INVESTIGATING THE USE OF ASYMMETRIC WINDOWS IN CELP VOCODERS

Dinei A. F. Florêncio

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, GA 30332

ABSTRACT

The behavior of several windows in relation to time delay and spectral estimation is analyzed in its application to CELP coders, with emphasis on time delay considerations. From the data, the importance of Asymmetric Windows in such context becomes evident. In order to get objective conclusions, we have used a perceptually weighted residual energy as the comparison criteria. It is shown that a reduction in delay of up to 8 ms can be obtained with no penalty in quality, simply with the correct choice of the window. Application to the military CELP standard, and considerations about the choice of asymmetric windows for other CELP implementations are also included.

1. INTRODUCTION

Due to the time-varying characteristics of speech, many speech coding algorithms process the data in a block fashion, usually using some kind of smoothing window. This is specially true for bit rates below 16 Kbps. In these coders, a windowing process is always present, be it in the direct form, be it in the form of a transform or even in the form of an IIR filter (i.e., an infinite length window). In most cases, the introduction of this window means averaging the parameters over a certain interval, producing an extra delay in the coder. Traditional choices for the window in speech coders include Hamming, Hann and Rectangular Windows. The use of windows with smaller sidelobes does not usually improve the performance of the coder.

With CCITT standards adopted for speech coding at bit rates above 16 Kbps, the research is now concentrated on bit rates between 2.4 and 12 Kbps. CELP coders have been considered the main option for speech coding around this rate, and have been object of intense research in recent years. In most applications (general telephony, mobile communications, etc) one of the important aspects of the performance of these coders is time delay, and every effort has to be done to keep the coding delay as low as possible. The 10 ms delay CCITT requirement for the 8 kbps codec [1] can be viewed as a goal, while most algorithms have total delays around 40ms.

In [2] Florêncio introduced the idea of using (Finite) Asymmetric Windows as a way of reducing the delay added to the spectral estimation by the windowing process. This can be explored in LP analysis in CELP coders.

The choice of (symmetric) windows for CELP coders has been analyzed by Bastiaan Kleijn, and the result incorporated in the 4800 kbps federal standard [3]. As one example, the result presented in this paper is able to reduce by 5-8 ms the delay of that algorithm.

In this paper we evaluate the potential of using Asymmetric Windows in CELP-based coders. Section 2 describes the objective comparison criteria we have used. Section 3 presents most of the results, while section 4 presents the results of implementing asymmetric windowing on the military standard CELP coder. Section 5 makes some considerations about the choice of asymmetric windows for real CELP implementations, and Section 6 presents the main conclusions of the work.

2. METHODS

CELP coders are usually evaluated using subjective criteria (MOS or alike). Nevertheless, this would not be appropriate for comparing windows, as we cannot expect dramatic differences among their performance (This is easily observed by informal listening tests). This way, we need an objective comparison criterion. To obtain this objective criterion, we looked at CELP coding as a simple (spectral weighted) vector quantization of the LP residual. At this point of view, it is clear

0-7803-0946-4/93 $3.00 © 1993 IEEE

that reducing the residual energy would imply a direct improvement of the coder performance. Therefore, it is reasonable consider the residual energy as a measure of the window performance (Fig. 1). The perceptual weighting filter includes also a low pass filter with cut-off frequency of 3.6 KHz, as it would be performed by the D/A converter. This criterion allow for objective comparison between the windows, while keeping a close relation to the overall (subjective) coder performance.

When dealing with a specific CELP implementation, the details of that implementation can be incorporated in this framework. That is, the inverse filter should use the same interpolation method, the weighting filter should use the same parameters, the frame (and sub-frame) rate should be the same, etc. Nevertheless, in general a sub-optimal choice is acceptable. The considerations of item 5 are rather general, and the conclusions about window choice made in that item will be appropriate for most CELP coders.

We have chosen the military standard CELP as an example implementation. The results obtained by introducing an asymmetric window in that vocoder are commented in section 4. The processed sentences do not include the sentences used in optimizing the window.

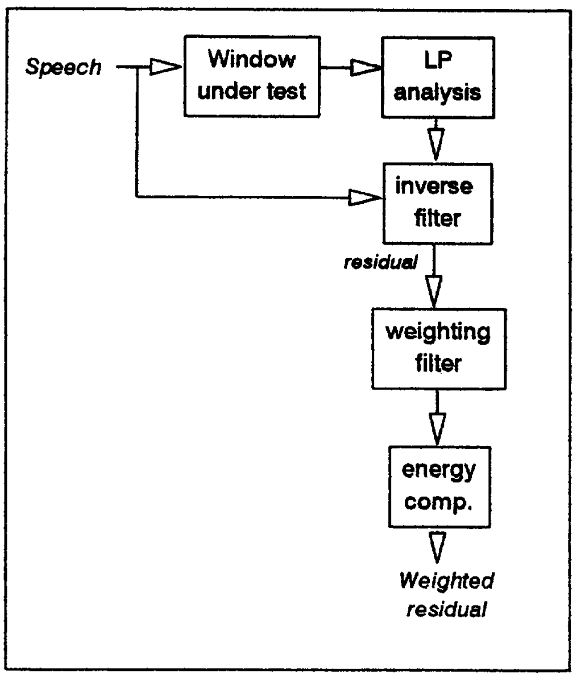

Figure 1 - Block diagram of the evaluation Method.

3. MAIN EXPERIMENTAL RESULTS

Using the above procedure (Fig. 1), we have collected data for several windows, with different settings for frame rate, allowed delay, window length, weighting factor (γ), etc. Table I show some of the most significant results (16 ms frame, $\gamma = 0.8$). In that table, '=' means that the window did not profit from the greater allowed delay. The values in the table represent the (weighted) residual energy, normalized in relation to the rectangular window.

The first point to note in the table is that the performance of the traditional windows significantly decrease when we impose a limit to the delay (Note that the 'ideal' delay for a symmetric window would be 50% of the difference between the window and frame lengths). This can be partially compensated reducing the size of the window. Its also clear that the Hann and Hamming windows have performed better than the rectangular window for any delay. But, the important point to note is that the asymmetric window has shown an even better performance. In fact, the asymmetric window with $\gamma = 3$ and *delay* $= 0$ has been out-performed only by symmetric windows with delays of 64 samples (8ms). And, even with that delay, an asymmetric window with $\gamma = 3$ and delay of 16 samples (2 ms) performed better. Although details of each individual algorithm would have to be considered, this show that changing a Hamming or Hann window for an Asymmetric Window can reduce the delay of CELP coders up to 64 samples, with no penalty in quality. It should be pointed out that this improvement is obtained with no cost, simply by changing the window coefficients in the coding algorithm. Also important, in many cases it can be implemented without affecting compatibility, since no modification is usually needed in the synthesizer. (see item 4 for a specific example).

In some low-delay coders, the delay has already been constrained, by imposing an limit (often zero) on the number of samples the LP analyzer can use from the next frame. In this case, the use of an asymmetric window cannot reduce it further. Nevertheless the introduction of an asymmetric window in these coders (provided one is not used yet) can improve the quality of the coded speech. In these coders backward LP analysis is most often used, and therefore modifications would have to be implemented in both coder and decoder. We should point out that many LD-CELP algorithms already use recursive windowing [5][7], and would not profit very much from the use of (finite) asymmetric windows.

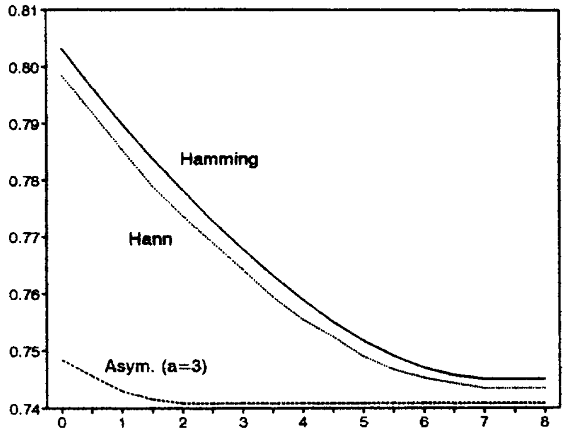

Fig. 2 - Residual Energy as a function of allowed delay for different windows

4. A SPECIFIC EXAMPLE

As an example of a real implementation, we have experimented with the introduction of an asymmetric window in the analysis part of an implementation of the Federal Standard 1016 (4800 bps CELP coder). We used an asymmetric window (240 points parabolic-exponential with $\alpha = 2.8$) and set the analysis window to use 45 futures samples less than the original Hamming window. This reduces the delay in the algorithm by 5.6 ms. For the sentences set processed, the segmented SNR has even increased 0.05 dB (from 7.55 dB to 7.60dB). Note that the only modifications that need to be implemented are in the buffer used for the LP analysis (to select the appropriate samples) and in the window coefficients. In particular, no modification is needed in the synthesis part of the vocoder.

For a delay reduction of 7.5 ms (60 samples), the segmented SNR decreased by 0.11 dB. As a comparison, using Hamming window with the same reduction in delay, decreased the SNR by 0.57 dB. Furthermore, we believe that even this small difference (0.11 dB) could be offset by reoptimizing the CELP codebook for the new window (but note that this would require making the same modification in the synthesizer). Informal listening tests did not show any difference in quality between speech processed by coders using the two window (Hamming and Asymmetric).

5. CHOOSING AN ASYMMETRIC WINDOW

The usual criteria for window selection includes parameters such as bandwidth and sidelobe level. This has been discussed in [2] for asymmetric windows and [6] is an excellent reference for traditional (symmetric) windows. Although ideal when dealing with applications like sinusoid detection, this is somewhat embarrassing when applied to speech analysis. ***"What is an appropriate sidelobe level?"*** and ***"Why get an window with narrow bandwidth and apply bandwidth expansion?"*** are common questions we ask ourselves in such cases. This is even more relevant when referring to asymmetric windows, as little information is available about them in the literature.

In this paper in have analyzed a family of asymmetric windows (parabolic-exponential), which is parameterized by the exponential decay rate α, namely:

$$w(n) = K\frac{nN - n^2}{N^2} exp(-\alpha(N - n)/N)$$

for $0 \leq n \leq N$, and where K is an optional constant (function of α) gain, to bring the energy of the windows samples to unit level, if necessary.

To clarify some points about the choice of this parameter α, we have used the same framework described in 2, computing residuals for allowed delays of 0 and 16 samples, for a window size of 256. Figure 3 shows a plot of this data for the case $\alpha = 1$. Note on that figure that the choice of α is not a critical one. For example, for an allowed delay of 16 samples (2 ms), the optimum value for α would be 2.5, but any value between 2.0 and 3.0 would perform reasonably well. From simulations with different settings, we concluded that a rather general choice would be $\alpha = 2.8$ for *delay* = 0 and $\alpha = 3.0$ for *delay* = 16 (2 ms).

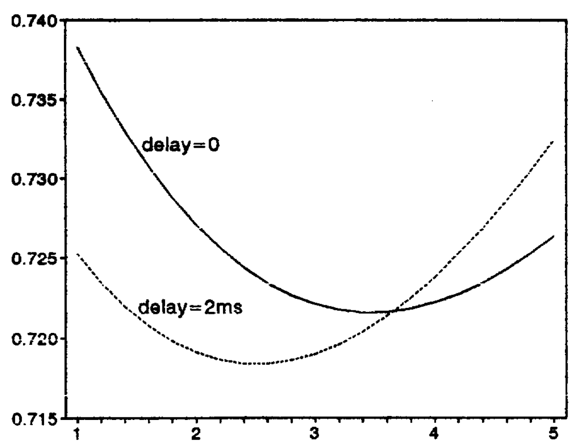

Fig. 3 - Residual Energy as a function of α.

Window	Lenght	Allowed Delay (samples)			
		0	16	32	64
Rectangular	128	1.0000	=	=	=
Hann	256	0.9363	0.9068	0.8849	0.8709
Hamming	256	0.9421	0.9123	0.8891	0.9734
"	192	0.9026	0.8852	0.8825	=
"	160	0.9002	0.8977	=	=
"	128	0.9336	=	=	=
par-exp a=1	256	0.8981	0.8791	0.8717	=
" a=2	256	0.8810	0.8694	=	=
" a=3	256	0.8716	0.8689	=	=
" a=4	256	0.8718	=	=	=
" a=5	256	0.8758	=	=	=
Barnwell		0.8745	=	=	=

Table 1 - Residual Energy for some Windows (gamma = 0.8).

6. CONCLUSIONS

In this paper we have analyzed the use of asymmetric windows in the LP analysis of CELP coders. The experimental data showed that the an appropriately chosen window can reduce the time delay by up to 50% of the analysis frame. Typical reductions in delay will be around 6-8 ms. We have observed that this delay reduction is obtained with no cost in complexity or speech quality, and that in many algorithms it can be implemented in the coder without affecting compatibility with previous versions of the same algorithm.

As one example of a specific implementation, an asymmetric window has been introduced in the Federal Standard 1016 CELP algorithm (4.8 Kbps CELP). This has shown to reduce the implicit delay of that algorithm by 5.6 ms with no loss in quality. A reduction by 7.5 can be obtained but may imply a slight loss in quality.

We have also analyzed the choice of the parameter in the parabolic-exponential window when being used in CELP coders. The analysis showed that appropriate values for are around =2.8 (for 2 ms allowed delay) and around =3.0 (for zero allowed delay).

We believe the results presented in this work are enough to discourage any application of traditional (symmetric) windowing for LP analysis in real time CELP coders.

REFERENCES

[1] CCITT Study Group XV, "Terms of Reference of the Ad Hoc group on 8Kb/s Speech Coding", **CCITT Study Group XV**, 1989.

[2] D. A. F. Florencio, "On the Use of Asymmetric Windows for Reducing the Time Delay in Real-Time Spectral Analysis", **in Proc. ICASSP'91**, pp 3261-3264, 1991.

[3] J.P. Campbell Jr. T.E. Tremain, V.C. Welch, "The Proposed Federal Standard 1016 4800 bps Voice Coder: CELP", **Speech Technology**, pp 58-64, Apr-May 1990.

[4] T.P. Barnwell, "Recursive Windowing for Generating Autocorrelation Coefficients for LPC Analysis", **IEEE Trans. on Acoustics, Speech and Signal Processing**, vol ASSP-29, pp 1062-1066, Oct 1981.

[5] V. Iyengar and P. Kabal, "A Low Delay 16 kb/s Speech Coder", **IEEE Trans. on Acoustics, Speech and Signal Processing**, vol 39, n. 5, pp 1049-1057. May 1991.

[6] F.J. Harris, "On The Use of Windows for Harmonic Analysis with the Discrete Fourier Transform", **Proceedings of the IEEE**, vol 66, pp 51-83, Jan 1978.

[7] AT&T, "Detailed Description of AT&T's LD-CELP algorithm", **CCITT study group XV/21**, contribution n. D.118. Nov 1989.