Canagarajah, C. N., & Rayner, P. (1991). A single-input hearing aid based on the auditory perceptual features to improve speech intelligibility in noise. In IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, 1991. (pp. 0_107 - 0_108). Institute of Electrical and Electronics Engineers (IEEE).

A Single-Input Hearing Aid Based on the Auditory Perceptual Features to Improve Speech Intelligibility in Noise

C. Nishan Canagarajah and Dr. Peter J.W.Rayner
University of Cambridge, Department of Engineering
Trumpington Street, Cambridge, England CB2 1PZ
Telephone: (0223)-332767 ; e-mail:cnc10@uk.ac.cam.phx

One of the main problems of the sensorineural hearing impaired listeners is the partial or complete loss of frequency selectivity. It is now well established that most of the auditory perceptual features are very well represented in the ear by the spectrum of the incoming sound signal. Thus the loss of frequency selectivity means that it is difficult for the impaired listener to discriminate between two sounds or to understand speech in a noisy environment. This handicap is referred to as the *cocktail party effect.* It is widely accepted, and proven by the experiments carried out on impaired listeners, that one of the main causes for this impairment is the broad and tilted auditory filter shapes in the damaged cochlea compared to an undamaged normal ear. As a result, in noisy surroundings these broad filters allow more noise than a normal ear making detection of signal in noise difficult. Therefore to improve intelligibility a hearing aid must, not only suppress the noise in speech but also alleviate the problems of reduced frequency selectivity.

There are a few hearing aids proposed in the literature to enhance speech in noise. Most of them are based on Adaptive noise cancellation or Adaptive beamforming principles. They have proved to be very useful in situations where there are few noise sources or when there is a reference noise available. Very often the environment contains many uncorrelated noise sources effectively creating a diffusive noise source. Hence obtaining a reference noise signal that is correlated with the noise in the other inputs is impossible. In these situations the above methods produce very little speech enhancement. There are many conventional single-input systems to suppress noise but like the multi-microphone methods mentioned above, they have proved to be of very little use in increasing the intelligibility of the speech for the hearing impaired. In this paper we illustrate how a single-input system incorporating the auditory perceptual features could be employed to increase intelligibility in hearing aids.

Spectral Subtraction (SS) is an efficient way of reducing noise in single-input systems. In this method an estimate of the magnitude spectrum of the noise, $\hat{N}(\omega)$, is obtained during nonspeech activity and is subtracted from the magnitude spectrum of the noisy speech, $X(\omega)$, to obtain the enhanced speech, $\hat{S}(\omega)$. This performs satisfactorily when the noise source is stationary. The main drawback of this system is it does not consider the problems of the hearing impaired and as a result is of very little benefit to them. Furthermore it introduces a ***residual or musical noise*** in the processed speech. It is shown in this paper that by incorporating the perceptual features like masking and excitation patterns the above problems can be eliminated.

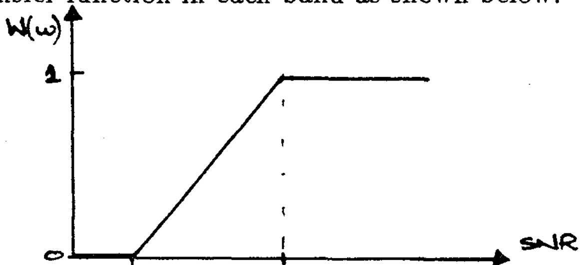
The technique proposed here, firstly transforms the power (not magnitude) spectrum of the noisy speech ($X(\omega)$) into auditory excitation patterns, $E(\omega)$. The auditory system consists of a bank of constant bandwidth band-pass filters on a logarithmic scale (Bark Scale). Excitation patterns are obtained by convolving the signal spectrum with these filters. $E(\omega)$ now represents the power spectrum of the signal as it would appear if it had been processed by a normal ear. These excitation patterns enable the hearing impaired to group frequency components quite successfully, thereby increasing their frequency selectivity in spite of the broad filters in their auditory system. This transformation also removes unwanted noise and speech that might have been processed by the impaired ear.

In conventional spectral subtraction the noisy signal spectrum is weighted based on the Signal-to-Noise Ratio (SNR) in each FFT bin. According to the present model the weighting, $W(\omega)$, is dependent on the SNR in each critical auditory filter band.

$$W(\omega) = f(\frac{E(\omega)}{\hat{N}(\omega)}) \tag{1}$$

It is postulated that the auditory system perceives sounds based on the SNR (or Signal to Masker Ratio) in each critical band. This concept is widely used in psychoacoustic experiments investigating the auditory perceptual features.

The key to the effectiveness of this processing comes from the ability to include the human *auditory masking* properties in the estimation of $W(\omega)$. Auditory Masking refers to the ability of the listener to attend to a signal in a critical band in the presence of other unwanted signals. In normal hearing subjects this property enables the speech to be intelligible even in low SNR situations. Thus masking enables the noise in a critical band to be suppressed when the signal contains relatively high power. Similarly if the noise in that band is very strong then the signal in that band will be inaudible. Therefore as far as the desired signal is concerned, in each critical band a unity weighting is applied when the SNR is very high and a zero weighting when the band contains just noise. Incorporating this knowledge in setting $W(\omega)$ produces a nonlinear saturation type transfer function in each band as shown below.



Hence when the SNR is very low the zero weight settings completely remove the noise in that band thus eliminating the residual noise that arise in conventional spectral subtraction systems. In addition, when estimating the excitation patterns local variations in the noise spectrum are averaged.

The above was implemented in real time on a DSP32C chip to reduce added white noise in the speech. The auditory filters were assumed to be symmetric with exponential slopes. The processing was performed on the power spectrum of the signal. Informal listening tests indicated that the processed output was more intelligible and of better quality than a conventional spectral subtraction processed output. Since the auditory system consists of overlapping filters the number of components in the excitation pattern was significantly reduced without a considerable loss in intelligibility. This meant that the proposed system produced an output of the same quality as that produced by a higher order conventional system, but with fewer excitation components. Results will be presented for different SNR conditions and in babble type noise to show the effectiveness of this approach.

The main conclusions reached in this paper include:

1. Using excitation patterns and masking properties of a normal ear improves intelligibility of the speech for the hearing impaired.
2. The excitation patterns can be derived with few spectral components thus reducing the complexity of the system.
3. Significant increase in intelligibility was obtained compared to conventional SS systems.
4. The proposed system does not introduce the distortions that occur in conventional systems.