17-9-2000

Improved signal analysis and time-synchronous reconstruction in waveform interpolation coding

N. R. Chong-White
University of Wollongong, uow_chongwhiten@uow.edu.au

I. Burnett
University of Wollongong, ianb@uow.edu.au

Recommended Citation

Chong-White, N. R. and Burnett, I.: Improved signal analysis and time-synchronous reconstruction in waveform interpolation coding 2000.
https://ro.uow.edu.au/infopapers/208

Improved signal analysis and time-synchronous reconstruction in waveform interpolation coding

Abstract

This paper presents a waveform-matched waveform interpolation (WMWI) technique which enables improved speech analysis over existing WI coders. In WMWI, an accurate representation of speech evolution is produced by extracting critically-sampled pitch periods of a time-warped, constant pitch residual. The technique also offers waveform-matching capabilities by using an inverse warping process to near-perfectly reconstruct the residual. Here, a pitch track optimisation technique is described which ensures the speech residual can be effectively decomposed and quantised. Also, the pitch parameters required to efficiently quantise and recreate the pitch track, on a period-by-period basis, are identified. This allows time-synchrony between the original and decoded signals to be preserved.

Disciplines

Physical Sciences and Mathematics

Publication Details

This paper originally appeared as: Chong-White, NR and Burnett, IS, Improved signal analysis and time-synchronous reconstruction in waveform interpolation coding, Proceedings. IEEE Workshop on Speech Coding, 17-20 September 2000, 56-58. Copyright IEEE 2000.

IMPROVED SIGNAL ANALYSIS AND TIME-SYNCHRONOUS RECONSTRUCTION IN WAVEFORM INTERPOLATION CODING

N. R. Chong-White, I. S. Burnett

Whisper Laboratories,
University of Wollongong, Australia

ABSTRACT

This paper presents a Waveform-Matched Waveform Interpolation (WMWI) technique which enables improved speech analysis over existing WI coders [1]. In WMWI, an accurate representation of speech evolution is produced by extracting critically-sampled pitch periods of a time-warped, constant pitch residual. The technique also offers waveform-matching capabilities by using an inverse warping process to near-perfectly reconstruct the residual. Here, a pitch track optimisation technique is described which ensures the speech residual can be effectively decomposed and quantised. Also, the pitch parameters required to efficiently quantise and recreate the pitch track, on a period-by-period basis, are identified. This allows time-synchrony between the original and decoded signals to be preserved.

1. INTRODUCTION

The quality of waveform coders, such as Code-Excited Linear Predictive (CELP) coders, degrades rapidly at rates below 4kbps. Conversely, parametric coders, such as Waveform Interpolation (WI) coders are limited at higher rates by the speech production model. To achieve toll-quality speech at 4kbps, the favorable attributes of both these coders are combined - the waveform matching properties of CELP, and the effective decomposition and quantisation of WI. The WMWI technique extends and improves upon other approaches reported earlier [2][3].

The proposed Waveform-Matched WI (WMWI) coder has two main advantages over standard WI methods:

1. Improved analysis and decomposition of speech, and
2. Ability to achieve waveform coding.

In standard WI, pitch-length segments (characteristic waveforms (CWs)) of the residual are extracted at a constant rate, and aligned via a rotation process. However, the variable length of the extracted segments results in poor analysis during regions of rapid pitch variation, and the alignment process destroys relative phase information.

In WMWI, the residual is continuously time-warped to a constant pitch period, then pitch cycles are critically-sampled to form an evolving surface. Hence, an accurate description of the signal evolution is produced, without errors due to cyclic rotation[1] or the repetition or omission of segments due to selective extraction[4]. This allows improved signal analysis.

Good scalability is desirable to accommodate a variety of applications and is best achieved if the coder satisfies the waveform-matching property. In contrast to the standard WI reconstruction, where transmitted CWs are continuously interpolated without regard to the original positioning of these periods within the frame, the WMWI synthesis aims to preserve the time-locations of the pitch periods. This allows waveform matching, while requiring only a moderate increase in bit rate.

In this paper, we firstly discuss the mapping of a signal to the warped time-domain enabling effective decomposition of the pitch periods. Secondly, a method to efficiently quantise and reconstruct the pitch track of WMWI is described, allowing pitch periods of the unwarped residual to be time-synchronised with corresponding periods of the input residual.

2. ACCURATE SIGNAL ANALYSIS

1.1 Characteristic Waveform Alignment

The linear prediction residual is warped in the time-domain to remove its pitch variations and enforce a constant pitch period. The CW surface is then formed by critical-sampling of CWs. For efficient quantisation, the pitch track used for warping must be accurate, ensuring the extracted CWs are phase-aligned, and hence can be decomposed into a slowly evolving waveform (SEW) and a rapidly evolving waveform (REW) [1].

The effect of an incorrect and correct pitch track for a section of voiced speech residual is shown in Figure 1. The well-aligned periods of Figure 1(b) lead to most of the signal energy being separated into the SEW, as desired. However, poor alignment (Fig. 1a), causes pulses to be decomposed into the REW, making REW quantisation difficult. It should be noted that to utilise effective VQ techniques, pitch pulses following an unvoiced region must also be aligned with pulses preceding that section.

1.2 Optimising the Pitch Track

The pitch track is designed to align all pitch pulse peaks to a fixed position in each warped period. To minimise discontinuities at the period boundaries, this position is chosen to be the central sample of the pitch period. The pitch optimisation method described here significantly improves on our previous approach [3], in which calculation of the pitch track required a second (corrective) iteration.

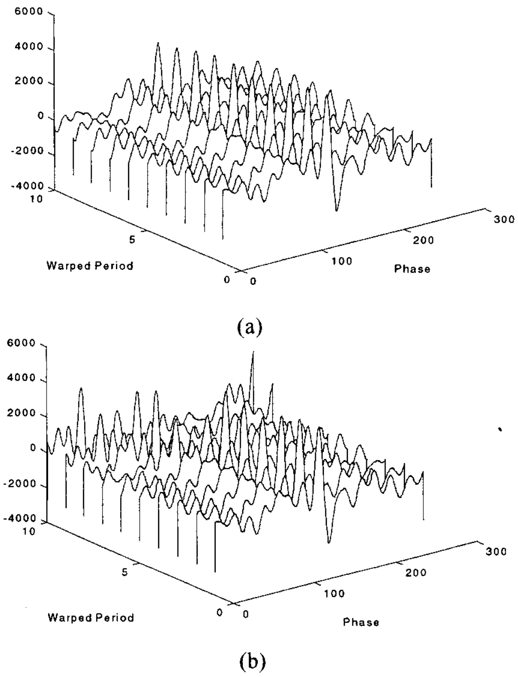
1.2.1 Definition of Terms

For the purpose of correctly warping to align pitch periods, the following terms are interpreted as follows:

a) Frames which contain sections of high periodicity and exhibit clear pulse peaks in the residual signal are labelled as *voiced*, otherwise they are *unvoiced.*

b) The *pitch period*, during voiced frames, is the distance between adjacent pulse peaks. During unvoiced frames, the pitch has no clear definition – it is simply assigned a value, to allow continuous time-warping.

0-7803-6416-3/00/$10.00©2000 IEEE



(a)

(b)

Fig. 1 CW surface in the case where the pitch track is (a) optimised, and (b) non-optimised (pulses not aligned).

1.2.2 Locating Pitch Pulses

To accurately determine the location of the pitch pulse peaks within the frame, the residual signal is lowpass filtered. A pulse detection algorithm, an extension of the technique described in [1], is then applied. Here, an initial pitch estimate for the frame, τ_{init}, is calculated from the autocorrelations of K segments, combined to form a composite function. For the case where K is odd, the composite autocorrelation function, R_C, for each candidate pitch value, d, can be expressed as:

$$R_C(d) = R_{\frac{K}{2}+\frac{1}{2}}(d) + \sum_{\substack{k=1 \\ k\neq\frac{K}{2}+\frac{1}{2}}}^{K} a_k \max\left[w(i)\cdot R_k(d-i)\right] \qquad -l(d)\le i\le l(d) \qquad (1)$$

where, for segment k, R_k is the autocorrelation function, a_k is a weighting factor determined by the voicing decision of the previous frame, $w(i)$ is a window function, and $l(d)$ is the window length.

The composite function is then recalculated (on an interpolated, filtered residual) for a small set of pitch period values surrounding the estimated pitch, τ_{init}, using segments of length equal to that value. If the refined R_C exceeds an adaptive threshold, it is proposed that the period contains a pulse, and the pulse peak location is determined at fractional sample resolution.

1.2.3 Pitch Track Calculation

Given the pitch pulse locations, the pitch track can then be formed. We define the pitch track for a set of four possible frame types: continuous voiced, continuous unvoiced, unvoiced-to-voiced, and voiced-to-unvoiced. It should be noted, that the true pitch contour, which reflects the nature in which the glottis opens and closes during speech production, may not be the optimum pitch track for good signal analysis and decomposition.

During a Continuous Voiced section, a simple, yet effective, technique is to simply allow the pitch to remain constant for the duration of the pitch period. During Continuous Unvoiced frames, the pitch takes on a nominal value.

For Unvoiced-to-Voiced frame transitions, the key requirement is to ensure that pitch cycles surrounding a variable duration unvoiced segment are aligned. Hence, the number of periods, n_u, and the pitch, τ_u, of the unvoiced section preceding the period with the first pulse must be chosen such that the first pulse peak is warped to the correct position. To minimise pitch variation, we solve

$$\underset{n_u}{\operatorname{argmin}} \left| (x_2 - x_1) - M\tau_u \right|, \qquad n_u = 1,2,3,... \qquad (2)$$

where,

$$M\tau_u = \frac{x_1 - \frac{x_2 - x_1}{2} - y}{n_u}, \qquad n_u = 1,2,3,.... \qquad (3)$$
$$\tau_{u\min} < \tau_u < \tau_{u\max}$$

where x_i is the position of the i^{th} pulse peak, y is the position of the end boundary of the last period of the previous (unvoiced) frame, and M is the interpolation constant. If x_1 is very close to the beginning of the frame, Equation 2 may be indeterminate due to the constraints on τ_u. In these cases, y is shifted back to the previous period boundary, and τ_u is recalculated.

3. TIME-SYNCHRONOUS RECONSTRUCTION

1.3 Unwarping

In WMWI, the residual is reconstructed by unwarping. Since no information is destroyed (or repeated) in the CW surface construction, near-perfect reconstruction can be achieved if the pitch track is accurately transmitted (note that the only source of error in unquantised WMWI is due to the filtering error of the warping/unwarping process). To obtain good speech quality, the analysis and synthesis pitch tracks do not need to be identical, however, if the two tracks differ significantly, distortions may result. If this occurs, time-synchrony may be lost, but the proposed pitch track reconstruction method will ensure it will be regained in the following frame.

1.4 Pitch Track Quantisation

If the analysis pitch track was to be perfectly recreated in the decoder, every pitch pulse position would need to be transmitted. Here, a method to construct a good representation of the analysis pitch track is described, which does not require large increases in the bit rate.

1.4.1 Pitch Parameters

In this approach, only the pitch of the final period of the frame is transmitted per frame. The pitch period value is quantised to half-sample resolution to reduce distortions resulting from the accumulation of rounding errors when integer pitch is used. However, to accurately recreate the pitch track, additional side information is required. The side information transmitted is:

a) Number of periods containing a pulse, including the periods overlapping each frame boundary, n_p

b) Number of periods containing no pulse, n_u

c) Pulsed/Unpulsed Classification

d) Period Boundary Information, x

Parameter d) is the most significant parameter which aids the waveform-matching objective. The number of warped samples at the end of the frame which do not make up a whole pitch period is transmitted to ensure input and output streams are synchronised at the beginning of every frame.

Bit allocations are shown in Table 1. A warped pitch period length, L, of 256 samples, a frame length, F, of 25ms and a sampling rate of 8kHz are used.

Parameter	**bits/frame**
Pitch Period Value	8
Period Boundary Information	8
No. of periods containing a pulse	4
No. of periods of noise	3
Pulsed/Unpulsed Classification	1
TOTAL BITS FOR PITCH	**24**

Table 1. Bit allocation required for pitch reconstruction

The side information specifies the configuration of periods within the frame, maximising the accuracy of the pitch track reconstruction. In the case of strongly voiced or unvoiced speech, not all parameters are significant. However, if the pitch has rapid variation, as transition regions often do, all five parameters are necessary to ensure the pitch track is accurately reconstructed.

1.5 Pitch Track Reconstruction

The pitch track reconstruction technique attempts to accurately reproduce the analysis pitch track on a period-by-period basis, rather than give a simplified interpolation or approximation, as in previous WI coders.

1.5.1 Continuous Voiced Frame

Fig. 2. shows a simple diagram of pulses in a continuous voiced frame. Given the pitch of the periods overlapping the frame boundaries, τ_{prev} and τ_{curr}, the pitch values of the intermediate periods are chosen to satisfy the constraints detailed in the side information. The pitch of the m^{th} intermediate period, τ_{p_m}, is calculated as follows:

$$\sum_{m=1}^{n_p-2} \tau_{p_m} = F - \left(\frac{\left(L - x_{prev,adj}\right)}{L} \tau_{prev} + \frac{x_{curr,adj}}{L} \tau_{curr} \right), \quad (4)$$

where
$$\tau_{p_m} = \frac{(n_p - 1 - m)\tau_{prev} + m\tau_{curr}}{n_p - 1}, \quad (5)$$
$$0 \le m \le n_p - 1$$

and
$$x_{adj} = \begin{cases} x + \frac{L}{2}, & x < \frac{L}{2} \\ x - \frac{L}{2}, & otherwise \end{cases} . \quad (6)$$

The adjusted value of x reflects the shift of the period boundaries to the pulse peak positions.

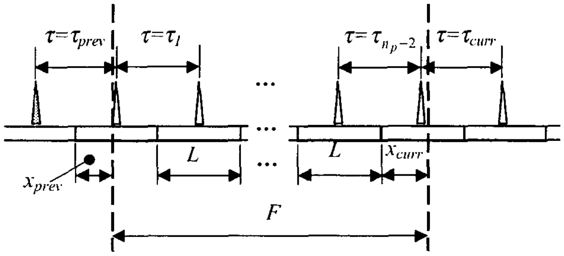


Fig. 2. Diagram of pitch periods within a "Continuous Voiced" frame in the warped domain.

1.5.2 Unvoiced-to-Voiced Transition

To improve accuracy and cater for pitch variations at the beginning of voiced regions better, an additional pitch, τ_{prev}, is transmitted during the previous unvoiced frame. The pitch of the n_u periods containing no pulse, τ_u, is calculated as follows:

$$n_u \tau_u = F - \left(\frac{\left(L - x_{prev}\right)}{L} \tau_{unv} + \frac{3}{2} \tau_{prev} + \sum_{m=1}^{n_p-3} \tau_{p_m} + \frac{x_{curr,adj}}{L} \tau_{curr} \right) \quad (7)$$

where τ_{p_m} is expressed in Equation 5, and τ_{unv} is the nominal unvoiced pitch period value.

4. CONCLUSION

WMWI provides improved analysis and decomposition of speech signals over standard WI, as well as satisfying the waveform coding objective. This technique, however, relies on accurate determination of the pitch track. We describe a method to reliably locate pulse peaks and construct an optimal pitch track to ensure the alignment of pitch pulses, even after unvoiced segments. We also define a set of transmission parameters that enables close reconstruction of the pitch track on a period-by-period basis, and hence maintains near time synchrony between the original and decoded speech signals

REFERENCES

[1] J.Haagen, W.B.Kleijn, "Waveform Interpolation" in "*Modern Methods of Speech Processing*" edited by R.Ramachandran and R.Mammone, Kluwer Academic Publishers, 1995.

[2] W.B.Kleijn, H.Yang, E. Deprettere "Waveform Interpolation Coding with Pitch-spaced Subbands", *Proc. 5th Int Conf. Spoken Language. Processing*, Dec 1998.

[3] N.R.Chong, I.S.Burnett, J.F.Chicharo "Adapting Waveform Interpolation (With Pitch Spaced Subbands) To Facilitate Vector Quantisation", *Proc. IEEE Workshop on Speech Coding*, Porvoo, Finland, pp96-98, June 1999.

[4] T.Eriksson, W.B.Kleijn "On Waveform-Interpolation Coding With Asymptotically Perfect Reconstruction", *Proc. IEEE Workshop on Speech Coding*, Porvoo, Finland, pp93-95, June 1999.