151, (1981).
67. R. Tyler and F. Kuk, *The effects of "noise suppression" hearing aids on sonsonant recognition in speech-babble and low-frequency noise,*. Ear and hearing **10**, 243-249, (1989).
68. A.R. Thornton and M.J.M. Raffin, *Speech discrimination scores modeled as a binomial variable*, Jour. Speech Hear. Res., vol. **21**, pp. 507-518, 1978.
69. B. Vidakovic, *Non-linear wavelet shrinkage with Bayes rules and Bayes factors*, Discussion Paper 94-24, Duke University (1994).
70. J. Verschure and P.P.G. Van Benthem, *Effect of hearing aids on speech perception in noisy situations*, Audiology, vol. **31**, pp. 205-221, 1992.
71. W. Voiers, *Diagnostic acceptability measure for speech communication systems*. In Proc. ICASSP 1977, pages 204-207, (1977).

72. S. Watanabe, *Karhunen-Loeve expansion and factor analysis: Theoretical remarks and applications*, Trans. 4th Prague Conf. Inform. Theory, Statist. Decision Functions, rand. Proc., 635-660, Prague Publishing House of the czechoslovak Academy of Sciences, (1967).
73. N. Whitmal, *Wavelet-Based Noise reduction for Speech Enhancement*, PhD thesis, Northwestern University, Evanston, Il. (1997).
74. N. Whitmal, J. Rutledge and J. Cohen, *Reducing correlated noise in digital hearing aids*, IEEE Eng. Med. Biol. Mag. 15, 88-96, (1996).
75. N. Whitmal, J. Rutledge and J. Cohen, *Reduction of autoregressive noise with shift-invariant wavelet packets*, Proc. IEEE-SP Symp. Time-Freq. Time-Scale Analysis, pp. 137-140, (1996).
76. G. Wornell (1990). *A Karhunen-Loeve-like expansion for 1/f processes via wavelets*, IEEE Trans. Inf. Theory 36, 859-861, (1990).
77. N. Whitmal and A. Vosoughi, Recruitment of loudness effects of attenuative noise reduction algorithms, J. Acoust. Soc. Amer. vol. **111,** Issue 5, p. 2380, (2002), (Conference Proceedings).