

A Speed-Invariant Temporal Feature Detector

Paul E. Cisek

Department of Cognitive and Neural Systems

Boston University

111 Cummington Street, Boston, MA 02215

(617) 353-6426

Fax: (617) 353-7755

e-mail: pavel@cns.bu.edu

Michael A. Cohen

Center for Adaptive Systems

Boston University

111 Cummington Street, Boston, MA 02215

(617) 353-9484

Fax: (617) 353-7755

e-mail: mike@cns.bu.edu

ABSTRACT

Temporal pattern classification normally operates upon vectors whose components are time-delayed samples, be they samples of spectral, cepstral, or LPC coefficients over time. Such a spatial representation of temporal patterns is very sensitive to speed variations, and requires computationally expensive time-alignment techniques such as Dynamic Time Warping in order to compare inputs to exemplars. This paper proposes a *speed-invariant* representation of temporal patterns using Taylor series expansion. A vector composed of successive time-derivative samples of the temporal signal is unique to the shape of the function in a region around the sampling point. Simple manipulations on such a “Taylor vector” yield a speed-invariant form. The degree of speed invariance can be controlled parametrically while retaining sensitivity to the direction of presentation. Simulations demonstrate that learning and recall of temporal patterns coded using this representation is accurate and does not require Dynamic Time Warping.

1. Introduction

The classification of temporal patterns into meaningful categories is a fundamental problem encountered in sensory processing. It has been addressed with dynamic techniques such as Hidden Markov Models (Rabiner et al., 1989) and with pattern classification algorithms. These latter approaches include template-matching methods (Sakoe and Chiba, 1978, Rabiner et al., 1979, Sakoe, 1979) and neural network approaches (Anderson et al., 1988, Waibel et al., 1989, Sakoe et al., 1989). All the pattern classification algorithms operate on *spatial* patterns, and thus require the conversion of the temporal signal into some spatial representation. The resulting vector representation is then compared to a stored exemplar using some appropriate metric.

One common and simple representation of a temporal pattern is a buffer of time-delayed samples. This scheme has the advantage of simplicity; a vector of time-delayed samples can be easily extracted from a temporal signal. However, before the input vector can be classified it must first be time-aligned with any exemplar vector with which it is to be compared. This is required since temporal patterns might be presented at a slightly different speed on different occasions, and classification on unadjusted time-delayed sample vectors would be extremely brittle. This is true regardless of whether the samples are spectral, cepstral, or LPC coefficients, and independent of the classification and learning algorithms used. To address this issue, sophisticated alignment techniques such as Dynamic Time Warping (Sakoe and Chiba, 1978, Sakoe, 1979, Sakoe et al., 1989) have been proposed. The issue of time-warping is implicitly addressed by Hidden Markov Models by maximizing the probability of an appropriate coding sequence.

Unfortunately, Dynamic Time Warping is quite computationally expensive. It may therefore be advantageous to formulate a spatial representation of temporal patterns that is *speed-invariant*, and thus facilitates comparison of exemplars and inputs without requiring time-alignment. One such representation is proposed below.

2. Taylor Vector Representation

From the theory of Taylor series, we know that any sufficiently differentiable function can be represented as a series of the form

$$f(x) = f(a) + f'(a)(a-x) + \frac{f''(a)}{2}(a-x)^2 + \dots + \frac{f^{(n)}(a)}{n!}(a-x)^n \quad (1)$$

expanded around a point $x = a$. This approximation is accurate within a region around a whose size is dependent on the number of terms in the series and on the smoothness of the function $f(x)$. Since the only elements of the Taylor series that are dependent on the choice of the function are the derivatives, then a vector composed of these derivatives uniquely specifies the function $f(x)$ in a region around a .

Such a vector of derivatives can be useful for representing a temporal signal $F(t) = \alpha(\beta + f(\gamma t))$ (with amplitude α , translation β , and speed γ) because simple

manipulations on it can yield a speed-invariant form. For example, if each component of the vector except the first is divided by the absolute value of the previous component plus some positive constant ϕ , and then normalized by dividing each component by the absolute value of their sum plus some positive constant θ , we obtain the form:

$$\frac{\left[\frac{\text{sgn}(\alpha\gamma)f'(\gamma t)}{|\beta + f(\gamma t)| + |\phi/\alpha|}, \frac{\text{sgn}(\alpha)f''(\gamma t)}{|f'(\gamma t)| + |\phi/\alpha\gamma|}, \frac{\text{sgn}(\alpha\gamma)f'''(\gamma t)}{|f''(\gamma t)| + |\phi/\alpha\gamma^2|}, \dots \right]}{\frac{\theta}{|\gamma|} + \frac{|f'(\gamma t)|}{|\beta + f(\gamma t)| + |\phi/\alpha|} + \frac{|f''(\gamma t)|}{|f'(\gamma t)| + |\phi/\alpha\gamma|} + \frac{|f'''(\gamma t)|}{|f''(\gamma t)| + |\phi/\alpha\gamma^2|} + \dots} \quad (2)$$

where $\text{sgn}(x) = x/|x|$.

Note that given small values of θ and ϕ , the form of equation (2) is approximately invariant with respect to amplitude α and speed γ , but sensitive to the translation β and the sign of the speed and amplitude (The presence of the speed factor γ within the argument of $f(\cdot)$ does not affect speed-invariance). Larger values of θ and ϕ yield more amplitude and speed sensitivity. We call this representation a “normalized skewed Taylor vector”.

Various other forms are possible. For example, a form that is very amplitude and translation-sensitive can be constructed by prepending the form of equation (2) with the unmodified 0th order derivative.

The Taylor vector represents the shape of the function in a region near the point at which derivatives are calculated. The size of this region depends on the number of derivatives in the vector and the local smoothness of the function. Because the contribution of each successive derivative decreases and because calculation of successive derivatives becomes less accurate, it’s usually easy to estimate the number of terms one wishes to use. However, it may also be possible to determine the order adaptively by truncating the expansion when some high derivative changes too fast. Comparison between an exemplar and input vector with different numbers of terms is possible if the longer one is truncated and renormalized, since, at least with the form (2), lower terms do not depend upon higher terms.

When representing temporal functions using time-delayed samples, one must first choose an appropriate window size within which to collect samples. With the Taylor vector representation, no window is chosen, but rather an effective window size exists that is dependent upon the number of terms and the local smoothness. Thus for gradual functions the effective window size is large and for rapidly changing ones it is small. In fact, the window size is also dependent upon properties of the derivative calculation, since these always involve multiple samples in time. Choice of the order of derivatives saved and the effective window size is a subject for further study.

If a classification metric such as Euclidean distance is applied to vectors of the form (2) then patterns that are amplitude or time-scaled versions of each other will be grouped together. Furthermore, patterns which are time-warped versions of each other will remain close in Taylor vector space as long as the time-warping does not alter the shape of the function too radically. In effect, the representation implements a similarity metric that

meaningfully groups together patterns with the same general shape, resistant to the timing of their presentation.

However, the Taylor series approximation is only good if the function is sufficiently differentiable. Thus, grossly discontinuous functions cannot be well represented with a Taylor vector, even if many terms are used. In particular, any region that spans a discontinuity in any of the component derivatives will not expand consistently into a Taylor vector. This essentially means that the time window within which a particular vector applies is highly dependent upon the local continuity properties of the function: it is wide for smooth functions and narrow for functions with rapidly changing derivatives. This is unavoidable for any representation that relies upon local information.

3. Temporal Feature Detector

We call a “temporal feature detector” any system that indicates the occurrence of some pattern in time. Each such detector may be tuned to a specific exemplar pattern and give a response that is a function of the similarity of patterns in the input signal to its exemplar. For example, one detector may be tuned to the occurrence of U-shapes in the signal, another to ramp-and-hold shapes, still others to more complex temporal patterns.

The Taylor vector representation described above facilitates a meaningful metric for comparison of temporal shapes. Consider a detector of the form

$$o_j = e^{-20\|c_j - v\|^2} \quad (3)$$

where v is the current Taylor vector calculated from the input signal, c_j is the exemplar Taylor vector that detector j is tuned to, and o_j is the output of the detector. This output will be maximal when the input Taylor vector is close in shape to the exemplar vector.

The exemplar can be learned using various methods, including supervised and unsupervised learning. We will demonstrate detection of temporal exemplar patterns that were learned using a simple supervised algorithm.

For recognition of sampled waveforms, the temporal feature detector must first low-pass-filter its input. Otherwise the high-frequency information in the signal will make derivative calculation highly inaccurate. Second, it will need to calculate the derivatives from non-local information, i.e. from several time-delayed samples. Although different time-delayed samples will be used as presentation speed varies, the derivatives so constructed will still facilitate speed-invariant Taylor vector construction.

To obtain successive derivatives of a temporal signal we use sets of 2-dimensional differential equations

$$\begin{aligned} \dot{x}_i &= -x_i + x_{i-1} - y_i \\ \dot{y}_i &= \epsilon(x_{i-1} - y_i) \end{aligned} \quad (4)$$

where $i \in 1, 2, \dots, N$. Cohen and Grossberg (1993) have shown that given input $x_0(t) = f(t)$ then $x_1(t)$ is approximately proportional to the derivative of the input at low input frequencies. Thus a cascade of such computations will yield higher order derivatives. Due to the time-averaging function of the differential equations this method also serves to low-pass filter the input signal.

Figure 1a shows an example waveform generated by moving a slider on a computer display. This waveform was sampled at 250 samples per second and low-pass filtered to smooth out discontinuities. Let's suppose we want to train a feature detector to extract the downward step at time $t = 2.15$. We do this by calculating 7th order Taylor vectors of the form (2) during an interval around $t = 2.15$ and storing their time-average as the vector c_j . Now, we can present the input to the system, calculating the Taylor vector v at each point and observing the activity given by equation (3). This is shown in Figure 1c. Note that the detector responds for both the learned downward step at time $t = 2.15$ and a similar one at around $t = 3.05$. This is due to the use of an amplitude-insensitive Taylor vector form, i.e. equation (2) with $\theta = 0.0001$ and $\phi = 0.0001$.

Figures 1b and 1d show the recall of an identical pattern speeded up by a factor of 2. Note that the recall remains accurate, though more spurious peaks occur. This makes it clear that the choice of thresholds is crucial.

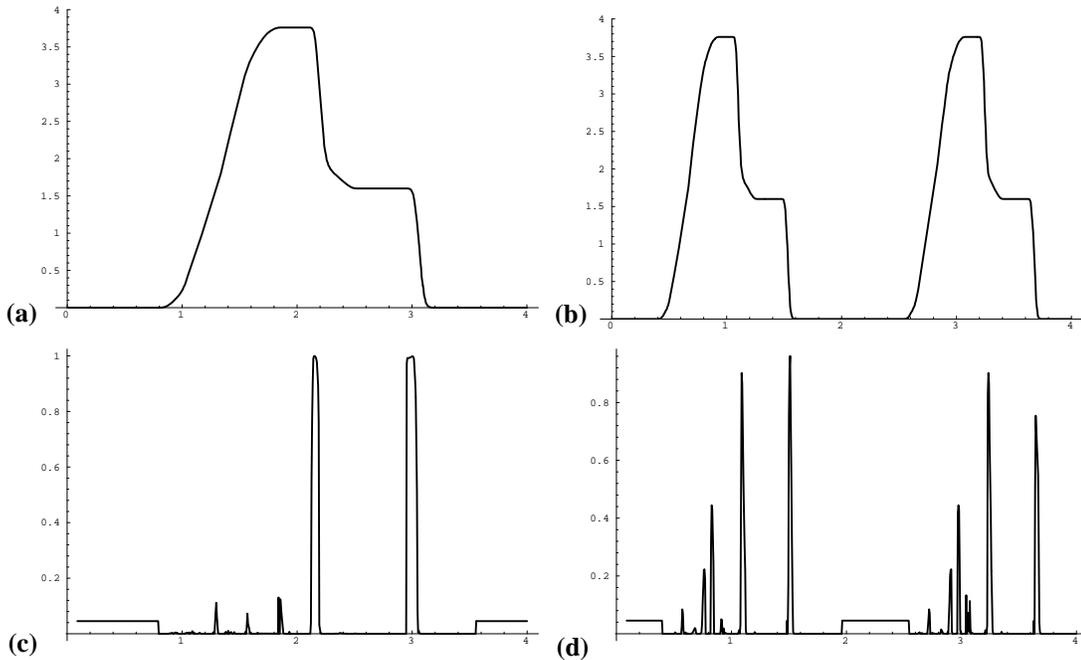


FIGURE 1. Recall of temporal features. (a) The original low-pass filtered pattern; (b) the same pattern speeded up by a factor of 2; (c) recall of the downward step from the original signal; (d) recall of the downward step from the speeded up signal.

Above we mentioned storing the time-average of the Taylor vector during an interval around the temporal feature of interest, rather than just storing a single vector calculated at a specific point in time. This is done because the vector moves through “Taylor vector space” fairly consistently, but noise in the signal and inaccuracies in derivative calculation cause it to fluctuate momentarily around its trajectory. Thus, a time-average of a portion of this trajectory will cancel out much of the effect of noise and produce a value that more properly identifies the shape of the function in that region. Comparison between a time-averaged exemplar and the time-average of the input is more likely to identify the gross features of the temporal pattern and be more resistant to inaccuracies in derivative calculation.

Issues for further investigation include adaptive methods for choosing the order of the vector and affecting the effective window size, other Taylor vector forms and their properties, and applications to signal processing.

References

- Anderson, S., Merrill, J. W. L., and Port, R. (1988). Dynamic speech categorization with recurrent networks. In *Proceedings of the 1988 Connectionist Models Summer School*. Morgan Kaufman.
- Cohen, M. A. and Grossberg, S. (1993). Parallel auditory filtering by sustained and transient channels separates coarticulated vowels and consonants. Technical Report CAS/CNS TR-93-051, Center for Adaptive Systems, Boston University, 111 Cummington Street, Boston MA 02215.
- Rabiner, L., Wilpon, J. G., and Soong, F. K. (1989). High performance connected digit recognition using Hidden Markov Models. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(8):1214–1225.
- Rabiner, L. R., Levinson, S. E., Rosenberg, A. E., and Wilpon, J. G. (1979). Speaker-independent recognition of isolated words using clustering techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-27(4):336–349.
- Sakoe, H. (1979). Two-level DP-matching - A dynamic programming-based pattern matching algorithm for connected word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-27(6):588–595.
- Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-26(1):43–49.
- Sakoe, H., Isotani, R., Yoshida, K., Iso, K., and Watanabe, T. (1989). Speaker-independent word recognition using dynamic programming neural networks. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 29–32.
- Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., and Lang, K. (1989). Phoneme recognition using time-delay neural networks. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*.