Proc. IEEE Workshop Speech Coding (Ste. Adèle, QC), pp. 39-40, Oct. 1993

Smooth Speech Reconstruction Using Prototype Waveform Interpolation[†]

M. Leong[1] and P. Kabal[1,2]

[1]Electrical Engineering, McGill University, Montreal, Quebec, H3A 2A7
[2]INRS–Télécommunications, Université du Québec Verdun, Quebec, H3E 1H6

Abstract

The Prototype Waveform Interpolation (PWI) speech coding technique aims to achieve natural sounding speech by producing smoothly evolving waveforms for voiced speech. The technique is to control the level of periodicity in reconstructing the voiced speech segments. It was found, however, that PWI does not always produce the smoothly evolving waveforms desired because the method does not truly control the level of periodicity. A revised PWI scheme is proposed which provides control over the level of periodicity and achieves smooth speech reconstruction.

1 Introduction

Current speech coders operating at very low rates produce highly intelligible speech but suffer from unnatural speech quality, especially for voiced segments of speech. A method for encoding voiced speech called prototype waveform interpolation (PWI) has recently been proposed by W.B. Kleijn to achieve high quality, natural sounding speech reconstruction at low rates (below 4.0 kb/s) [1, 2]. The method is integrated with code-excited linear prediction (CELP) [4] for encoding the unvoiced speech segments.

The fundamental idea behind PWI is to produce smoothly evolving, highly periodic waveforms for voiced speech. The method aims to control the level of periodicity in the reconstructed speech. Our investigation has revealed that PWI does not fully control the level of periodicity and consequently does not always produce smoothly evolving waveforms. Audible warble is occasionally present in the voiced speech segments. An improvement to the PWI scheme is suggested which provides control over the level periodicity and produces the smoothly evolving waveforms desired.

2 PWI Overview

The approach of PWI is to take better advantage of the highly periodic nature of voiced speech by encoding only a single pitch cycle prototype per update frame. Traditional coders perform waveform matching on an entire update frame and fail to produce high quality speech at very low rates. PWI reconstructs the speech by interpolating the prototypes. The interpolation procedure produces a smoothly evolving waveform and PWI furthermore provides control over the level of periodicity of the reconstructed signal by controlling the cross-correlation between quantized prototypes [1, 2].

The basic steps of PWI are summarized as follows: 1) filter the speech with a short-term linear predictive coding (LPC) filter to remove spectral redundancy [3] (the LPC coefficients must of course be encoded), 2) interpolate between samples to double the sampling resolution (from 8 kHz to 16 kHz), 3) determine the pitch period and extract a pitch cycle prototype, 4) encode the prototype by taking its discrete Fourier transform (DFT) and differentially encode the DFT using the quantized prototype of the previous frame [1, 2], 5) construct the excitation signal by interpolating between prototypes, and 6) reconstruct the speech by passing the excitation through the LPC synthesis filter. For step (5) the prototypes can be interpolated in the DFT domain using the inverse DFT. The DFT coefficients and the pitch periods are interpolated. The interpolation formula below is in the continuous-time Fourier series domain (refer to [5] for the discrete-time formulation):

$$
\begin{aligned}
&e(t) = \\
&\sum_{m=0}^{M} [(1-\beta(t))F_c(m)+\beta(t)G_c(m)] \cos\left(\phi_0+\int_0^t \frac{2\pi m d\tau}{(1-\beta(\tau))T_f+\beta(\tau)T_g}\right) \\
&+ [(1-\beta(t))F_s(m)+\beta(t)G_s(m)] \sin\left(\phi_0+\int_0^t \frac{2\pi m d\tau}{(1-\beta(\tau))T_f+\beta(\tau)T_g}\right)
\end{aligned}
$$

where $F_c(m)$, $F_s(m)$, $G_c(m)$, and $G_s(m)$ are the DFT coefficients of the previous and current prototypes, T_f and T_g are the periods, $\beta(t)$ is a function increasing from 0 to 1, and ϕ_0 is the ending phase from the previous frame.

3 Improving PWI

We have found in our simulations that the PWI encoding method proposed by Kleijn does not always achieve its objective of smooth speech reconstruction. The reconstructed speech can exhibit an inconsistent evolution of the wave-

[†] This work was supported by NSERC.

form amplitude, even with unquantized prototypes. The undesired envelope variations can result in audible warble.

Our initial attempt to improve PWI was to apply energy smoothing on the output speech waveform. This post-processing approach gives only small improvements and does not get to the source of the problem.

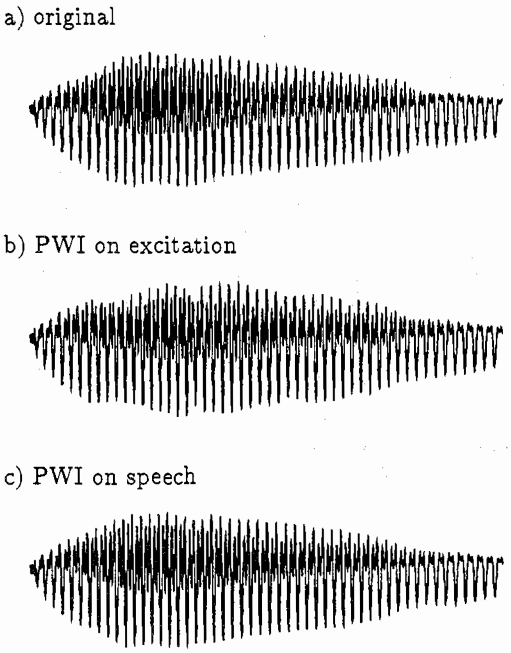
The problem was identified to caused by the time-varying nature of the short-term LPC filter. The filter coefficients are updated every frame and may actually change on subframe intervals if a scheme which interpolates the LPC coefficients is used. This filtering provides a lower energy signal on which prototype extraction and encoding is performed. PWI achieves smooth waveform reconstruction with control over the level of periodicity of the LPC excitation signal but this is not always achieved in the reconstructed speech. Adjacent prototypes are filtered with different LPC coefficients and segments of a single prototype may even be filtered with different coefficients. Linear interpolation, for example, in the excitation domain does not correspond to linear interpolation in the speech domain. Undesired variations in the envelope of the reconstructed speech are sometimes experienced and this can produce warble in the speech.

We suggest that in order to reliably achieve high quality speech reconstruction the PWI method should be applied directly on the unfiltered speech, thus avoiding the non-linear effects of the time-varying LPC filter. By definition of the PWI procedure, applying PWI in the speech domain using a linear interpolation function produces linearly interpolated speech frames. This approach therefore reconstructs the speech with a smooth envelope which does not produce warble. It provides the desired control over the periodicity of the reconstructed speech. The speech may have a smoother envelope than the original which is much preferred over having an erratic envelope. An example is shown in Fig. 1 in which the prototypes were unquantized. The original speech segment is shown in Fig. 1a. In Fig. 1b the reconstructed speech has an undesired envelope which results from applying PWI on the excitation. In Fig. 1c the speech has been reconstructed by applying PWI directly on the speech, producing a smoothly evolving waveform.

The quantization of the prototypes for the PWI approach proposed here is not less efficient compared to the approach of extracting the prototypes after formant filtering. Our PWI approach can still apply linear predictive filtering to predict a prototype from the previous one. For each prototype, only one set of filter coefficients is used to filter and resynthesize it so the unwanted time-varying effects are completely avoided.

4 Conclusion

In summary, this work proposes that the prototype waveform interpolation method should be applied in the unfiltered speech domain to avoid the non-linear effects of the time-varying LPC filter which can produce audible warble. Kleijn [1, 2] suggests that the level of periodicity in voiced speech should be preserved to achieve natural sounding speech but his scheme falls a little short of providing this control. Our modified PWI scheme provides true control over the periodicity and reliably produces the smoothly evolving speech waveforms desired.



a) original

b) PWI on excitation

c) PWI on speech

Fig. 1 PWI on speech vs. PWI on excitation. The speech was reconstructed with unquantized prototypes.

References

[1] W. B. Kleijn, "Continuous representations in linear predictive coding," *Proc. Int. Conf. Acoust. Speech and Sign. Process.*, pp. 201–204, Toronto, 1991.

[2] W. B. Kleijn, "Methods for waveform interpolation in speech coding," *Digital Signal Processing,* pp. 215–230, Sept. 1991.

[3] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *Proc. Int. Conf. Acoust. Speech and Sign. Process.,* pp. 661–664, Toronto, 1991.

[4] T. E. Tremain, J. P. Campbell, V. C. Welch, "Federal Standard (FED-STD) 1016: analog-to-digital conversion of voice by 4800 bits/sec code excited linear prediction (CELP) coding," U.S. Government, Dept. of Defense, U.S.A., Aug. 1989.

[5] M. Leong, "Representing voiced speech using prototype waveform interpolation for low-rate speech coding," M. Eng. thesis, McGill University, Montreal, 1992.