Wavelet Analysis in Recruitment of Loudness Compensation

Laura A. Drake, *Student Member, IEEE,* Janet C. Rutledge, *Member, IEEE,* and Jonathan Cohen

***Abstract*—Wavelet-based multiband dynamic range compression is developed to compensate for a common hearing impairment known as recruitment of loudness. The algorithm combines standard compression with intensity-level dependent gain calculation. Complexity and performance are similar to traditional techniques. Further, the methods established here can be applied to the more adaptive wavelet packets and local cosine bases which model the speech signal more closely.**

I. Introduction

In this work, we consider the efficacy of using a wavelet representation of speech in a compensation algorithm for the hearing-impaired. The particular hearing impairment that we address is recruitment of loudness. Although many previous approaches have been used to compensate for recruitment of loudness, none are consistently good for the full range of sounds found in a language like English for all listening environments [1]. Preliminary experiments with wavelet-based amplitude compression produced results which are comparable to previous compensation schemes. The wavelet-based methods developed here can be adapted to wavelet packets and local cosine bases which allow more flexible choice of frequency bands and better matching of the speech waveform.

A. Recruitment of Loudness

Recruitment of loudness commonly occurs in conjunction with sensorineural hearing impairments which are caused by prolonged or excessive exposure to noise, Ménière's syndrome, congenital or hereditary factors, and aging [2]. Here, the word *recruitment* means a growth or increase. By our nomenclature, *loudness* refers to the subjective magnitude of sound; intensity refers to the measured power of the acoustic signal. If a listener suffers from recruitment of loudness, perceived loudness grows more rapidly with an increase in sound intensity than it does in the normal ear.

Manuscript received September 1, 1992; revised June 10, 1993. The Guest Editor coordinating the review of this paper and approving it for publication was Dr. Ahmed Tewfik. This work was supported in part by the National Science Foundation.

L. A. Drake and J. C. Rutledge are with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208-3118.

J. Cohen is with the Department of Mathematics, DePaul University, Chicago, IL 60614.

IEEE Log Number 9212191.

For the purposes of building compensation systems, it has been shown that recruitment of loudness can be modeled as an expander followed by an attenuator [3]. The expander causes speech to cover a greater portion of a normal listener's dynamic range of hearing. The attenuator simulates raising the threshold of hearing causing low intensity speech to be below the threshold of hearing.

Equal loudness curves are functions mapping loudness to intensity at each frequency such that any two tones played at the associated intensities along a given curve will sound equally loud. The thresholds of hearing and pain bound the range of sounds that a person can hear. This range is called the *dynamic range of hearing*. For normal-listeners, equal increments in sound intensity produce equal increments in loudness perception uniformly across all frequencies of importance to speech signals. Since the threshold of hearing is usually raised with recruitment of loudness while the threshold of pain may remain constant or even be lowered, the dynamic range of hearing is reduced. Thus, the loudness curves are compressed together, causing relatively small changes in intensity to give larger corresponding changes in perceived loudness.

It is possible to simulate the dominant effects of recruitment of loudness in normal-hearing listeners by adding background noise to the audio signal. Models of masking effects often use noise to approximate the contributions made by the inner ear malfunctions [4]. Thus, compensation algorithms can be tested with normal-hearing listeners by masking the compensated speech with appropriately shaped background noise.

B. Traditional Compensation Techniques

A variety of techniques have been used to compensate for recruitment. These fall under two general categories: linear amplification with automatic gain control and amplitude compression with equalization [1]. In either category, compensation algorithms can have one or many bands.

1) Linear Amplification: Linear amplification with automatic gain control has shown the best performance under the controlled conditions of laboratory tests [1]. The base amplification level is set to raise as much of the speech signal above the threshold of hearing as possible. The automatic gain control keeps the signal below the threshold of pain. In a study performed by Lippmann *et*

1053-587X/93$03.00

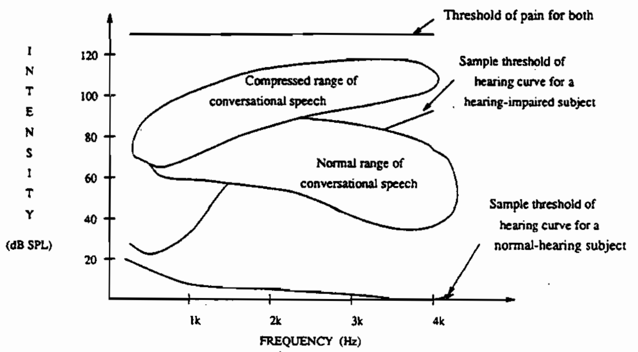


Fig. 1. Typical hearing threshold and discomfort threshold for a subject having a high frequency sloping sensorineural hearing loss with recruitment of loudness. The range of conversational speech is shown before and after compression processing.

al. [1], amplitude compression was shown to perform better when the input speech had a low overall intensity and/or had variations in intensity level that were closer to normal conversational speech than the prerecorded speech that is often used in lab tests. Since this prerecorded speech has a limited dynamic range, it can be considered equivalent to precompressed speech. Amplitude compression was also shown to compensate better for subjects with significant losses of dynamic range [5].

2) Amplitude Compression: Compression with equalization was first proposed by Villchur who showed that expansion followed by attenuation was a reasonable model of recruitment of loudness [3]. Thus, this compensation technique was designed specifically to invert the effects of recruitment of loudness. Its effect is to map speech from the normal conversational speech range into the reduced range of an impaired listener so that the impaired listener will hear a sound at a comparable loudness to that at which the normal listener hears the uncompressed sound. Fig. 1 shows the normal range of conversational speech and the compressed range superimposed on typical hearing profiles of normal-hearing and hearing-impaired subjects. With most schemes for implementing compression, the parameters are chosen based on normal conversational speech and remain fixed over time. Low level input may not receive enough boost, and high intensity sounds may be severely clipped. Therefore, a compensation scheme which adapts to calculate the amount of compression gain based on input should perform better.

3) Parametric Compression: Compression of speech parameters can also serve to compensate for recruitment of loudness. An interesting example, and the one that is the basis for the wavelet-based algorithm, is TVFD (time-varying frequency-dependent) processing [6]. This processing is based on a sinusoidal model of speech [7] in which speech is represented as the sum of sinusoids with various amplitudes, frequencies and phases. The components are calculated from the short-time Fourier transform using a 20 msec window. TVFD is a digital, multiband compression system in which each sinusoidal component corresponds to a separate frequency band. The sinusoidal amplitude parameters are given compression gain based on an intercomponent masking model of both normal-hearing and hearing-impaired listeners. The masking model incorporates the listener's thresholds of hearing and pain. Calculating the gain for each frequency component based on preserving the relative distance above masked threshold allows the algorithm to adjust to the time-varying qualities of the input speech.

II. Wavelet-Based Compression

Wavelet-based compression uses the TVFD-style gain calculation, i.e., gain is calculated separately for each wavelet coefficient based on its intensity level, with the wider bandwidths of traditional multiband amplitude compression. The parameters of this parametric compression algorithm are wavelet coefficients. As a compression technique, the strategy of wavelet-based compression remains to invert the expander/attenuator model of recruitment of loudness.

The wavelet analysis/synthesis in this algorithm is performed via the multiresolution analysis wavelet decomposition/reconstruction algorithm developed by Mallat [8]. It is also called a fast wavelet transform (FWT) [9]. Multiresolution analysis decomposes a signal onto a hierarchical system of subspaces of $L^2(\boldsymbol{R})$ consisting of a multiresolution subspace and an orthogonal subspace at each resolution level. At the m^{th} level, the multiresolution space, V_m, is spanned by the basis functions $\{2^{m/2}\ \phi(2^m x - n);\ n \epsilon \boldsymbol{Z}\}$ and the space, W_m, orthogonal to V_m in V_{m+1} is spanned by $\{2^{m/2}\ \psi(2^m x\ -\ n);\ n \epsilon \boldsymbol{Z}\}$, where $\phi(x)$ is called the scaling function and $\psi(x)$ is called the wavelet function. Mallat's algorithm allows wavelet coefficients, $c_{mn} = \langle f(t), \psi_{mn} \rangle$, and scaling coefficients, $d_{mn} = \langle f(t), \phi_{mn} \rangle$ at the m^{th} scale to be calculated recursively from the representation of the speech signal, $f(t)$, at the preceding, finer scale, $d_{m+1,n}$, through the following filtering operation:

$$d_{mn} = \sum_{k=-\infty}^{\infty} \tilde{h}(2n - k)\, d_{m+1,k}$$

$$c_{mn} = \sum_{k=\infty}^{\infty} \tilde{g}(2n - k)\, d_{m+1,k}$$

where

$$\begin{aligned} \tilde{h}(n) &= h(-n) \\ h(n) &= \langle \phi_{-1,0}, \phi_{0,n} \rangle \\ \langle \phi_{-1,0}, \phi_{0,k-2n} \rangle &= \langle \phi_{mn}, \phi_{m+1,k} \rangle \end{aligned} \tag{1}$$

and

$$\begin{aligned} \tilde{g}(n) &= g(-n) \\ g(n) &= \langle \psi_{-1,0}, \phi_{0,n} \rangle \\ \langle \psi_{-1,0}, \phi_{0,k-2n} \rangle &= \langle \psi_{mn}, \phi_{m+1,k} \rangle. \end{aligned}$$

The synthesis algorithm recursively regenerates the $d_{m+1,n}$ coefficients from the coefficients at the next coarser level, d_{mn} and c_{mn}. Note, $\langle f(t), \phi_{0,n} \rangle$ is the original sam-

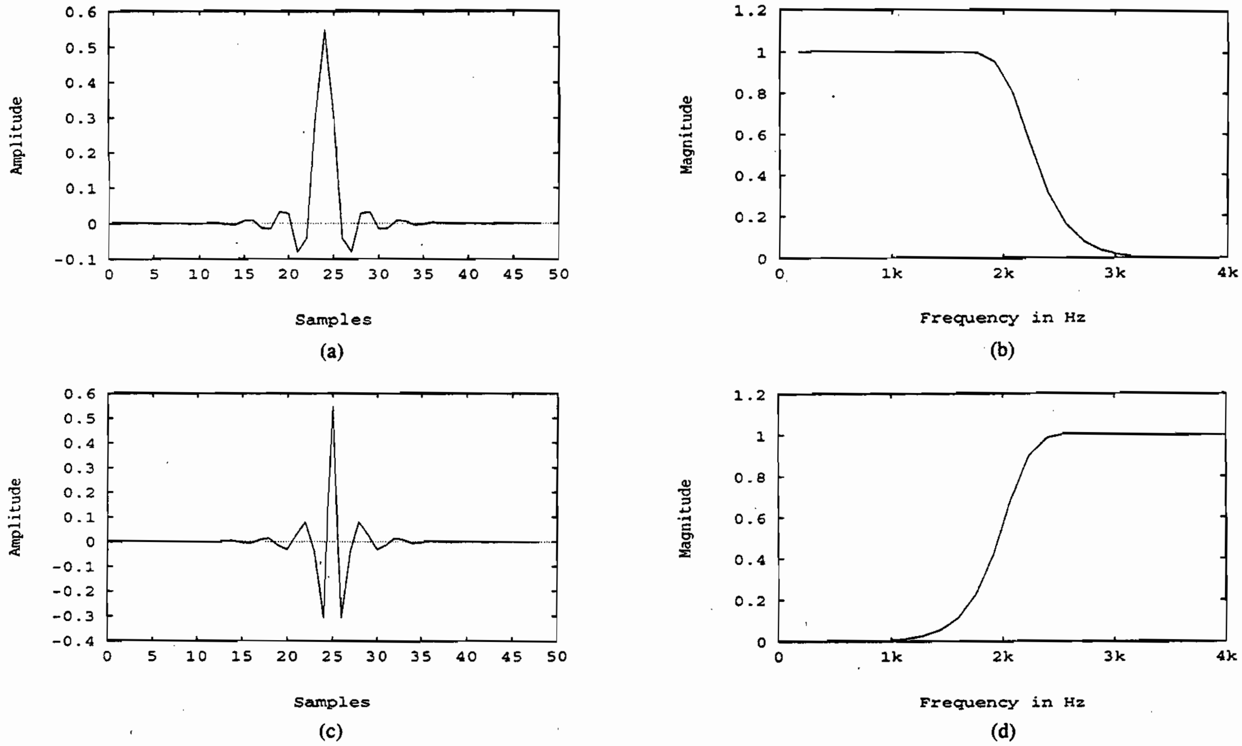


Fig. 2. (a) The impulse response, $h(n)$, of the filter corresponding to the scaling function, $\phi(x)$. The sampling rate is 8 kHz. (b) The transfer function of $h(n)$. (c) The impulse response, $g(n)$, of the filter corresponding to the wavelet function, $\psi(x)$. The sampling rate is 8 kHz. (d) The transfer function of $g(n)$.

pled speech signal and $m = 0, -1, -2, \cdots$ in this paper. The nth coefficient at the $m + 1^{st}$ scale is

$$d_{m+1,n} = \sum_{k=-\infty}^{\infty} h(n - 2k)\, d_{mk} + \sum_{k=-\infty}^{\infty} g(n - 2k) c_{mk},$$

where $h(n)$ and $g(n)$ are the impulse responses of quadrature mirror filters. The filters used for the testing of the wavelet-based compression algorithm are those used in Mallat's algorithm [8]. Their impulse responses were truncated to 49 samples. They are derived from the scaling function, $\phi(x)$, of a class of multiresolution approximations of $L^2(\boldsymbol{R})$ studied by Lemarie [10] and Battle [11]. $\phi(x)$ is a continuously differentiable and exponentially decreasing function. For details about this scaling function, refer to [8, Appendix A]. Fig. 2 shows the impulse responses and the corresponding transfer functions of the filters used in this implementation. The final filter bank, corresponding to two levels of multiresolution decomposition, is shown in Fig. 3.

The results of FWT analysis/synthesis are shown in Fig. 4. Fig. 4(a) shows the original time waveform of the sentence, *Cats and dogs each hate the other*. The same sentence after analysis/synthesis is shown in Fig. 4(b).

To compensate for recruitment of loudness, compression is applied to each wavelet coefficient c_{mn} and then the signal is reconstructed from a wavelet series with the modified coefficients. Gain is applied in such a way as to amplify the coefficient from a given equal loudness curve in the normal-hearing person's hearing profile to the corresponding equal loudness curve in the hearing-impaired

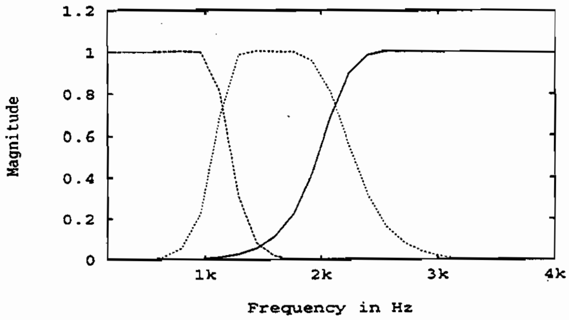


Fig. 3. Filters corresponding to two levels of wavelet analysis.

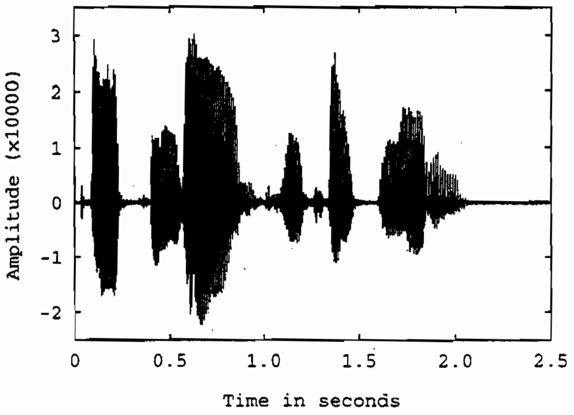
person's profile. Since each wavelet coefficient corresponds to a band of frequencies, the equal loudness curves have been arranged within each band to also give one value per band.

Gain is calculated for each wavelet coefficient in each band such that the ratio of log intensity above hearing threshold to dynamic range of hearing is the same for the hearing-impaired listener as the corresponding ratio is for the normal-hearing listener. In other words, the compression gain is found so the desired relation specified in [2] and pictured in Fig. 5 holds. Mathematically,

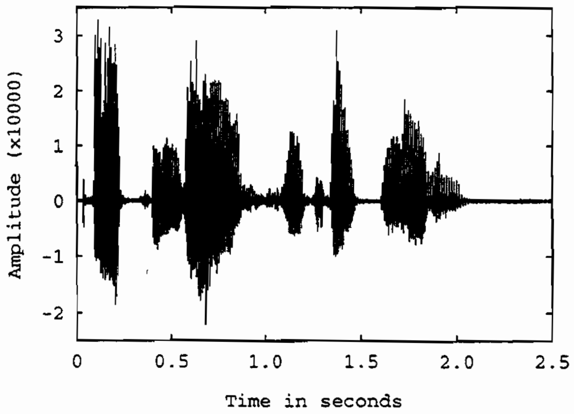
$$\frac{\delta^*}{\delta} = \frac{\Delta^*}{\Delta} \tag{2}$$

where

- $\delta = C_{mn} - T_{nor}$ = the distance the unamplified wavelet coefficient is above a normal hearing person's threshold of hearing (in dB).



(a)



(b)

Fig. 4. (a) Original time waveform of the sentence, *Cats and dogs each hate the other*. (b) Time waveform of the same sentence processed using wavelet analysis/synthesis with no compression.

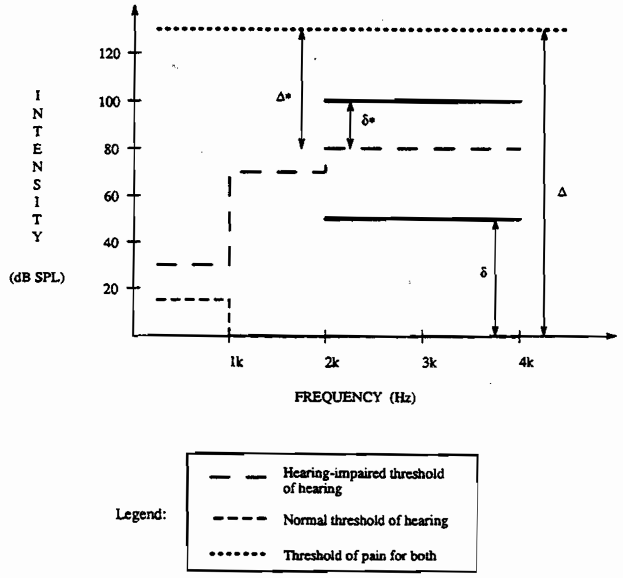


Fig. 5. Calculation of gains based on the distance the wavelet coefficient is above the approximated threshold of hearing for the normal and impaired listener.

• δ^* = the distance the compensated wavelet coefficient is above a hearing-impaired person's threshold of hearing (in dB).

• $\Delta = T_{pain} - T_{nor}$, the normal-hearing person's dynamic range of hearing (in dB).

• $\Delta^* = T_{pain} - T_{im}$, the hearing-impaired person's dynamic range of hearing (in dB).

• $C_{mn} = 20 \log c_{mn}$.

• T_{im} = the hearing-impaired person's threshold of hearing (in dB).

• T_{nor} = the normal-hearing person's threshold of hearing (in dB).

• T_{pain} = the threshold of pain for both (in dB).

With the above-defined relation, compression gain for a coefficient can be derived as follows. Let $C^*_{mm} = 20 \log c^*_{mn}$ be the compensated coefficient in dB. Then,

$$C^*_{mn} = T_{im} + \delta^* = T_{im} + \delta \frac{\Delta^*}{\Delta}. \tag{3}$$

Substituting the definition of δ into (3)

$$C^*_{mn} = T_{im} + (C_{mn} - T_{nor}) \frac{\Delta^*}{\Delta}. \tag{4}$$

Define the thresholds t_{im} and t_{nor} such that $T_{im} = 20 \log t_{im}$ and $T_{nor} = 20 \log t_{nor}$, and substitute into (4) to find the compensated coefficient

$$20 \log c^*_{mn} = 20 \log t_{im} + 20 \frac{\Delta^*}{\Delta} \log c_{mn} - 20 \frac{\Delta^*}{\Delta} \log t_{nor}.$$

Thus, the amplitude compression operator maps the coefficient, c_{mn} to the compressed coefficient, c^*_{mn} as follows:

$$c^*_{mn} = t_{im} \left(\frac{c_{mn}}{t_{nor}} \right)^{\Delta^*/\Delta}. \tag{5}$$

The transformation $c_{mn} \rightarrow c^*_{mn}$ defined by [5], could be applied anytime the speech signal is represented as a superposition of basis functions, all of which are reasonably well localized in frequency. For example, this transformation is used in TVFD where speech is modeled as a linear combination of sinusoids. With TVFD the thresholds include the effects of intercomponent masking. The transformation could be applied equally well to a wavelet packet representation of the speech signal or a superposition of local cosine functions.

Upon reconstruction of the compensated wavelet coefficients similar results are achieved to amplitude compression. This is illustrated in Fig. 6 which shows the original spectrum for the vowel /ʌ/ in the word *other* along with the vowel using the wavelet-based compensation technique and multiband amplitude compression for a flat-loss subject. With both compensation techniques the general formant structure is maintained.

III. Comparison with Other Techniques

As discussed in the previous section, this wavelet-based processing technique bears a close resemblance to multiband amplitude compression. Therefore, as a preliminary

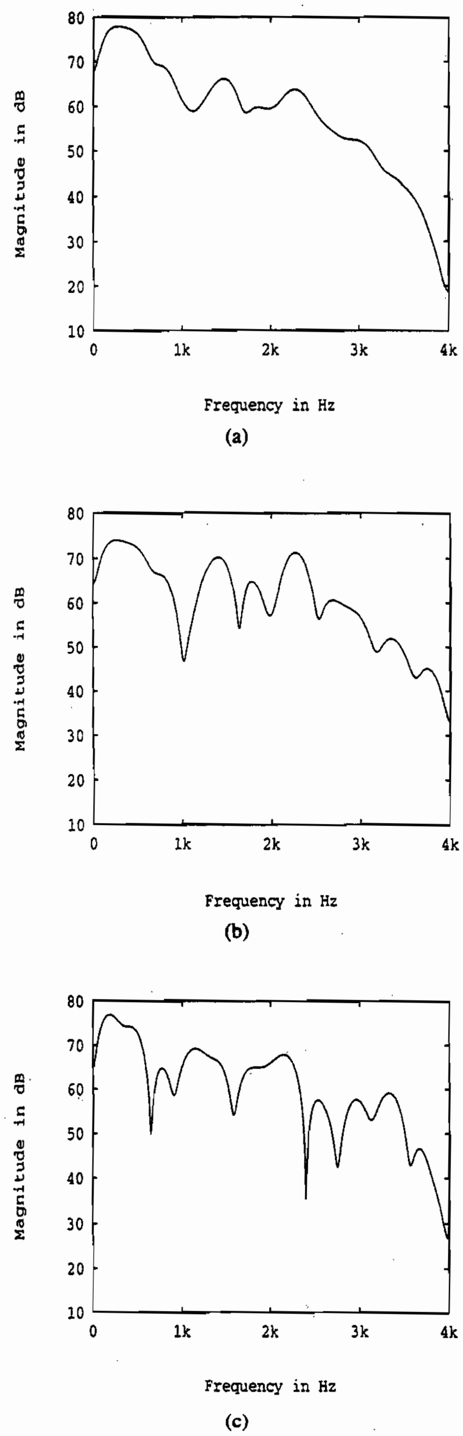


Fig. 6. Spectra of the vowel /ʌ/ based on an FFT magnitude with a 49 sample Hanning window. (a) Original vowel spectrum. (b) Wavelet-based compensation vowel spectrum. (c) Amplitude compression vowel spectrum.

evaluation, this new technique was compared to compression from both a subject-based performance and an implementation perspective. The compression algorithm used 2 : 1 compression in each band. The bands were the same as those with the wavelet processing: 0–1000 Hz, 1000–2000 Hz and 2000–4000 Hz which are shown in Fig. 3. A summary of those results is presented here.

A. Subject-Based Testing

Listening tests were conducted using four normal-hearing subjects between the ages of 30 and 50 years with noise masking to simulate a flat-loss hearing impairment. The stimuli were sentences from the 500 Harvard Sentences listed in the *IEEE Recommended Practice for Speech Quality Measurements*. Each sentence has five key words to test for correctness. This provides analysis of the speech reception abilities in conversational speech, where contextual information can be used.

Training consisted of ten sentences (50 words) presented at the most comfortable listening level for each processing technique. Subjects made little adjustment to listening level between techniques. Testing consisted of 20 sentences per system. Training and testing on one system was completed before proceeding to the next system for each subject. The listener was instructed to write down each sentence after it was played twice.

The results, given in Table I, show that performance on both multiband amplitude compression and the wavelet-based processing was similar. The average score on the compression processing was 80.6% words correct versus 78.1% words correct for the wavelet-based processing. Subject JC showed the most variation between systems with a score ten percentage points higher on the compression processing. TR showed almost no difference between systems. LD and EC had a six-percentage-point spread between systems, one favoring the wavelet processing and the other favoring compression. One comment made by TR was that he found the compression processing to be a little clearer, while the wavelet processing was more pleasant.

B. Implementation Differences

An obvious question in light of the above results is what advantage does wavelet processing have over traditional multiband filtering approaches. From a complexity perspective, both are fairly similar. In the processing stage before applying the compression gain, both mainly involve convolution. Four convolutions are needed to get three levels within the wavelet processing, while only three are needed to get the filter outputs. However, since with the wavelet processing the number of data points is halved at each resolution scale, so are the lengths of the convolutions. With a parallel bank of bandpass filters the amount of information (and thus memory) is increased in proportion to the number of bands used in the system, whereas the wavelet processing will have the same number of coefficients as data points. The computation of the gains is similar in complexity for the two techniques. During reconstruction, the wavelet processing requires additional convolutions; the multiband compression merely requires addition of the filtered time waveforms. On a DSP microprocessor, the additional convolutions of wavelet processing would not be as significant as the increased memory required by multiband compression. In relation to TVFD, processing complexity is lower. With

TABLE I
PERCENT CORRECT WORDS FROM SENTENCE TESTING

Subject	JC	TR	EC	LD	Avg
Compression	72%	88%	74.5%	88%	80.6%
Wavelet	62.5%	87%	81%	82%	78.1%

TVFD, a set of joint linear equations must be solved to model the intercomponent masking relationship.

An advantage to using wavelet processing is that the variable support of the wavelet in the time domain is dependent upon its scale. Thus, calculation of gain will be customized to the characteristics of the speech waveform in each frequency region. The gain for consonants and stops, which contain predominantly high frequencies, will be calculated from a smaller window than vowels, which contain more low frequencies. With multiband compression processing, gains are based on the energy envelopes which follow variations in the waveform according to predetermined time constants. With other parameterization techniques the window size is the same for each frequency band.

The calculation of the gains is designed so that each coefficient will be within the impaired listener's residual dynamic range regardless of the intensity level of the input speech. Multiband filter compression systems are time-invariant. They are based on the average intensity level of input speech and may have some reduction in performance for very low level inputs [1].

IV. CONCLUDING REMARKS

The wavelet-based multiband compression algorithmn is a viable alternative to traditional multiband filtering techniques. The structure lends itself quite well to this application since the wavelet decomposition provides access to both time-domain and frequency-domain information. While performance and complexity are comparable, the wavelet-based processing allows TVFD style customization of the gain calculation at each point. Furthermore, there are more sophisticated signal decomposition systems, such as wavelet packets and local cosine bases, which are more adaptive and could provide more efficient and flexible signal representations. Wavelet packets are especially well suited to this problem since it is possible to decompose the signal into frequency bands that are as fine as desired. Rather than being confined to the octave band format of the wavelet transform, compensation gain can be calculated more carefully in regions where the threshold of hearing changes rapidly. Also, experiments have shown that, for the same amount of data compression, wavelet packets result in less distortion. This suggests that the wavelet packets have a better ability to match the speech waveform. In this paper, a method of operating on the wavelet coefficients to perform multiband compression has been established. This method could be applied equally well to coefficients of the better adapted wavelet packets or local cosine bases.

ACKNOWLEDGMENT

The authors wish to thank Apichat Tungthangthum for his gracious and expert technical assistance during the preparation of this paper and the subject-based test materials. The authors also wish to thank the participants in the listening tests.

REFERENCES

[1] R. P. Lippmann, L. D. Braida, and N. I. Durlach, "Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss," *J. Acoust. Soc. Am.*, vol. 69, pp. 524–534, Feb. 1981.

[2] H. Davis and S. R. Silverman, *Hearing and Deafness*. New York: Holt, Rinehart and Winston, 4th ed., 1978.

[3] E. Villchur, "Simulation of the effect of recruitment on loudness relationships in speech," *J. Acoust. Soc. Am.*, vol. 56, pp. 1601–1611, Nov. 1974.

[4] L. E. Humes, B. Espinoza-Varas, and C. S. Watson, "Modeling sensorineural hearing loss. I. Model and retrospective evaluation," *J. Acoust. Soc. Am.*, vol. 83, pp. 188–202, 1988.

[5] D. K. Bustamante and L. D. Braida, "Principle-component amplitude compression for the hearing impaired," *J. Acoust. Soc. Am.*, vol. 82, no. 4, pp. 1227–1242, 1987.

[6] J. C. Rutledge and M. A. Clements, "Compensation for recruitment of Loudness in sensorineural hearing impairments using a sinusoidal model of speech," in *Proc. IEEE ICASSP*, 1991, pp. 3641–3644.

[7] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 744–754, Aug. 1986.

[8] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 11, pp. 674–693, July 1989.

[9] M. A. Cody, "The Fast Wavelet Transform," *Dr. Dobb's J.*, pp. 16-28, Apr. 1992.

[10] P. G. Lemarie, "Ondelettes a localisation exponentielles," *J. Math. Pures Appl.*, to be published.

[11] G. Battle, "A block spin construction of ondelettes, Part 1: Lemarie functions," *Commun. Math. Phys.*, vol. 110, pp. 601–615, 1987.

Laura Drake (S'91) received the B.A. degree from the Department of Mathematics the Johns Hopkins University, Baltimore, MD, and the M.S. degree from the Department of Electrical Engineering and Computer Science Northwestern University, Evanston, IL. She is currently completing the Ph.D. degree in the EECS department of Northwestern University.

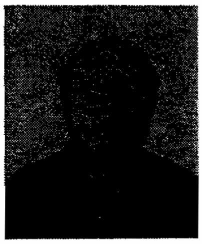
Her research interests include psychoacoustics and speech processing.

Janet C. Rutledge (S'81, M'90) received the B.S. degree in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, in 1983. She received the M.S. and Ph.D. in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, in 1984 and 1990, respectively.

Since August 1990, she has been an Assistant Professor in the Electrical Engineering and Computer Science Department at Northwestern University, Evanston, IL, with an adjunct appointment in the Audiology and Hearing Sciences Dept.

Her research interests include modeling and compensating for the effects of sensorineural hearing loss, speech recognition, speech enhancement, and audio signal processing.

Professor Rutledge is a member of the speech technical committee of the Acoustical Society of America and a Fellow of the Hugh Knowles Center for Clinical and Basic Science in Hearing and its Disorders at Northwestern University. She serves on the NIH NIDCD Ad Hoc Committee on Hearing Aid Research.



Jonathan Cohen was born in New York City. He received his Ph.D. in mathematics form Washington University, St. Louis, MO, in 1977.

He is currently an associate professor at DePaul University, Chicago, where he has taught since 1987. His research has been in the area of harmonic analysis and has included work on singular integrals, boundary value problems in partial differential equations, potential theory, uniqueness of multiple trigonometric series, and wavelet-based signal processing.

Professor Cohen is a member of the AMS and SIAM.