ISSN 0031-9155 Volume 39a 1994 Part 1

Physics in Medicine & Biology

Physics in Medicine & Biology

Abstracts of the World Congress on Medical Physics and Biomedical Engineering

RIO 94

21 - 26 August 1994 Rio de Janeiro Brazil

International Organization for Medical Physics

International Union for Physical and Engineering Science in Medicine

International Federation for Medical and Biological Engineering

X International Conference on Medical Physics
XVII International Conference on Medical and Biological Engineering

II National Forum of Science and Technology in Health
V Brazilian Congress of Physicists in Medicine
XIV Brazilian Congress on Biomedical Engineering
II Argentinean Congress on Bioengineering and Medical Physics

An official Journal of the International Organization for Medical Physics

OS12 - 2.6

THE ANALYSIS OF FACIAL PROFILES

J.C.Campos [1], A.D.Linney [1] and J.P.Moss [2]
Department of [1]Medical Physics and [2]Orthodontics, University College London,
1st floor Shropshire House, 11-20 Capper Street, London WC1E 6A, England.

[Introduction] The description of the face has, over the past few years, been an intensive topic of research receiving considreable attention in areas such as psychology, biostereometry, pattern recognition, forensic science, orthodontics and computer vision. In the case of clinical analysis of facial form, increasing iportance has been attached to the ability of orthodontic and surgical planning methods to provide a prediction in terms of facial appearance, considering its vital importance to patients undergoing facial surgery. Landmarks play an important role in segmenting the profile as a pre-cursor for analysis and an automatic method is thought to remove subjectivity inherent in this process so that quantitative analysis has a high repeatability. The work presented here has the aim of producing an objective method of identifying landmarks to segment the human profile into regions of interest to the clinicians(e.g. nose, chin, etc.), for the purposes of assessing changes in the profile due to surgery or growth. We are primarily interested not so much in relative movements of landmarks as in the changes in shape of the segments between the landmarks.
[Material and Methods] The approach adopted on this work combines the aspects of: the well recognized importance of curvature variation along the contour, the view of curvature as a result of processes acting on the shape, use of criteria that involve metric information extracted from curvature analysis and the use of contour segments bounded by perceptually relevent points(e.g. inflection and extremal points). The method uses Scale Space techniques, extensively explored in the fields of digital image and signal processing. These make use of filtering the signal across a continuum of scales by applying Gaussian filters and then tracking the extremal points and their derivatives as they move with scale changes, allowing the curvature values to be computed. The result is a description called Curvature Scale Space Image(CSSI), which is independent of profile orientation. The CSSI may be reduce to a simple interval tree representing a qualitative description of the profile simultaneoulsy at all scales. This description is then used to automatically identify and localize features in the facial profile. A difference metric is then derived using the following techniques: bending energy, spatial differences and curvature values. A medical graphics workstation is used for data analysis allowing the definition and extraction of a number of arbritrary sections(sets of x-y coordinate points) from a three dimensional model of the face generated from facial surface data. The data used in the analysis consited of mid-line facial profiles representing the pre and post-treatment states of a patient following surgery to the middle third of the face.
[Results] The results show the automatic segmentation of the profiles on a series of eight convex and concave curves corresponding to: soft tissue nasion, nose, nasio-labial fold, upper lip, mouth, lower lip, labio-mental fold and chin. The curvature value for each point along the profile is plotted against the path length and are used to quantify the changes occurred. The bending energy is used to express whether the segment has been elongated or compressed.
[Conclusion] The idea of segmenting the profile using scale space techniques proved to be efficient as it avoids the problems of identification of landmarks by using mathematically constructed points. The reproducibility of this method was tested by repeat recording and measurementes on several separet occasions. Although the contours in the CSSI may vary, the segments can always be identified. The curvature values within the segment gives a valuable shape description corresponding to the clinical perception of the profile.

OS12 - 2.7

COMPARISON BETWEEN WAVELETS AND FOURIER ANALYSIS AS SPEECH RECOGNITION PREPROCESSING TECHNIQUES

L. Rufiner, L. Gamero, M.E.Torres, D.Zapata, A. Sigura
Facultad de Ingeniería, Bioingeniería, Universidad Nacional de Entre Ríos - Ruta 11 Km 10, E.Ríos, Argentina-
TE (54)(43)230992 FAX (54)(43)230884 - Postal Address: CC 57 Suc. 3 (3100) Paraná

| Introduction | The use of neural networks for speech recognition was fundamentally oriented towards stationary patterns. To deal with the dynamic aspects of speech signal, the use of Time Delay Neural Network (TDNN) that takes simultaneous information of different instants was proposed. The good use of this feature depends strongly on how adequately these events can be presented to the network. In particular, the Discrete Wavelet Transform (DWT) is of interest for the analysis of non-stationary signals, because it provides an alternative to the classical Short Time Fourier Transform (STFT).
This paper studies the performance of analysis through DWT of speech signal against the STFT. By performance, in this context, we means the quality of preprocessing that makes the important characteristics of voice signal evident in order to achieve its automatics recognition through a TDNN. This improvement is seen as a decrease in training times or an increase in the recognition rate.
| Analysis Technique and Methods | As a recognition task, the speaker-dependent recognition of the diphones 'be', 'de' and 'ge' was chosen. Two training experiments on the same data have been prepared. Half of the data was used for training and the other half for validation. In the first case an FFT of 256 points with 20 ms Hamming window and 10 ms overlap were used as a preprocessing block. The quantity of points was chosen to obtain no more than 128 coefficients since this makes the network structure more complex. These results were compared with the DWT, with the same quantity of coefficients. We have considered the special case in which the basic functions are cubic polynomial splines. The network structure was the same in both experiments and they were trained through the backpropagation algorithm in identical situations.
| Results | One of the major drawbacks is the TDNN training times. According to preliminary results based on this reduced set of training sequences that included the diphones mentioned, a better performance of the Wavelet approach of about a 10% in the recognition rate for the same number of training cycles could be observed, after the stabilization of connection weights.
| Conclusions | The TDNNs has proved to bee efficient in speech recognition due to their ability to identify relationships among near transitory events. This is ostensibly improved with the use of preprocessing techniques oriented to the analysis of transitory signals such as the DWT in contrast to the classics techniques. To make a more complete assessment of the characteristics and advantages of the DWT for this type of tasks, it would be necessary to increase the quantity of training data and training cycles as well as to use diphones with a different transitory evolution.

COMPARISON BETWEEN WAVELETS AND FOURIER ANALYSIS AS SPEECH RECOGNITION PREPROCESSING TECHNIQUES

L. Rufiner, L. Gamero, M.E.Torres, D.Zapata, A. Sigura
Facultad de Ingeniería, Bioingeniería, Universidad Nacional de Entre Ríos - Ruta 11 Km 10, E.Ríos, Argentina-
TE (54)(43)230992 FAX (54)(43)230884 - Postal Address: CC 57 Suc. 3 (3100) Paraná.

[Introduction] The harmonic nature of the sound and the characteristics of the source of stimulus of the human speech production system lead us to think that these aspects should be analyzed in the frequency domain. But the voice signal is the product of a complex generative process whose parameters vary in time in a continuos form. For example, the particular movement of a formant in time is an important cue for identifying a voice stop. This fact reveals that it is necessary to relate temporal events represented in terms of frequency features.

The use of neural networks for speech recognition was fundamentally oriented towards stationary patterns. To deal with the dynamic aspects of speech signal, the use of Time Delay Neural Network (TDNN) that takes simultaneous information of different instants was proposed [1, 2]. This type of neural net allows to discover acoustic-phonetic features and their time relationships. The good use of this feature depends strongly on how adequately these events can be presented to the network. The TDNN learns decision surfaces automatically using error backpropagation. The morphology of the solution space and the separability of the classes are fundamentally important in order to decrease training time.

This paper studies the performance of analysis through Discrete Wavelet Transform (DWT) of speech signal against the Short Time Fourier Transform (STFT). By performance, in this context, we means the quality of preprocessing that makes the important characteristics of voice signal evident in order to achieve its automatics recognition through a TDNN. This improvement is seen as a decrease in training times or an increase in the recognition rate.

[Analysis Technique and Methods] Historically, digital processing of voice signals for coding, synthesis or recognition has been based on the adaptation of *long term* signal processing techniques to the analysis of nonstationary characteristics through the concept the short time analysis. Such is the case of the Fourier Transform and its STFT version, which, together with the Short Time Linear Prediction Coding (STLPC) analysis, have been widely applied in speech processing. In spite of the good results achieved, this type the analysis has not been originally thought for this type of task. That is why it is expected that better results will bee achieved using techniques specifically designed to deal with the transient aspects of the signal.

In particular, the Wavelet Transform (WT) is of interest for the analysis of non-stationary signals, because it provides an alternative to the classical STFT or to the Gabor Transform. The basic difference is as follows: in contrast to the STFT, which uses a single analysis window, the WT uses short windows at high frequencies and long windows at short frequencies. The wavelet decomposition of a continuous-time signal g(x) is an expansion of the form

$$g(x) = \sum_{i \in Z} \sum_{k \in Z} d_{(i)}(k)\, \psi(2^{-i} x - k)$$

where the basis functions are generated by dilatation and translation of a single prototype $\psi(x)$. The wavelet function $\psi(x)$ must satisfy certain properties: there must exists a linear one-to-one mapping between a function $g(x) \in L_2(R)$ and its wavelet coefficients $\{d_{(i)}(k), (i,k) \in Z^2\}$; this mapping defines the discrete wavelet transform. We have considered the special case in which the basic functions are cubic polynomial splines and we have used Fast computational algorithms [3, 4].

The discrete wavelet representation has a number of attractive features that have contributed to its recent growth in popularity among mathematicians and signal processors. First, its hierarchical

decomposition that enables the characterization of signal over scale (multiresolution analysis) [5]. Second, the wavelet transform is in essence a subband signal decomposition; in fact it is closely related to a variety of multirate decomposition techniques [6]. Finally, there is a fast wavelet transform algorithm. All these features have led us to the use of the multiresolution analysis as a preprocessing technique for the recognition of diphones through a neuronal network.

The performance of the system was assessed using TDNN. As a recognition task, the speaker-dependent recognition of the diphones 'be', 'de' and 'ge' was chosen. Two training experiments on the same data have been prepared. The emissions of one male speaker who pronounced each diphone 20 times at different moments were digitized totaling 60 emissions taken at 12 KHz, 16 bits. Half of the data was used for training and the other half for validation. In the first case an FFT of 256 points with 20 ms Hamming window and 10 ms overlap were used as a preprocessing block. The quantity of points was chosen to obtain no more than 128 coefficients since this makes the network structure more complex. These results were compared with the DWT, with the same quantity of coefficients. The network structure was the same in both experiments and they were trained through the backpropagation algorithm in identical situations. Each network output node correspond to each of the training diphones.

[Results] One of the major drawbacks is the TDNN training times. According to preliminary results based on this reduced set of training sequences that included the diphones mentioned, a better performance of the Wavelet approach of about a 10% in the recognition rate for the same number of training cycles could be observed, after the stabilization of connection weights.

[Conclusions] The TDNNs has proved to bee efficient in speech recognition due to their ability to identify relationships among near transitory events. This is ostensibly improved with the use of preprocessing techniques oriented to the analysis of transitory signals such as the wavelet transform in contrast to the classics techniques. To make a more complete assessment of the characteristics and advantages of the WT for this type of tasks, it would be necessary to increase the quantity of training data and training cycles as well as to use diphones with a different transitory evolution.

[References]

1. Waibel A., Hanazawa T., Hinton G., Shikano K. and Lang J., "Phoneme Recognition Using Time Dealy Neural Networks", IEEE Trans. on Acoustic Speech and Signal Processing; Vol ASSP 37, No 3, 1989.
2. Waibel A. H. Sawai, and Shikano, "Consonant recognition by modular construction of large phonemic time-delay neural networks", Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Glasgow, Scotland, vol. 1, pp. 112-115, 1989.
3. Unser M., Aldroubi A., Murray E., A family of polinomial spline wavelet Transform, Signal Processing 30, 141-162, Elsevier, 1993.
4. Unser M., Aldroubi A. and Eden M., "Fast B-Spline transform for Continuous Image Representation and Interpolation", IEEE Trans. "Pattern Anal. and Machine Intell., Vol 13, No 3, pp 277-285, March 1991.
5. Mallat, S.G. "A theory of multiresolution of signal decpomposition: the wavelet representation", IEEE Trans. Patt.Anal. Mach.Intell.; Vol.PAMI-11, Nº 7, 1989.
6. Rioul O. and Vetterli M., Wavelets and Signal Processing, IEEE Signal Processing Magazine, October 1991.