A NEW TIME-SCALE WARPING ALGORITHM FOR SINGLE DIMENSIONAL AND MULTIDIMENSIONAL SPEECH PARAMETER CONTOURS

A. Maheswaran (*), R. E. Bogner (**)

(*) Systems Engineering Division, Computer Sciences of Australia Pty. Ltd.
(**) Department of Electrical and Electronic Engineering, University of Adelaide.

ABSTRACT - In this paper a new sample association approach to be known as the Hilbert Warping (HW) algorithm is described. This algorithm is chosen from the observation that signals of similar form but with different time scales appear as similar trajectories when represented by suitable two dimensional plots in the X-Y plane, and overcomes difficulties such as identification of signal endpoints and assumptions about the smooth nature of warping that are permissible, associated with dynamic programming algorithms. The HW algorithm can be applied to both single dimensional and multi-dimensional signals as in dynamic programming algorithms.

INTRODUCTION

It is well known[1-3] that speaker verification experiments require warping of speech parameter contours for time registration purposes. The warping is expected to compensate for normal expected variations in speaking behaviour, which may displace corresponding events from repetition to repetition of a given utterance. In associating speech signal samples, dynamic programming approaches[1-5] have been found to be effective. These assign correspondences in such a way that the distance between two signals (vectors) is minimized subject to certain constraints on the permitted amount of warping or shift. The two signals, the 'test signal' and the 'reference signal', are both warped in a comparison. Although computational efficiency of the dynamic programming algorithms[1-5] may be improved, for example by the use of preconditioning through the identification of signal endpoints or fixed points, it is inherently computationally expensive. The dynamic programming algorithms implicitly use 'a priori' information about the smooth nature of warping that are permissible by imposing constraints on the associations that can be made; i.e., how abruptly the time-scale of one signal can be changed relative to the other signal.

In the HW algorithm the signals of similar form but with different time scales have coinciding trajectories in the X-Y plane. For example a sinewave $x(t) = A \cos \omega t$ is frequently represented by the X-projection of $c(t) = Ae^{j\omega t} = x(t) + jy(t)$ where j is $\sqrt{-1}$ and $y(t) = A \sin \omega t$. The trajectory of c(t) is the same circle for any value of ω. More complicated signals x(t) are similarly represented by their complex envelopes or analytic signals[6,7] in which $y(t) = \hat{x}(t)$, the Hilbert transform of x(t).

The performance of the new HW algorithm is found to be superior for warping speech parameter contours due to its computational efficiency and accurate representation of the warping neighbourhood, when compared with the conventional dynamic time warping (DTW) algorithm, where computational efficiency can only be improved by 'a priori' information on the warping neighbourhood, and by the use of preconditioning through the identification of signal endpoints or fixed points.

THE HILBERT WARPING (HW) PRINCIPLE

Curve theory principles[8] can be applied to derive the HW principle. We consider both linear and non-linear warping problems.

We start from the fact that the continuous analytic contour in the $s - \hat{s}$ plane can be represented by a vector function:

$$s + j\hat{s} = \underset{\sim}{s}(t) \quad \text{Interval:} \quad a \leq t \leq b \tag{1}$$

where the components of $s(t)$ are $s(t)$ and $\hat{s}(t)$. Then other functions are obtainable from eqn. (1) by imposing a transformation $t = h'(\tau)$ where τ is a warped version of t, which does not alter the shape of the analytic contour given in eqn. (1). A point set H on the analytic contour is defined to be 'Hilbert Warpable' if the point set represented by the new vector function $s(h'(\tau))$ remains the same under the transformation given by eqn. (1). Transformations that satisfy this specification exactly have been proven to be strictly linear for 'Hilbert Warpable' signals. However the transformation function can be chosen more arbitrarily if we make point to point correspondences between the chosen curve and the transformed curve.

THE HILBERT WARPING (HW) ALGORITHM

Sample association

In associating samples of s_T and s_R, two general approaches are possible. The first is to make direct association between samples of s_T and s_R by localized searches in the complex plane for nearest neighbours. This approach is basically sound, but it involves the processing of every reference signal with every test signal, and the distances associated with warped patterns will vary if the reference and test signal are interchanged such that the test signal is considered to be the reference signal.

An alternative efficient approach is to use a third set of phase values defined as the 'neutral phase sampler'. With this we pre-determine a fixed set of permissible phases:

$$\Phi_P(i) = i\Delta\Phi, \qquad i = 0, 1, \ldots N \tag{2}$$

with $\Delta\Phi$ set at some fixed increment. At this stage we just assume that the increment $\Delta\Phi$ is sufficient to provide permissible phase values in the range of reference and test phase values. Table 1 illustrates the phase warping approach.

We compare each value of $\Phi_P(i)$ with the phases $\Phi_R(k_R)$ of s_R (k_R) observed at samples k_R. $\Phi_R(k_R)$, $s_R(k_R)$ are respectively total phase and signal amplitude values of the reference signal. That value of k_R which gives the closest value of Φ_R is then associated with the phase index i, which becomes the sample index for the warped signal.

This allocation of phases to s_R is done once only for each reference, thus producing a set of 'neutral phase' references. To make the notation explicit, we denote the m^{th} reference by $s_{Rm}(k_R)$ in its original form, and by $s_{Rm}(i)$ in its warped form. All values of i will have some associated $s_{Rm}(i)$ for the uniform phase sequence $\Phi_P(i)$ to be maintained. Not all values of $s_{Rm}(k_R)$ appear in $s_{Rm}(i)$ because some are discarded in the

closest phase selection process. This phenomenon is illustrated by the non-appearance of values in $\Phi_R(i)$ and i columns of Table 1 alongside some values of k_R.

Each reference is processed only once, and the set

$$s'_{Rm}(i), \qquad m = 1, \ldots M \tag{3}$$

is stored, M being the total number of references. A similar process is used to associate values of test signal phase $\Phi_T(k_T)$ with $\Phi_P(i)$ and i. We obtain for each test signal $s_T(k_T)$ a warped set of data $s_T(i)$.

The comparison of test and reference signals is accomplished by comparing $s'_T(i)$ with $s'_{Rm}(i)$ for m = 1, ... M directly by correlation or, equivalently by computation of the Euclidean norm of their differences. The latter has the advantage of evaluating the comparison as the warping progresses.

AMBIGUITIES ASSOCIATED WITH HW

Figure 1(a) shows a signal $s_T(k_T)$ (test signal amplitude s at sample number k_T) in which the complex plane plot (plot of $s_T(k_T)$ vs. $\hat{s}_T(k_T)$) has a loop. It is not immediately apparent how the several values of $s_T(k_T)$ that occur near $\Phi = 50^o$ should be allocated to the warped signal $s'_T(i)$. The sign (polarity) of the signal phase difference ($\Delta\Phi(k_T) = \Phi(k_T) - \Phi(k_T - 1)$) changes from positive to negative and negative to positive on the loop. This disturbs the monotonicity of the signal phase curve, as shown in Figure 1(b), and leads to ambiguities in selection of $s_T(k_T)$ for the particular phases $\Phi_P(i)$ in the neighbourhood of 50^o.

The non-monotonic phase curves of the signals have two properties which should be understood so that they may be HW'd successfully by the HW algorithm. The information on the characteristic of the analytic contour (i.e., analytic contour with a loop) is shown by the peaks and valleys of the non-monotonic analytic phase curve (1) and (2) of Figure 1(c). The curves (1) and (2) are smoothed until monotonicity is achieved. These curves are illustrated by phase curves (3) and (4) of Figure 1(c), and illustrate the non-linear variation of phase between the signals. This non-linear variation is due to distortion of the time scale only, and the purpose of the Hilbert phase warping algorithm is to warp signals which undergo this non-linear variation and preserve the non-monotonic characteristics of each phase curve. Hence, we can choose either of the phase curve (3) or (4) to be the neutral phase sampler, and compare it with the other.

MULTIDIMENSIONAL HW (MHW)

In MHW we associate phase vectors, rather than phase points as in single dimensional warping. A phase vector at a particular time instant is defined by the phase value in each dimension at that time instant. The phase vectors are compared as for the single dimensional case in a monotonic fashion, and from the minimum distance phase vectors corresponding signal samples in each dimension they are associated.

PRACTICAL APPLICATION OF THE HW

The reflection coefficient time contours were obtained for the digitized speech data "the tighter the traveller wrapped" by sliding a 128 point Hamming window, and evaluating 14 order reflection coefficients (k_1, k_2,

... k_{14}) for each analysis segment obtained by a 32 point shift of the window in the direction of time. The reflection coefficient contours were chosen in favour of linear predictor coefficients because the stability condition on k_i, $|k_i| < 1$ is simple to preserve under quantization. Hence, smoothing of the k contours can be performed without affecting the stability of the all-pole filters; small changes in linear predictor coefficients can lead to instability of the filter.

As an example to illustrate the HW procedure, we consider the first reflection contour of repeated utterances by a speaker. Figure 2(a) illustrates the smoothed first reflection coefficient (k_1) time contour of the speech data "the tighter the traveller wrapped" spoken by the same speaker in two different sessions. The HW'd version of the signals is shown in Figure 2(b).

As an example to illustrate the superiority of the HW algorithm when compared to DTW algorithm, we consider the "k_1" contour of the speech data 'traveller' uttered twice by a speaker. Figure 3(a) illustrates the smoothed "k_1" contours. Figures 3(b), 3(c) and 3(d) illustrate respectively HW'd contours, DTW'd contours with no slope constraints and DTW'd contours with slope constraint. By comparing Figure 3(b) with Figures 3(c) and (d), the HW algorithm is shown to be superior to the DTW algorithm, due to its consideration of an accurate warping neighbourhood of the signal points without any 'a priori' information on the signals. The DTW algorithm, due to its arbitrary restriction of the warping neighbourhood, does not provide adequately warped signals.

CONCLUSION

Beginning with a simple illustration of the circular analytic contour for all sinewaves, we formulated the HW algorithm for classes of signals with non-linear monotonic phase curves and complex classes of signals with non-linear non-monotonic phase curves. It was observed that the latter class of signals consists of analytic contours with a loop. A novel, practical, HW algorithm, which consists of transformation of non-monotonic signal phase curve to monotonically smooth phase curve and choice of an appropriate common phase scale (neutral phase sampler), was developed.

The application of HW algorithm was demonstrated to be promising for time aligning speech parameter contours such as reflection coefficient contours, and observed to be comparable to the DTW algorithm. The disadvantages suffered by DTW algorithms, such as the requirement of 'a priori' information on warping neighbourhood and 'ad hoc' endpoint synchronization procedures, have been shown to be overcome by the HW algorithm. The HW algorithm is more efficient than the conventional DTW algorithm for warping one dimensional and multidimensional signals.

ACKNOWLEDGEMENT

The support of the Australian Radio Research Board is acknowledged with pleasure.

REFERENCES

1. ITAKURA, F., "Minimum prediction residual principle applied to speech recognition", IEEE Trans. vol. ASSP-23, pp. 67-72, Feb. 1975.

2. LUMMIS, R.C., "Speaker verification by computer using speech intensity for temporal registration", IEEE Trans. vol. AU-21, pp. 80-89, April

1973.

3. ROSENBERG, A.E., "Automatic speaker verification - review", Proc. IEEE vol. 64, pp. 475-486, April 1976.

4. SAKOE, H. and S. CHIBA, "Dynamic Programming algorithm optimization for spoken word recognition", IEEE Trans. vol. ASSP-26, pp. 43-49, Feb. 1978.

5. KUHN, H. and H. H. TOMASCHEWSKI, "Improvement in isolated word recognition", IEEE Trans. vol. ASSP-31, Feb. 1983.

6. STEPHEN, O.R., "Envelopes of narrow band signals", Proc. IEEE vol. 70, pp. 692-699, July 1982.

7. BOLTON, R.J., "Representation and pattern of Hilbert transformed electrocardiograms", Dept. of Electrical Engineering, University of Queensland, Brisbane, Australia, Report No. EE 83/9, July 1983.

8. KREYSZIG, E., "Differential Geometry", University of Toronto Press, 1959.

k_R	$\phi_R(k_R)$ [Deg]	$\phi_p(i)$ [Deg]	i
0	1	0	0
1	2		
2	6	5	1
3	7		
4	8		
5	8.5		
6	9	10	2
7	11.1		
8	11.5		
9	12.5	15	3
10	21	20	4
11	22		
12	22		
13	27	25	5
13	27	30	6
14	37	35	7
14	37	40	8
15	50	45	9
15	50	50	10
15	50	55	11
16	62	60	12

Table 1: Illustration of association of reference signal samples with neutral phase index i.

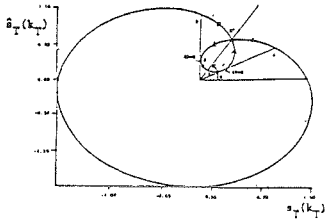

Fig.1(a): The complex plane plot of signal $s_T(k_T)$.

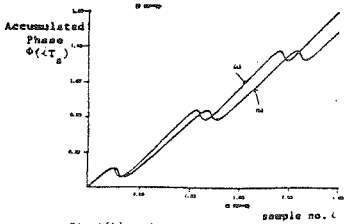

Fig.1(b): Non-monotonic phase curves of signals corresponding to the analytic contour in Fig.1(a).

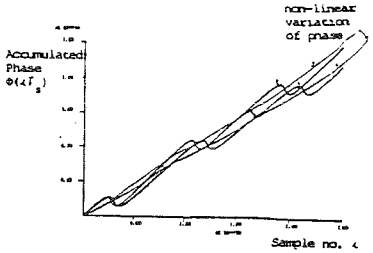

Fig.1(c): Non-monotonic phase curves (1,2) of the signals in Fig.1(a), and monotonically smoothed phase curves (3,4).

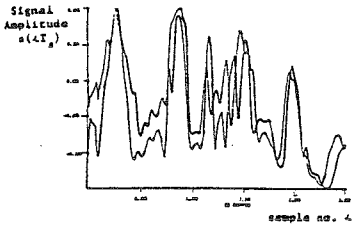

Fig.2(a): First reflection coefficient time contour: Speaker 1. Sessions 1 and 2.

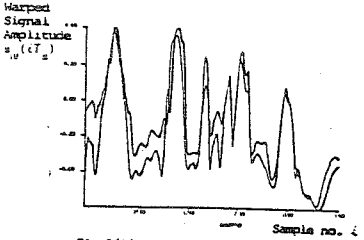

Fig.2(b): HW'd version of the signals in Fig.2(a).

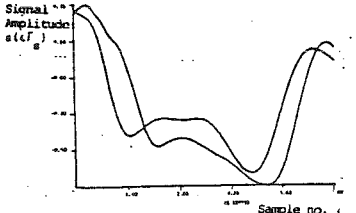

Fig.3(a): First reflection coefficient (k_1) time contour of the speech data "traveller" uttered by the same speaker in Sessions 1 and 2.

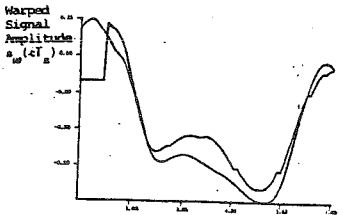

Fig.3(b): The HW'd version of the signals in Fig.3(a).

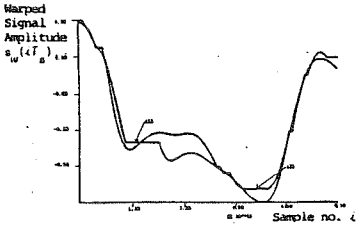

Fig.3(c): DTW'd signals of Fig.3(a) with no slope constraint on warping curve.

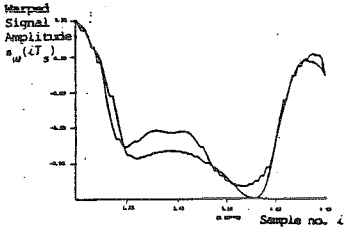

Fig.3(d): DTW'd signals of Fig.3(a) using the slope constrained warping curve.