34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Augmenting the Decomposition of EMG Signals Using Supervised Feature Extraction Techniques

Hossein Parsaei, Mehrdad J. Gangeh, *Member*, *IEEE*, Daniel W. Stashuk, *Member*, *IEEE*, and Mohamed S. Kamel, *Fellow*, *IEEE*

Abstract— **Electromyographic (EMG) signal decomposition is the process of resolving an EMG signal into its constituent motor unit potential trains (MUPTs). In this work, the possibility of improving the decomposing results using two supervised feature extraction methods, i.e., Fisher discriminant analysis (FDA) and supervised principal component analysis (SPCA), is explored. Using the MUP labels provided by a decomposition–based quantitative EMG system as a training data for FDA and SPCA, the MUPs are transformed into a new feature space such that the MUPs of a single MU become as close as possible to each other while those created by different MUs become as far as possible. The MUPs are then reclassified using a certainty–based classification algorithm. Evaluation results using 10 simulated EMG signals comprised of 3-11 MUPTs demonstrate that FDA and SPCA on average improve the decomposition accuracy by 6%. The improvement for the most difficult-to-decompose signal is about 12%, which shows the proposed approach is most beneficial in the decomposition of more complex signals.**

I. Introduction

An Electromyographic (EMG) signal acquired during a muscle contraction is the superposition of background noise and motor unit potential trains (MUPTs) of the motor units (MUs) that are active throughout the contraction. EMG signals contain valuable information on activity, state of health, and the characterization of the muscle from which they are detected [1]. For example, the MU firing pattern of these MUs can assist with the better understanding of the neural control of movement [2]. The characteristics (amplitude, duration, complexity, number of phases, and number of turns) of the motor unit potential (MUP) templates of the MU can assist with the diagnosis of neuromuscular disorders [1]. An efficient technique for extracting such information is EMG signal decomposition.

EMG signal decomposition is a process by which MUPTs are extracted from an EMG signal such that each extracted MUPT estimates the actual MUPT generated by a single MU. The purpose of EMG signal decomposition is to provide an estimate of the firing pattern and MUP template of each active MU that contributed significant MUPs to the EMG signal. In general, EMG signal decomposition is accomplished by employing digital signal processing and pattern recognition techniques in four/five steps [3]: signal preprocessing, signal segmentation and MUP detection, feature extraction, clustering of detected MUPs and/or supervised classification of detected MUPs . The first two steps are to remove noise from the signal, make the MUPs into narrow spikes, and detect the MUPs of the MUs that contribute to the signal. The last three steps are to group the detected MUPs into several MUPTs such that each MU has only one corresponding MUPT in the decomposition results.

Numerous EMG signal decomposition algorithms have been developed [3]. In fact, various feature extraction, clustering and supervised classification techniques have been employed to reduce the level of false-classification error (FCE) rate and missed-classification error (MCE) rate of the extracted MUPTs. This paper demonstrates the possibility of improving the decomposition results using Fisher discriminant analysis (FDA) and supervised principal component analysis (SPCA). Specifically, we will show that the combined FDA or SPCA and a certainty-based classifier (CBC) technique [4] will improve both FCE and MCE rates in the extracted MUPTs.

A challenge in using SPCA and FDA for EMG decomposition is that these techniques need labeled (training) data while such data is not available *a priori*. In fact, in EMG decomposition, neither the number of MUPTs nor the labels of the MUPs are known in advance. In this work, the decomposition results provided by a decomposition-based quantitative EMG (DQEMG) system [5] is used as training data for FDA and SPCA. Following is a brief discussion of the steps of using FDA or SPCA for improving EMG decomposition accuracy.

II. Methods

The presented decomposition system involves three steps. First, the signal is decomposed using the DQEMG system [5]. Second, the class assignments are used to estimate parameters required for FDA/SPCA and transform the MUPs into a new feature space. Third, the MUPs are reclassified using the CBC algorithm in the new feature space.

A. DQEMG System

The DQEMG system decomposes a given EMG signal off-line in the five steps discussed above. Here, we briefly explain this system and more details can be found in [5].The signal is filtered using a first-order low pass difference filter to decrease MUPs temporal overlap, sharpen MUPs, and increase the differences between MUPs created by different MUs. To identify the positions of suitable MUPs in the prefiltered signal, the signal is scanned for the peaks that satisfy several criteria

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

H. Parsaei was with the Dept. of Systems Design Eng., University of Waterloo, Waterloo, ON, N2L 3G1, Canada. He is now with the Technical & Eng. College of Shiraz, Vocational & Technical University, Iran (e–mail: hparsaei@uwaterloo.ca)

D. W. Stashuk is with the Dept. of Systems Design Eng., University of Waterloo,Waterloo, ON N2L 3G1, Canada (e–mail: stashuk@uwaterloo.ca)

M. J. Gangeh and M. S. Kamel are with the Dept. of Electrical and Computer Eng., University of Waterloo, Waterloo, ON N2L 3G1, Canada (e–mail: {mgangeh, mkamel}@pami.uwaterloo.ca)

(e.g., having suitable peak values) [5]. Each detected MUP is then represented by 2.56 ms of filtered data (i.e., 80 samples at a 31250 Hz sampling rate), centered about the position of its peak. These 80 time-samples are used as a feature vector representing each MUP and, ultimately, to sort MUPs into several MUPTs.

Grouping the detected MUPs into sets of MUPTs are conducted using a clustering and then a supervised classification algorithm in several iterations. In general, the objective is to assign a given MUP to a MUPT with which its time of occurrence and shape are more consistent than the firing pattern and MUP shape of the other MUPTs. The purpose of clustering is to provide the necessary initial information required for supervised classification such as estimates of the number of MUPTs, their prototypical MUP shapes, and their MU firing pattern statistics. In DQEMG, MUP clustering is conducted using a shape and temporal-based clustering (STBC) algorithm. The STBC is partially based on *k*-means clustering algorithm, i.e., it groups MUPs based on their MUP shape similarity. Moreover, the firing pattern information is used to test the validity of the assignments in STBC [6].

Using the information obtained regarding possible MUPTs of the given EMG signal, the remaining unclassified MUPs are assigned to the extracted MUPTs via the CBC algorithm that employs both MUP shape and MU firing pattern information to determine the class label of a MUP. The CBC algorithm first identifies the MUPTs with the most and the next most similar MUP templates to the MUP to be classified (denoting by M) using Euclidean distance similarity measure. Then it calculates the certainty values of assigning M to one of these two trains. The certainty values are calculated by combining MU firing pattern (C_{FC}) and MUP shape (C_{SC}) certainties as

$$C_i = C_{\mathrm{FC}_i} \times C_{\mathrm{SC}_i}; i = 1, 2 \tag{1}$$

where C_i is the certainty of assigning M to MUPT_i which is one of the two closest MUPTs to M. Firing pattern certainty C_{FC_i} measures the consistency of the occurrence time of M relative to the established MU firing pattern of MUPT_i. MUP shape certainty C_{SC_i} measures the consistency of the shape of the M to that of the MUPs in MUPT_i and is estimated by multiplying normalized absolute shape certainty (C_{ND}) and relative shape certainty (C_{RD}). The C_{ND} represents the distance from M to the template of a train, normalized by the energy of the template. The C_{RD} represents the distance from M to the most similar MUP template relative to the distance of M to the next most similar MUP template.

Having the C_i values calculated, the M is assigned to the MUPT which has the greatest C_i values, if $\max(C_1, C_2)$ is greater than a certainty threshold (C_{AT}). Otherwise, M is left unassigned.

B. Fisher Discriminant Analysis

Fisher discriminant analysis (FDA) is a classical supervised dimensionality reduction algorithm, which was first introduced by Fisher [7]. FDA looks for a transformation U that projects data to a subspace where data samples from different classes are as far as possible from each other and those inside a class as close as possible. This can be formulated using the within-class (S_W) and between-class (S_B) covariance matrices. In other words, the optimization problem in FDA is as follows

$$\arg\max_U \frac{U^{\mathrm{T}} S_B U}{U^{\mathrm{T}} S_W U}. \tag{2}$$

The solution for this optimization problem is the eigenvectors of $S_W^{-1} S_B$ [8]. The dimensionality of obtained subspace is at most the same as the number of classes.

C. Supervised Principal Component Analysis

Principal component analysis (PCA) is a classical unsupervised dimensionality reduction technique that projects data into a maximum variance subspace. Representing data in this projected subspace is optimal in reconstructing data in mean-squared-error sense (for example see [8]). This is important in applications such as denoising and coding. However, this representation is not optimal for classification tasks. This is mainly because PCA is an unsupervised approach that does not take into account the category information in finding the low dimensional subspace.

To overcome this major drawback of PCA, which is important in classification tasks, supervised PCA (SPCA) has been recently proposed [9]. Since SPCA is based on Hilbert-Schmidt independent criterion (HSIC), we first briefly explain HSIC and then describe the formulation for SPCA.

HSIC is a kernel-based method to measure the dependency of two random variables $\mathcal{X}$ and $\mathcal{Y}$ [10]. To this end, HSIC computes the Hilbert-Schmidt norm of the cross-covariance operators in reproducing kernel Hilbert spaces (RKHSs) [10, 11]. Since its introduction, the HSIC has been used in many applications including feature selection [12], independent component analysis [13], and sorting/matching [14].

For practical purposes, HSIC has to be estimated using a finite number of data samples. Consider $\mathcal{Z} = [(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_n, \mathbf{y}_n)] \subseteq \mathcal{X} \times \mathcal{Y}$ as n independent observations drawn from $P_{\mathcal{X}\times\mathcal{Y}}$. An empirical estimate of HSIC is as follows

$$HSIC(\mathcal{Z}) = \frac{1}{(n-1)^2}\mathrm{tr}(KHLH), \tag{3}$$

where tr is the trace operator, $H, K, L \in \mathbb{R}^{n\times n}$, $K_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$, $L_{i,j} = l(\mathbf{y}_i, \mathbf{y}_j)$, and $H = I - n^{-1}\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}$ (I is the identity matrix, $\boldsymbol{e}$ is a vector of n ones, and hence H is the centering matrix). It is important to notice that according to (3), to maximize the dependency between two random variables $\mathcal{X}$ and $\mathcal{Y}$, the empirical estimate of HSIC, i.e., $\mathrm{tr}(KHLH)$ should be maximized. Next, we describe SPCA using HSIC.

Essentially, SPCA addresses the problem of finding an orthonormal transformation U (i.e., $U^{\mathrm{T}}U = I$), which projects data to a space where the dependency between data and their corresponding labels is maximum. Thus, considering a finite training set of n data points, each of which consisting of d features, i.e., $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d\times n}$, the objective is to find the subspace $U^{\mathrm{T}}X$ where the dependency of projected data $U^{\mathrm{T}}X$ is maximized with respect to the labels Y. Based on what was explained above, this can be done using HSIC by maximizing $\mathrm{tr}(KHLH)$, where K is a kernel defined on projected data $U^{\mathrm{T}}X$ such as $K = X^{\mathrm{T}}UU^{\mathrm{T}}X$, and L is a kernel on the labels Y, e.g., $L = Y^{\mathrm{T}}Y$.

TABLE I. The performance of the DQEMG compared to that of the FDAEMGD and SPCAEMGD for 10 simulated EMG signals.

Signal	Intensity (pps)	No. of MUPTs	Jitter (μs)	IDI_CV	DQEMG			FDAEMGD			SPCAEMGD		
					A_r (%)	A_c (%)	CC_r (%)	A_r (%)	A_c (%)	CC_r (%)	A_r (%)	A_c (%)	CC_r (%)
1	30.5	3	100	0.15	92.1	97.2	89.5	97.4	99.3	96.7	97.7	98.7	96.4
2	41.8	5	100	0.15	87.8	97.5	85.6	92.4	97.9	90.5	94.5	98.0	92.6
3	45.6	4	50	0.15	88.8	93.1	82.7	91.5	96.4	88.2	95.6	96.6	92.3
4	54.0	6	50	0.15	92.6	98.0	90.7	94.1	98.2	92.4	94.8	98.1	93.0
5	59.4	7	100	0.15	90.7	95.9	87.0	93.6	96.9	90.8	92.8	97.1	90.1
6	68.2	7	50	0.15	90.2	96.4	86.9	92.4	96.8	89.4	92.5	97.5	90.2
7	82.5	8	100	0.15	83.5	89.8	75.0	92.0	93.7	86.2	90.6	91.4	82.8
8	85.2	9	50	0.15	89.3	87.3	78.0	89.8	89.9	80.8	88.7	89.8	79.7
9	97.5	10	100	0.15	86.0	84.8	74.3	88.0	93.2	82.1	87.1	93.3	81.3
10	105.2	9	50	0.15	80.6	87.1	70.2	89.3	92.5	82.6	87.0	92.7	80.6
		Mean			88.2	92.7	82.0	92.1	95.5	88.0	92.1	95.3	87.9
		STD			3.8	5.0	7.1	2.7	3.0	5.1	3.7	3.2	6.1

Replacing these kernels on projected data and labels in $\mathrm{tr}(KHLH)$, and knowing that the value of trace operator is invariant to the rotation of its arguments, we obtain

$$\begin{aligned}\mathrm{tr}(KHLH) &= \mathrm{tr}(X^{\mathrm{T}}UU^{\mathrm{T}}XHY^{\mathrm{T}}YH)\\ &= \mathrm{tr}(UU^{\mathrm{T}}XHY^{\mathrm{T}}YHX^{\mathrm{T}})\\ &= \mathrm{tr}(U^{\mathrm{T}}XHLHX^{\mathrm{T}}U)\end{aligned} \tag{4}$$

Now, the optimization problem is to find the orthonormal transformation U, which maximizes this trace function, i.e.,

$$\begin{aligned}&\underset{U}{\mathrm{argmax}}\ \mathrm{tr}(U^{\mathrm{T}}XHLHX^{\mathrm{T}}U)\\ &\mathrm{s.t.}\ \ U^{\mathrm{T}}U = I\end{aligned} \tag{5}$$

where the constraint of the optimization problem is due to the orthonormality of the transformation U.

It is well known that the solution for this optimization problem is the eigenvectors of $\varphi = XHLHX^{\mathrm{T}}$. Hence, if we want the subspace to be of p dimension, where $p \leq d$, then transformation U is consisting of the top eigenvetors corresponding to the largest eigenvalues of φ.

Both SPCA and FDA need labeled data. Therefore, these methods in general cannot be directly applied to EMG decomposition problem because such training data is not available *a priori*. In this work, the classification results provided by the DQEMG system are used as training data for FDA and SPCA. Specifically, the MUPs that their C_i values > 0.5 are used to estimate the parameters of the FDA and SPCA that ultimately transform the MUPs into a new feature space. The FDA/SPCA will improve the separability of MUPs created by different MUs and subsequently will improve the classification of the MUPs. In the new feature space, a final inspection is made on the MUPs. However, to speed up decomposition process, reclassifying the MUPs whose certainty values are smaller than 0.5 has been only considered. The CBC algorithm is used to classify MUPs where the certainties for assigning each of these MUPs are re-calculated as explained in Section II-A. The MUP will be moved from its current MUPT, to the MUPT that has the greatest new certainty value, if the new certainty value is greater than *maximum* of C_{AT} and current certainty value. In the remaining of this paper, the decomposition algorithm that uses FDA to transform MUPs into a new feature space is denoted as FDAEMGD (FDA-based EMG decomposition) and the algorithm that employs SPCA is called SPCAEMGD (SPCA-based EMG decomposition).

III. Results and Discussion

The effectiveness of the FDA and SPCA in improving the decomposition of EMG signals was studied using 10 simulated EMG signals that were generated using a physiologically–based EMG signal simulation algorithm [15]. This EMG simulator creates intramuscular EMG signals with different complexities such as different numbers of active MUs, different degrees of MUP shape and/or firing pattern variability (represented by the amount of jitter and coefficient of variation (CV) of intern discharge interval (IDI)), and different signal intensities (represented by the average number of MUPs per second (pps)). The parameters of the 10 signals used in this work are given in columns 2 to 5 of Table I.

The following three performance measure indices were used to study the advancement achieved by using FDA or SPCA.

$$\mathrm{A_r}\% = \frac{\text{Number of MUPs assigned}}{\text{Total number of MUPs detected}} \times 100 \tag{6}$$

$$\mathrm{A_c}\% = \frac{\text{Number of MUPs correctly classified}}{\text{Total number of MUPs assigned}} \times 100 \tag{7}$$

$$\mathrm{CC_r}\% = \frac{\text{Number of MUPs correctly classified}}{\text{Total number of MUPs detected}} \times 100 \tag{8}$$

The two indices $\mathrm{A_r}$ and $\mathrm{A_c}$, in fact, express the level of MCE rate and FCE rate in the MUPTs obtained from decomposing a signal, respectively. High value of $\mathrm{A_r}$ shows low MCE rate level, likewise high value of $\mathrm{A_c}$ indicates low FCE rate.

The results for DQEMG, FDAEMGD, and SPCAEMGD applied to the 10 EMG signals are summarized in Table I. The overall mean and standard deviation (STD) for the three performance indices used are provided as well. The statistical comparison of the average values was conducted using analysis of variance (ANOVA) where $\alpha = 5\%$ and the Tukey-Kramer honestly significant difference test has been used for pair-wise comparison of the mean values.

Based on the results presented in Table I, the FDA and SPCA performed similarly on the signals used. Compared to the DQEMG, a significant improvement was achieved in $\mathrm{A_r}$ by using FDA or SPCA. The accuracies of the extracted MUPTs ($\mathrm{A_c}$) are slightly improved. Such improvements, can lead to a better estimation of MUP templates and MU firing patterns of the MUs because the accuracy of the error filtered estimation algorithm [16] in estimating MU firing pattern statistics increases when the MCE rate in the train decreases [16].

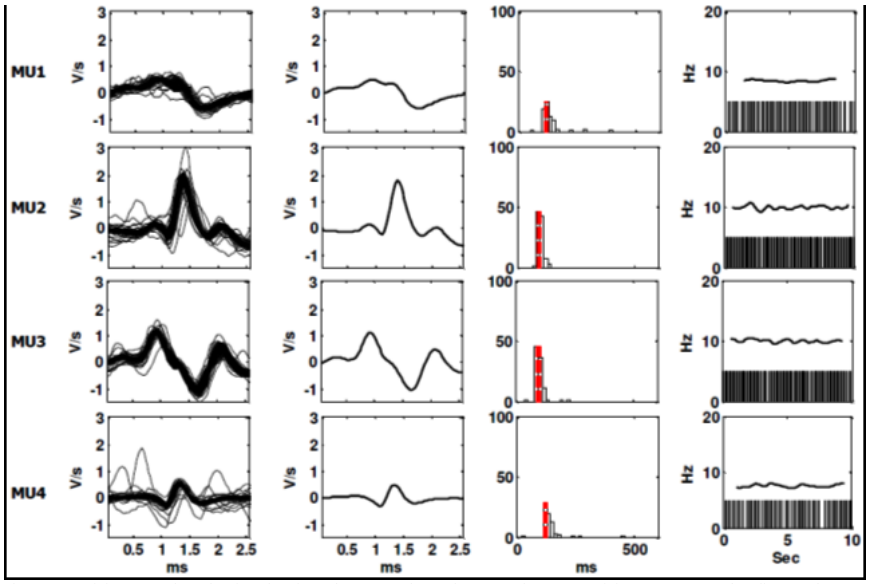

Figure 1. An example of decomposition results obtained using the FDA-based EMG decomposition system. The first column shows the shimmer plot of the slope of the MUPs assigned to each MUPT. The second column shows the template for these MUPs. The third column presents the IDI histogram of each MU. Finally, the last column presents the firing pattern of each MU.

The improvement in decomposition results increases with the complexity of the signal. For example, the improvement in CC_r (as an example) for signal #10 is 12.4% which is about 2 times of the average improvement in CC_r while for signal # 1, which has the highest decomposability among these 10 signals, is about 7.0%.

Figure1 presents an example of the decomposition results obtained by the FDAEMGD. The histogram of the IDIs and also the firing pattern plots reveal that the obtained MUPTs have low MCE and FCE rates.

The presented FDA- and SPCA-based decomposition systems are for decomposing intramuscular EMG signals mainly for clinical applications where the MUP template and mean MU firing rate for each MUPT are required. Since these parameters can be estimated from incomplete MUPTs, superimposed MUPs were not resolved. In fact, the majority of MUPs left unassigned are superimposed MUPs.

The main disadvantages of the system developed using FDA and SPCA is that it is computationally more complex than the DQEMG. Calculating the covariance matrix of the MUP features and its singular value decomposition takes time. Nevertheless, the system is still fast enough to be used in clinical environments.

IV. Conclusion

In this paper, we showed that using FDA and SPCA could lead to an improvement in EMG decomposition results. Specifically, these two supervised feature extraction techniques assisted with significantly reducing the MCE rates in the MUPTs obtained by the DQEMG. The FCE rates in the MUPTs are improved too but not as much as the MCE rate. Finally, both FDA and SPCA performed almost the same on the signals used in this work.

References

[1] C. Farkas, A. Hamilton-Wright, H. Parsaei, and D. W. Stashuk, "A review of clinical quantitative electromyography," *Crit Rev Biomed Eng*, vol. 38, no. 5, pp. 467–485, 2010.

[2] P. Contessa, A. Adam, and C. J. De Luca, "Motor unit control and force fluctuation during fatigue," *J Appl Physiol*, vol. 107, no. 1, pp. 235–243, Jul. 2009.

[3] H. Parsaei, D. W. Stashuk, S. Rasheed, C. Farkas, and A. Hamilton-Wright, "Intramuscular EMG Signal Decomposition," *Crit Rev Biomed Eng*, vol. 38, no. 5, pp. 435–465, 2010.

[4] D. Stashuk and G. Paoli, "Robust supervised classification of motor unit action potentials," *Med Biol Eng Comput.*, vol. 36, no. 1, pp. 75–82, Jan. 1998.

[5] D. W. Stashuk, "Decomposition and quantitative analysis of clinical electromyographic signals," *Med Eng Phys.*, vol. 21, no. 6, pp. 389–404, Jul. 1999.

[6] D. Stashuk and Y. Qu, "Adaptive motor unit action potential clustering using shape and temporal information," *Med Biol Eng Comput.*, vol. 34, no. 1, pp. 41–49, Jan. 1996.

[7] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics* vol. 7, pp. 179–188, 1936.

[8] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. 2nd ed., New York: John Wiley and Sons, 2001.

[9] E. Barshan, A. Ghodsi, Z. Azimifar, M. Z. Jahromi, "Supervised principal component analysis: visualization, classification and regression on subspaces and submanifolds," *Pattern Recognition*, vol. 44, no. 7, pp. 1357–1371, 2011.

[10] A. Gretton, O. Bousquet, A. J. Smola, B. Scholkopf, "Measuring statistical dependence with Hilbert–Schmidt norms," in: *Proceedings Algorithmic Learning Theory (ALT)*, vol. 3734, 2005, pp.63–77.

[11] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, no. 3, pp. 337–404, 1950.

[12] L. Song, J. Bedo, K. M. Borgwardt, A. Gretton, and A. J. Smola, "Gene selection via the BAHSIC family of algorithms," *Bioinformatics*, vol. 23, pp. i490–i498, 2007.

[13] H. Shen, S. Jegelka, and A. Gretton, "Fast kernel-based independent component analysis," *IEEE Trans Signal Processing*, vol. 57, no. 9, pp. 3498–3511, 2009.

[14] N. Quadrianto, A. J. Smola, L. Song, and T. Tuytelaars, "Kernelized sorting," *IEEE Trans on Pattern Analysis and Machine Intelligence*, vol. 32, no. 10, pp. 1809–1821, 2010.

[15] A. Hamilton-Wright and D. W. Stashuk, "Physiologically based simulation of clinical EMG signals," *IEEE Trans Biomed Eng*, vol. 52, no. 2, pp. 171-183, 2005.

[16] D. W. Stashuk and Y. Qu, "Robust method for estimating motor unit firing-pattern statistics," *Med Biol Eng Comput.*, vol. 34, no. 1, pp. 50–57, Jan. 1996.