IMPROVING SUBJECTIVE QUALITY IN WAVEFORM CODERS BY THE USE OF POSTFILTERING

A. Perkis, ELAB
B. Ribbum, ELAB
T. Ramstad, Dept. of Elec. Eng. & Comp. Science

The Norwegian Institute of Technology
Trondheim, Norway

ABSTRACT - Adaptive postfiltering is shown to significantly enhance the perceived speech quality of medium bit rate coders. The postfilter, utilizing auditory masking properties, provides an adaptive shaping of the noise and signal spectra, thus reducing the perceived quantization noise level at the cost of introducing some extra signal distortion. This paper will discuss the effectiveness of postfiltering in three distinctly different coding schemes. These are Regular Pulse Excited linear predictive coding (RPE) representing LPC based coding schemes, and Adaptive Sub Band Coding (SBC) and Adaptive Transform Coding (ATC) representing frequency domain coders.

1 INTRODUCTION

As digital communication has become an increasingly important field, there is a great need for high quality speech coders at medium bit rates. For 16 kbit/s and lower, linear predictive coding (LPC) based speech coders and frequency domain coders are both possible candidates. However, they all tend to have a perceivable level of quantization noise. The concept of postfiltering (Jayant, 1981) has been shown to improve the subjective quality of some coding schemes, viz. ADPCM and Code Excited Linear Predictive coding (CELP). The postfilter, utilizing auditory masking properties, provides an adaptive shaping, of the noise and signal spectra (Atal & Schroeder, 1979, Paliwal, 1987). One might thus reduce the subjectively perceived quantization level while introducing time-varying signal distortion.

This paper will demonstrate the effectiveness of postfiltering in three different coding schemes at bitrates from 13-15 kbit/s. These are Regular Pulse Excited linear predictive coding (RPE), Adaptive Sub Band Coding (SBC), and Adaptive Transform Coding (ATC). The synthesized speech data from these coding schemes all have a perceivable noise level. Improvement of the subjective quality by postfiltering will be demonstrated by using both long term and short term spectral information to shape the noise spectrum. The improvement is in general obtained with no increase in bitrate, as long as the postfilter parameters in the SBC and ATC coders are derived from the reconstructed speech (i.e. the LPC parameters are estimated from the noisy data).

In designing the postfilter it is important to obtain a balance between quantization noise suppression and signal distortion. Different simulation results obtained by varying two parameters α and ε, and thereby obtaining different degrees of postfiltering are discussed. Due to the signal distortion introduced the results must be evaluated subjectively. The postfiltered utterances are ranked using a paired comparison listening test (Sreenivas, 1987).

Chapter 2 gives the theoretical background of postfiltering, while the actual coding schemes are briefly described in chapter 3. At last subjective evaluations of postfiltering are discussed.

2 THE CONCEPT OF POSTFILTERING

Noise shaping is a commonly used technique in ADPCM coders where the noise spectrum is shaped according to the signal spectrum to increase the subjective quality based on auditory masking. The signal spectrum is unchanged in this technique. Postfiltering will modify the noise and signal spectra similarly and thus introduce some signal distortion. Both methods will result in a poorer signal to noise ratio.

Adaptive postfiltering was first introduced by Jayant in (Jayant, 1981) where the structure depicted in Figure 1 was proposed.

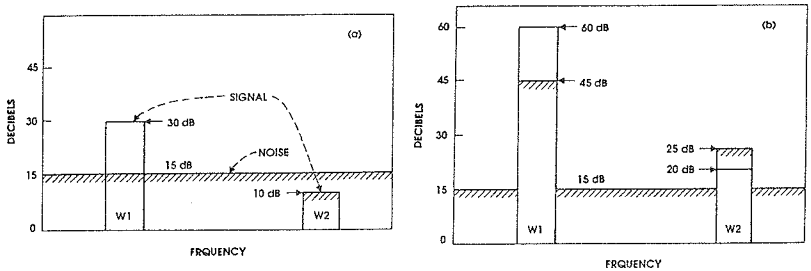


Figure 1. An idealized explanation of the effects of postfiltering, assuming a signal with two narrowband components and a white noise spectrum. (a) Signal and noise spectra at the input to the postfilter, showing signal-to-noise ratios of 15 dB and -5 dB in signal frequency bands W_1 and W_2. (b) Spectra of postfiltered signal and postfiltered noise, assuming a postfilter transfer function identical to the signal spectrum in (a). Regions W_1 and W_2 continue to have local signal-to-noise ratios of 15 dB and -5 dB as in (a), but the signals are now 45 dB and 10 dB above the out-of-band noise level. In (a) the corresponding numbers are only 15 dB and -5 dB. The overall effect is a reduction of perceived noise, but the price paid is a change in the relative strengths of the signal components in W_1 and W_2.

The most general postfiltering scheme is based on both long term and short term spectral information of the synthesized speech by using the general structure in Figure 2. The long term correlation as represented by the pitch parameters give the fine spectral information, while the short term predictor represented by the LPC coefficients give the global spectral shape (Kroon & Atal, 1987).

$$\hat{s}(n) \xrightarrow{G} \boxed{\frac{1}{P'(z)}} \longrightarrow \boxed{\frac{1}{A'(z)}} \longrightarrow \hat{s}_p(n)$$

Figure 2. General postfilter structure.

If the optimal long term predictor is expressed as

$$P(z) = 1 - \beta z^{-M} \qquad (1)$$

and the short term predictor as

$$A(z) = 1 - \sum_{i=1}^{n} q_i z^{-i} \qquad (2)$$

then the corresponding postfilter is composed of

$$\frac{1}{P'(z)} = \frac{1}{P(\varepsilon^{1/M} z)} \quad \text{and} \quad \frac{1}{A'(z)} = \frac{1}{A(z/\alpha)} \qquad (3)$$

Where $0 \leq \varepsilon < 1$ and $0 \leq \alpha < 1$.

These parameters give the flexibility of varying the impulse response of the postfilter between that of an allpass filter ($\alpha = \varepsilon = 0.0$ i.e. no post-filtering), and making the filter equal to LPC filter ($\alpha = \varepsilon = 1.0$). Figure 3 shows the impulse response of the postfilter for various values of α.

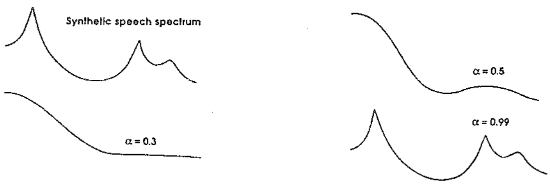


Figure 3. Frequency response of the short-term postfilter for varying values of α.

When designing the postfilter it is important to obtain a balance between noise suppression and signal distortion. Thus by varying the two parameters, α and ε, the degree of noise shaping and signal distortion is varied. Due to the lack of any good objective criterion for subjective quality, the optimum value of α and ε must be determined by subjective assessment.

Eq. 1 is based on the fine spectral structure, i.e. pitch information. Given a suitable value of ε, the postfiltering will "attenuate the valleys" in the comb filter at the cost of increasing the bandwidth by a factor ε. Eq. 2 is based on the spectral envelope, i.e. LPC parameters, and will thereby affect the formant structure. Values of the parameter α in the range $0<\alpha<1$ amplify the magnitude of the resonance and increase the bandwidth at the resonance frequencies. Figure 4 gives a theoretical example of postfiltering.

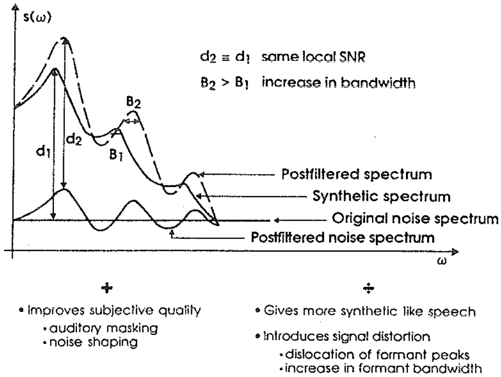


Figure 4. The postfilter, utilizing auditory masking properties, provides an adaptive shaping of the noise and signal spectra.

The postfiltering in general introduce an amplification of the signal which can be controlled by the factor G given by:

$$G = \frac{\sigma_{\hat{S}(n)}}{\sigma_{\hat{S}_p(n)}} \tag{4}$$

3 THE CODING SCHEMES

The sub band coder utilizes a 16 channel filterbank constructed as a tree structure of quadrature mirror filters (QMF). The coding gain is obtained by adaptively allocating bits to the quantizers in the different bands. The coder is basically as described in (Ramstad, 1982) with the exception that the spectral information is represented by vector-quantization (Perkis, 1987). The bitrate is 15.0 kbit/s.

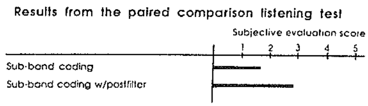
The adaptive transform coder is based on (Krishnan & Paliwal, 1986) and the simulation is described in (Ribbum, 1988). The block size is 128 samples and the number of sub-blocks is kept at 16 giving a total bitrate of 15.0 kbit/s.

The regular pulse excited (RPE) linear predictive coder is the recommended standard for the GSM mobile telephone system. The coder is described in (P. Vary et al., 1987). The bitrate is 13.0 kbit/s.

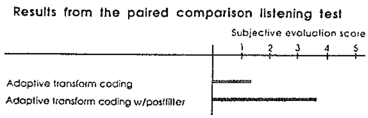
4 PERFORMANCE RESULTS

Subjective evaluations by means of randomized paired comparison tests have served two purposes. Firstly to find the optimum degree of postfiltering i.e. the best subjective values of α and ε, and secondly to get an indication if postfiltering is preferred or not.

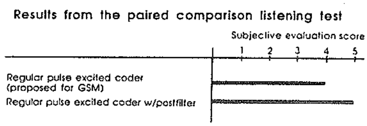
The speech material used is at sampled 8 kHz and IRS filtered. For the ATC coder a female Norwegian sentence is used while the two others use an English conversation between a man and a woman. The sentences are of equal length. Results of the tests are shown in Figure 5. The listening tests are carried out for the three coding schemes separately, and give no information on their relative preference.



a) Sub-band coding.



b) Adaptive transform coding.



c) Regular pulse exited linear predictive coder

Figure 5. Performance results from the listenings tests.

NOTE: The subjective evaluation scores only give a relative measure of preference between the noisy speech with and without postfiltering, and not between the coding schemes.

For all three coding schemes it is clear that postfiltering does indeed improve the subjective quality, and a reasonable choice of α and ε would be $\alpha = \varepsilon = 0.3$ as reported for the Code Excited Linear Predictive coder (Kroon & Atal, 1979)

It seems that the "bird song" like background noise present in the ATC coder can be successfully masked by using a stronger degree of postfiltering ($\alpha = \varepsilon = 0.5$). The effect of a stronger signal distortion, generally giving the postfilter a stronger low pass characteristic and thereby "muffling" the speech, seems to be subjectively more tolerable for this background noise.

To prevent an increase in bit-rate, the LPC parameters in the waveform coders must be found based on the synthetic speech data, i.e. estimated from a noisy signal. By instead estimating the parameters from the original speech data, thus adding the LPC-parameters as extra side-information, a small increase in signal to noise ratio is obtained, and subjectively gives a slightly more "crisp" sounding speech.

5 CONCLUSION

The effect of postfiltering has been investigated for three different coding schemes; Adaptive Transform Coding (ATC), Adaptive Sub-Band Coding (SBC) and Regular Pulse Excited linear predictive coder (RPE), two waveform coders and an LPC based coder, respectively.

Postfiltering will, in all the three coding schemes represent a significant improvement in subjective quality by choosing appropriate values of the parameters α and ε. To obtain an optimum balance between noise suppression and signal distortion, α and ε must be found by subjective evaluations.

By using a paired comparison listening test, $\alpha = \varepsilon = 0.3$ were found to be a subjectively good combination. The test also showed that the "bird song" like background noise present in the ATC coder can be successfully masked by using a stronger degree of postfiltering ($\alpha = \varepsilon = 0.5$).

Acknowledgements: This work was supported by the multiclient project in speech processing at ELAB, Norwegian Institute of technology, Trondheim, Norway.

REFERENCES

Atal, B.S. and Schroeder, M.(1979) *Predictive coding of speech signals and subjective error criteria,* IEEE, ASSP, pp. 247-254.

Jayant, Ramamoorthy (1981) *Adaptive Post-Filtering of ADPCM speech,* BSTJ, vol. 60, no. 5.

Krishnan, S. and Paliwal, K.K. (1986) *An optimization study of Adaptive transform coder,* Indian J. Technol., Vol. 24, pp. 777-782.

Kroon, P., Atal, B.S.(1987) *Quantization procedures for the excitation in CELP coders,* IEEE Conf. proc. ICASSP'87.

Paliwal, K.K. (1987) *Stochastic, multipulse and self excited linear predictive coders for low bit-rate coding of speech,* ELAB report STF44 F87108.

Perkis, A. (1987) *Available simulation modules for sub-band coding of speech,* ELAB report STF44 F87118.

Ramstad, T. (1982) *Sub band coder with a simple adaptive bit-allocation algorithm a possible candidate for Digital Mobile Telephony* IEEE ICASSP, Paris, May 1982.

Ribbum, B. (1988) *ATC coder simulation program at ELABs VAX 11/750, ELAB* project memo AN88038.

Sreenivas, T.V. (1987) *Modelling LPC-residue by components for good quality speech coding,* ELAB report STF44 F87147.

Vary, P. et al. *The Candidate for the GSM Radio Communication system",* International Conference on Digital Land Mobile Radio Communication 30 June-3 July, 1987, Venice, Italy.