Proceedings of the Institute of Acoustics

BINAURAL ENVELOPE EXPANSION FOR SPEECH DEREVERBERATION

P.D.Stringer and A.I.Tew

Signal Processing: Voice and Hearing Research Group,
Department of Electronics, University of York, Heslington, York, YO1 5DD.

ABSTRACT

The effects of room reverberation on speech intelligibility are well studied. This paper argues that it may be beneficial for a speech dereverberation algorithm to preserve binaural cues as far as possible along the signal path, possibly as far as the listener. Such a binaural dereverberation scheme is discussed which is based on the method of envelope expansion.

1. INTRODUCTION

It has been shown that reverberation has a deleterious effect upon speech intelligibility [1,2,3,4]. Many signal processing algorithms have been investigated which attempt to remove the effects of reverberation from speech, for example [5,6], but these techniques often require *a priori* knowledge of the room acoustics. For a head worn device, acoustic parameters would change quite rapidly, and in this case fine measurement of the room acoustics is not practical. Instead, techniques must be developed that require little or no *a priori* knowledge of the room acoustics, but are capable of preserving binaural cues.

When listening to a speaker in a room, an observer first receives the direct sound followed by a large number of reflections, arriving as echoes from the different surfaces in the room. To the observer the sound appears to come from the direction of the speaker, even though most of the received energy is a result of reflections in the room. In determining the direction of a sound source in a room the ear apparently makes use of the amplitude and time differences of the first echoes of every speech sound [7]. Echoes arriving within a certain period are integrated with the direct sound and increase the speech intelligibility. Echoes arriving after this period are considered as masking sounds and may well overlap with other speech sounds, causing a lowering of speech intelligibility [1]. For binaural listening in reverberation Moncur and Dirks [1] showed that there was on average a +10dB intelligibility improvement over monaural near-ear listening for several reverberation times, although scores were similar for the anechoic situation. It may be argued therefore that there are considerable advantages to be gained from developing a binaural signal processing system which can reduce the deleterious effects of reverberation and present a binaural output to the listener. The work of Nabelek and Mason [8] and Nabelek and Pickett [9] showed that even people with various degrees of hearing impairment can still obtain a binaural advantage when listening in reverberation.

This paper describes a binaural speech dereverberation technique which reduces the effects of room reverberation whilst preserving cues that are important for binaural listening.

2. BINAURAL ENHANCEMENT SCHEME

In terms of the modulation transfer function [10]; when a signal is passed through an enclosure, the envelope of the received signal is, in general, a smoothed version of the original envelope; the degree of smoothing being dependent upon the room acoustics. It is reasonable to conclude therefore, that attempting to increase the modulation depth of the processed signal may increase intelligibility. The envelope expansion scheme discussed below is designed to achieve such an increase.

The proposed binaural speech enhancement scheme is shown in Figure 1. The scheme consists of two identical structures, each processing the signals received in one of the channels. The reverberant signal received by the left channel, $S_L(n)$, is first filtered into N uniform contiguous frequency bands spanning the frequency range $0 - F$ Hz. If we consider the kth band, the output signal from the kth filter, $s_k(n)$, can be represented as:

$$s_k(n) = A_k(n)\cos(\theta_k(n)) \tag{1}$$

where $A_k(n)$ is a slowly varying envelope function and $\theta_k(n)$ is an instantaneous phase function [11].

The envelope function $A_k(n)$ is determined by the use of the analytic signal $x_k(n)$, of $s_k(n)$, such that:

$$x_k(n) = s_k(n) + j\hat{s}_k(n) \tag{2}$$

where $\hat{s}_k(n)$ is the Hilbert transform of $s_k(n)$.

The Hilbert transform frequency response $H(e^{j\omega})$ is defined as

$$H(e^{j\omega}) = \begin{cases} -j & \text{if } \omega \geq 0 \\ +j & \text{if } \omega < 0 \end{cases} \tag{3}$$

This transfer function represents a $\pm 90°$ phase shift, with unity gain at all frequencies. Since $s_k(n)$ is real,

$$\hat{s}_k(n) = A_k(n)\sin\theta_k(n) \tag{4}$$

Substituting Equations 1 and 4 into Equation 2, the envelope function $A_k(n)$ can be determined,

$$A_k(n) = \sqrt{s_k(n)^2 + \hat{s}_k(n)^2} \tag{5}$$

The resulting envelope is then subjected to a non-linear expansion, by raising it to a power v, where $v > 1$, such that $\hat{A}_k(n) = A_k^v(n)$. Following this expansion the envelope is low-

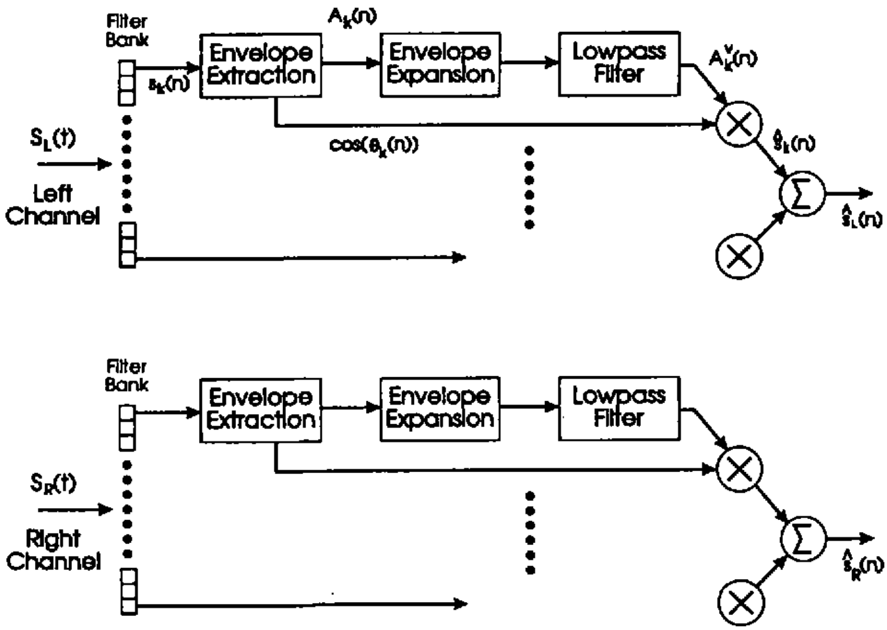

Figure 1: Binaural Speech Enhancement Scheme based on Envelope Expansion

pass filtered. The inclusion of the low-pass filter after the expansion is intended to remove harmonic components arising from the envelope expansion. After filtering, the envelope is recombined with the degraded phase and then summed with the contributions from the other bands.

2.1 Effects of binaural envelope expansion on sound localisation

In order for the binaural expansion scheme to preserve cues important for sound localisation, two potential problems need to be addressed:

- Harmonics introduced into the reconstructed waveform due to the envelope expansion.
- Envelope alignment. (When the speech waveform is reconstructed it is necessary to ensure that relative timing information between the two channels is preserved.)

By considering the simple example of a modulated sine wave which is received at two receivers but delayed at one receiver by time γ, it is possible to demonstrate that binaural localisation cues are preserved.

After processing by the envelope expansion scheme, the modulated signals may be altered such that,

$$y_R(t) = [1 + \alpha \sin(\omega_e t)] \sin(\omega_s t) \tag{6}$$

$$y_L(t) = [1 + \alpha \sin(\omega_e (t + \gamma + \beta))] \sin(\omega_s (t + \gamma)) \tag{7}$$

where α = modulation depth (ranges from $0 - 1$)
γ = receiver time difference
β = envelope timing error due to processing
ω_e = envelope frequency
ω_s = carrier frequency.

If there are no timing errors introduced by the processing then $\beta = 0$ and $y_R(t) = y_L(t + \gamma)$; the sidebands generated by the envelope expansion have the relative interaural time delays used for sound localisation.

If the processing introduces a timing error in the envelope, then with further analysis it can be shown that the phase error introduced into the side bands is equivalent to a time difference of $\omega_e(\beta - \gamma)/(\omega_e \pm \omega_s)$. It therefore becomes important to keep both ω_e and $(\beta - \gamma)$ small in order for the sidebands to maintain the correct binaural relationship.

ω_e, the envelope bandwidth, is low-pass filtered in the envelope expansion process, by a filter having a cut-off frequency of 20Hz. This ensures that the envelope bandwidth is much lower than the center frequency of any of the band-pass filters in the filter bank.

The above argument has been extended to more complex signals, and informal listening tests have shown that localisation is possible as long as the expansion rate does not become too high. At higher expansion rates distortion effects on the speech become more dominant.

2.2 Algorithm implementation and speech intelligibility testing

The binaural expansion scheme has been coded in 'C' on a 33MHz 486 PC. Sentence material was recorded in a reverberant room with an average reverberation time of 1.8s. The reverberated data was transferred to the PC via the CARDD audio processing package and waveform analysis conducted using the public domain Speech Filing System [12].

Informal listening tests conducted on the binaural envelope expansion scheme indicate that even at expansion rates exceeding $v = 1.8$, sound localisation is preserved, although a disturbing degree of distortion is introduced. Consequently, the choice of expansion rate must be a compromise between the objective of increasing the modulation depth of the reverberant signal and restricting distortion effects.

An example output from the system is shown in Figure 2. The top two signals are the original reverberant speech waveforms for the left and right channels. The lower two waveforms show the output from the binaural envelope expansion scheme which used an expansion rate of $v = 2$, and ten 1.5kHz bands. The lower traces clearly show how the reverberant tails, present in the original signals, have been suppressed.

3. DISCUSSION

In this implementation of the envelope expansion scheme the applied expansion rate, v was strictly dependent upon the amplitude of the signal in each frequency band. This may have the effect of suppressing signals in the higher frequency bands, since the average speech spectrum of natural speech rolls off in a manner which is shown in Figure 3. With further work it may be possible to develop a variable expansion rate scheme based on the average speech spectrum.

4. CONCLUSION

This paper has discussed a binaural envelope expansion scheme which may be used to reduce the effects of reverberation whilst preserving cues necessary for sound localisation. Informal listening tests conducted on the algorithm indicate that the perceived level of reverberation is reduced and sound localisation cues are preserved. At the present time a formal speech intelligibility test is being conducted which aims to determine whether the algorithm improves speech intelligibility.

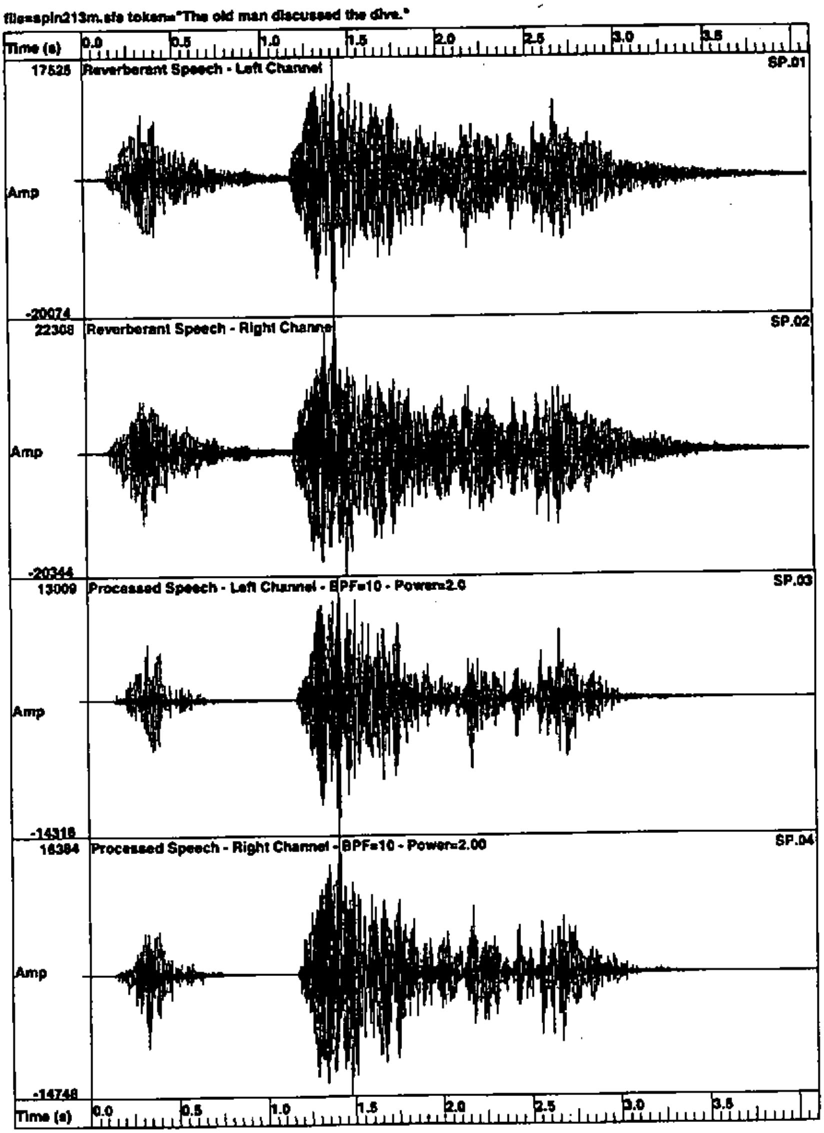

Figure 2: Example output of the Binaural Envelope Expansion Scheme

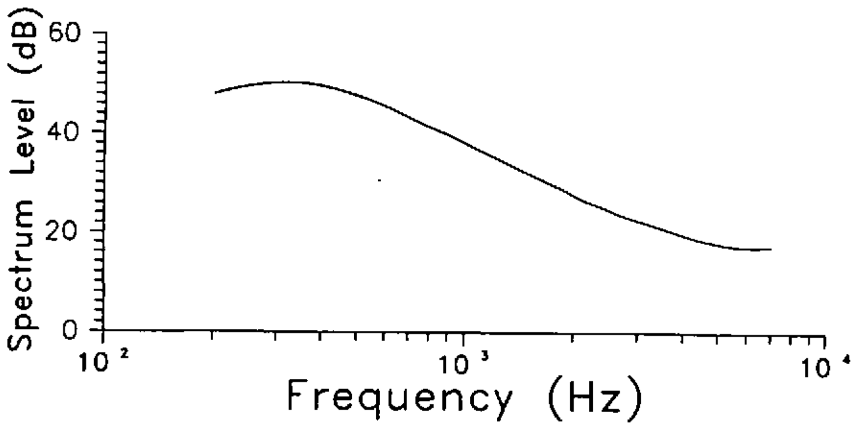

Figure 3: Average speech spectrum of male speech (from Kryter [13])

5. ACKNOWLEDGEMENTS

The authors wish to acknowledge the support of this work by the Science and Engineering Research Council and the IBM(UK) Scientific Centre.

6. REFERENCES

[1] J.P.Moncur, D.Dirks, *Binaural and Monaural Speech Intelligibility in Reverberation*, J.Speech and Hearing Research,10,186-195,1967.

[2] A.K.Nabelek, J.M.Pickett, *Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation, and hearing aids.*, JASA,56(2),628-639,1974.

[3] S.A.Gelfand, S.Silman, *Effects of small room reverberation upon the recognition of some consonant features*, JASA,66(1),22-29,1979.

[4] J.P.A.Lochner, J.F.Burger, *The intelligibility of speech under reverberant conditions*, Acustica,11,195-200,1961.

[5] J.B.Allen, D.A.Berkley, J.Blauert, *Multimicrophone signal processing technique to remove room reverberation from speech signals*, JASA,62(4),912-915,1977.

[6] K.D.Farnsworth, P.A.Nelson, S.J.Elliot, *Equalisation of room acoustic response over spatially distributed regions*, Proc. IOA, 7(3), 245-252, 1985.

[7] J.P.A.Lochner, J.F.Burger, *The subjective masking of short time delayed echoes by their primary sounds and their contribution to the intelligibility of speech*, Acustica, 8(1), 1-10, 1958.

[8] A.K.Nabelek, D.Mason, *Effects of noise and reverberation on binaural and monaural word identification by subjects with various audiograms*, J.Speech and Hearing Research, 24,375-383, 1981.

[9] A.K.Nabelek, J.M.Pickett, *Monaural and Binaural Speech Perceptions through hearing aids under noise and reverberation with normal and hearing-impaired listeners*, J.Speech and Hearing Research,17,724-739,1974.

[10] T.Houtgast, H.J.M.Steeneken, *The modulation transfer function in room acoustics as a predictor of speech intelligibility*, JASA, 28, 66-73, 1973.

[11] J.Mourjopoulos, J.K.Hammond, *Modelling and Enhancement of Reverberant Speech Using an Envelope Convolution Method.*, Proc. IEEE ICASSP, 1144-1147, 1983.

[12] M.D.Edgington, C.M.Barnes, P.D.Stringer, D.M.Howard, *The speech filing system: A tool for cooperative speech research.*, Proc. I.O.A, 14, 1992, (in press).

[13] K.D.Kryter, *Methods for the calculation and use of the articulation index*, JASA, 34(11),1962.

[14] H.Kurtovic, *The influence of reflected sound upon speech intelligibility*, Acustica, 33, 32-39, 1975.