Journal of Zhejiang University SCIENCE B
ISSN 1673-1581 (Print); ISSN 1862-1783 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Characterization of surface EMG signals using improved approximate entropy*

CHEN Wei-ting, WANG Zhi-zhong†, REN Xiao-mei
(*Department of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*)
†E-mail: zzwang@sjtu.edu.cn

Received Jan. 13, 2006; revision accepted Apr. 10, 2006

Abstract: An improved approximate entropy (ApEn) is presented and applied to characterize surface electromyography (sEMG) signals. In most previous experiments using nonlinear dynamic analysis, this certain processing was often confronted with the problem of insufficient data points and noisy circumstances, which led to unsatisfactory results. Compared with fractal dimension as well as the standard ApEn, the improved ApEn can extract information underlying sEMG signals more efficiently and accurately. The method introduced here can also be applied to other medium-sized and noisy physiological signals.

Key words: Surface EMG (sEMG) signal, Nonlinear analysis, Approximate entropy (ApEn), Fractal dimension
doi:10.1631/jzus.2006.B0844 **Document code:** A **CLC number:** R318.04

INTRODUCTION

Surface EMG (sEMG) signals recorded from skin surface have been widely used in fields such as prosthesis control, rehabilitation, muscle fatigue analysis and clinical diagnosis (Chang *et al*., 1996; Park and Stelmach, 2006; Zijdewind *et al*., 1998; Doorenbosch and Harlaar, 2004; Georgakis *et al*., 2003; Abel *et al*., 1996). When it comes to controlling a prosthesis what action to carry out next, recognizing the rehabilitation condition of post-operational muscles, or judging the degree of muscle fatigue (Sparto *et al*., 2000) etc., it is of great importance to accurately extract the signature information from sEMG signals.

Many parameters have been successfully applied to classify sEMG signals (Jang *et al*., 1994; Zardoshti-Kermani *et al*., 1995; Kang *et al*., 1995; 1996; Gupta *et al*., 1997; Englehart *et al*., 1999; Hu *et al*., 2005; Crawford *et al*., 2005; Huang *et al*., 2005; Kim *et al*., 2006), among which nonlinear dynamical analysis is a quite powerful approach. The calculations of most nonlinear dynamic measures, however, are frequently confronted with insufficient data points and noisy backgrounds. Wolf *et al*.(1985) pointed out that the required data points to get a reasonable estimate of correlation dimension were up to 30^m for an m-dimensions attractor. In 1991, Pincus (1991) proposed approximate entropy (ApEn) for a measure of system complexity that is applicable to noisy and short datasets with data points typically between 100 and 5000 (Pincus and Goldberger, 1994). Given N points, $ApEn(m, r, N)$ is approximately equal to the negative average natural logarithm of the conditional probability that vectors similar for m points remain similar at the next point, where the similarity of two vectors is judged in terms of their absolute coordinates.

We present here an improved ApEn where two vectors' similarity is based on their shapes rather than their absolute coordinates, and then employed it to extract features from sEMG signals. Results showed that, compared with fractal dimension as well as the standard ApEn, the improved ApEn can more effi-

* Project supported by the National Natural Science Foundation of China (No. 60171006) and the National Basic Research Program (973) of China (No. 2005CB724303)

ciently characterize the same two different patterns of sEMG signals once dealt with by Hu *et al*.(2005).

MATERIALS AND METHODS

Subjects and data acquisition

The signal collection was completed at the EMG room of Shanghai Huashan Hospital in China. Thirty healthy volunteers participated in this study, all of whom gave their informed consent before participating. Each subject was requested to execute two different kinds of actions: forearm supination (FS) and forearm pronation (FP), during which sEMG signals were collected. Skin surface of the area of interest was abraded with alcohol beforehand, and two sets of 5-mm diameter discs electrodes were placed over the flexor carpi radialis and the extensor carpi radialis longus on the right forearm. Signals were digitally sampled at a rate of 1000 Hz and the bandwidth of the amplifier-filter was 10 Hz to 500 Hz.

Improved ApEn

For a time series of N data $\{u(i):1\leq i\leq N\}$, form vector sequences $\boldsymbol{x}(1)$ through $\boldsymbol{x}(N-m+1)$ as follows:

$$\boldsymbol{x}(i)=\{u(i), u(i+1), \ldots, u(i+m-1)\}. \quad (1)$$

Here, $\boldsymbol{x}(i)$ represents m consecutive u values, starting with the ith point. Generalize vector sequences $\{\boldsymbol{x}(i), 1\leq i\leq N-m+1\}$ by removing a baseline as follows:

$$\begin{aligned}\boldsymbol{V}(i)&=\{u(i)-v_0(i), u(i+1)-v_0(i), \ldots, u(i+m-1)-v_0(i)\}\\&=\{v(i), v(i+1), \ldots, v(i+m-1)\},\end{aligned} \quad (2)$$

where the baseline can be obtained by averaging the m consecutive u values in vector $\boldsymbol{x}(i)$

$$v_0(i)=\frac{\sum_{j=0}^{m-1}u(i+j)}{m}. \quad (3)$$

Define the distance $d[\boldsymbol{V}(i), \boldsymbol{V}(j)]$ between vectors $\boldsymbol{V}(i)$ and $\boldsymbol{V}(j)$

$$d_{ij}=d[\boldsymbol{V}(i),\boldsymbol{V}(j)]=\max_{k\in(0,m-1)}|v(i+k)-v(j+k)|. \quad (4)$$

Given tolerance r, construct $C_i^m(r)$ for each $i\leq N-m+1$

$$C_i^m(r)=\frac{\text{Number of } j\ (j\leq N-m+1 \text{ and } d_{ij}\leq r)}{N-m+1}. \quad (5)$$

Average the natural logarithms of $C_i^m(r)$ over i

$$\phi^m(r)=\frac{1}{N-m+1}\sum_{i=1}^{N-m+1}\ln C_i^m(r). \quad (6)$$

Then the ApEn of the sequence can be estimated by

$$ApEn(m, r, N)=\phi^m(r)-\phi^{m+1}(r). \quad (7)$$

Choices of parameters

How the input parameters are chosen (Pincus and Goldberger, 1994) gives rise not only to statistical estimation issues but also to parameter issues. The parameters m and r must be fixed for each calculation of ApEn, where m is the length of compared runs, and r is the tolerance for accepting matches. A choice of m=2 is superior to m=1, in that it allows more detailed reconstruction of the joint probabilistic dynamics of the process. Typically, m>2 is unfavorable due to the need of very large values of N or r, while N is often small for a physiological dataset, and too large an r value may lead to the loss of detailed information. Accordingly, we set m=2 in our experiments.

As an effective filter, the tolerance r should also be carefully selected. Too small an r value will result in salient influence from noise, while too large an r value, as mentioned above, is supposed to be avoided for fear of information loss. Pincus (1995) concluded that a value between (0.1~0.25)×SD of r produces good statistical validity of $ApEn(m, r, N)$ for m=1 and m=2 when dealing with most datasets, where SD is the standard deviation of the original datasets. We found in our experiments that a value of r=0.2SD worked well for the characterization of sEMG signals, which was thus set as the tolerance in our calculation of ApEn.

RESULTS AND DISCUSSION

For the two patterns of sEMG signals (Hu *et al*.,

2005), we first calculated the approximate entropy values using the improved ApEn, and then compared the results with those of fractal dimension. The approximate entropy values of FP and FS signals distribute among 0.7~0.9 and 1.1~1.3 respectively. No mix region exists in the approximate entropy values of the two motions, and the division of the two clusters is apparent (Fig.1). Though fractal dimensions (Hu *et al*., 2005) of filtered FP and FS signals can be distinguished more possibly than those of raw FP and FS signals (Fig.2), there are still some mix points of the two clusters, so it is somewhat ambiguous when judging which cluster some of the points fall in. Comparing the parameters of approximate entropy with those of fractal dimension when characterizing the two motions graphically, we can see that the former works better than the latter.

In addition, the efficiency of the two parameters as features of sEMG signal classification was also tested on the basis of two statistics (Hu *et al*., 2005), namely, dispersion of the clusters and distance between different cluster centers. The lower the dispersion and the larger the distance are, the more appropriate a parameter is to be a feature. Employing standard deviation in one cluster as the estimation of cluster dispersion and the average value as the estimation of cluster center, we calculated these two statistics of the two parameters (Table 1). The cluster-to-cluster distance of fractal dimension appears much larger than that of approximate entropy, but it does not necessarily mean that fractal dimension is superior to approximate entropy as a feature of sEMG signal classification, because the two parameters bear different numerical values. If generalized through the division by the mean value of the two clusters' centers, the cluster-to-cluster distance of approximate entropy will then be closer to that of fractal dimension, 0.4105 and 0.5061 respectively. When taking the other statistic, cluster dispersion which is more important than cluster distance to measure the efficiency of a feature, into account, we can find that the standard deviations of approximate entropy for the two clusters (0.0330 for FP and 0.0515 for FS) are much smaller than those of fractal dimension (0.2286 for FP and 0.3166 for FS). In consequence, approximate entropy is more appropriate to be a feature of sEMG signal than fractal dimension.

Cluster dispersion and cluster-to-cluster distance

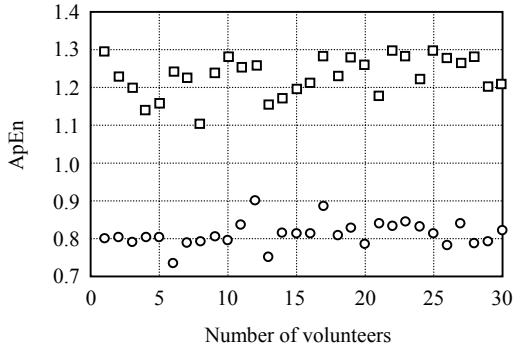

Fig.1 Distribution of approximate entropy calculated by improved ApEn. Squares (□) and circles (○) symbolize approximate entropy values of FS and FP signals respectively

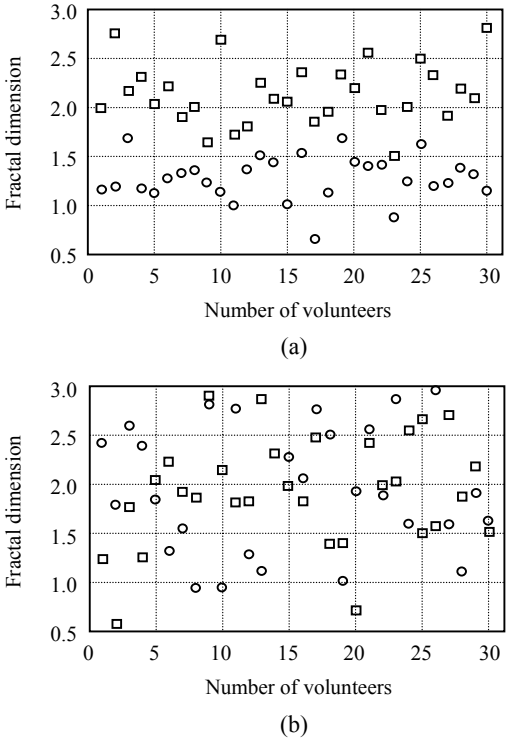

Fig.2 Distribution of fractal dimensions for (a) filtered sEMG signals and (b) raw sEMG signals (Hu *et al*., 2005). Squares (□) and circles (○) symbolize FS and FP signals, respectively

Table 1 Statistics of fractal dimensions for filtered signals and approximate entropy for raw signals

	D_g (FP, FS)	D (FP, FS)	*SD* (FP)	*SD* (FS)
Fractal dimension	0.5061	0.8653	0.2286	0.3166
Improved ApEn	0.4105	0.4189	0.0330	0.0515

D_g: Generalized cluster-to-cluster distance; D: Original cluster-to-cluster distance; *SD*: Standard deviation

were also calculated for approximate entropy parameters obtained by the standard ApEn, and are listed in Table 2 along with those by the improved ApEn. Smaller cluster dispersion values for FP and FS of the improved ApEn algorithm, together with larger cluster-to-cluster distance of it, imply that features by the improved ApEn can identify the two kinds of actions more efficiently than those of the standard ApEn.

Table 2 Statistics of approximate entropy for the standard ApEn and the improved ApEn

	D (FP, FS)	*SD* (FP)	*SD* (FS)
Standard ApEn	0.3291	0.0375	0.0600
Improved ApEn	0.4189	0.0330	0.0515

D: Cluster-to-cluster distance; *SD*: Standard deviation

CONCLUSION

Due to the low pass filter effect of tissues between motor units and the surface electrode on motor unit action potentials (MUAPs), MUAPs with different distances from surface electrode have different amplitudes and frequencies. Rather than simple summation of MUAPs, surface EMG signal is nonlinear combination of all MUAPs generated within the pick-up area of the electrode. The non-linearity underlying surface EMG signals can hardly be described by simple linear superposition models. Successful application of nonlinear dynamic analysis to other biological signals such as EEG and ECG provides a new way to characterize complex EMG signals.

Most nonlinear dynamic measures like K-S entropy and fractal dimensions need very large data volume to describe a strange attractor, which is hard to satisfy when dealing with surface EMG signals. Another problem the most nonlinear parameters frequently come across in dealing with sEMG signals is the noisy background. Unlike most entropy and fractal dimensions, ApEn stops short via small embedding dimension m and coarse tolerance r, sacrificing an attempt to reconstruct the full process dynamics. In consequence, ApEn is applicable to datasets with a relatively small number of data points in addition to its robustness to noise through the selection of tolerance r. Less data points needed together with smaller embedding dimension in the calculation of ApEn also greatly shorten the calculation time, which makes real-time application possible. Much smaller cluster dispersions of approximate entropy than fractal dimension indicate that ApEn is more suitable for characterizing sEMG signals. Besides, smaller cluster dispersions and larger cluster-to-cluster distance of the improved ApEn than the standard ApEn imply that the former can more efficiently extract information from sEMG signals.

In the improved ApEn, similarity between two vectors is based on their shapes rather than their absolute coordinates, and thus two vectors' approximation can be characterized more reasonably, and the measurement of complexity for the time series analyzed is more satisfying. The method presented here can also be applied to other physiological signals with medium-sized data length in noisy background, which are hard to deal with through most linear and nonlinear algorithms.

References

Abel, E.W., Zacharia, P.C., Forster, A., Farrow, T.L., 1996. Neural network analysis of the EMG interference pattern. *Med. Eng. Phys.*, **18**(1):12-17. [doi:10.1016/1350-4533(95)00032-1]

Chang, G.C., Kang, W.J., Luh, J.J., Cheng, C.K., Lai, J.S., Chen, J.J., Kuo, T.S., 1996. Real-time implementation of electromyogram pattern recognition as a control command of man-machine interface. *Med. Eng. Phys.*, **18**(7):529-537. [doi:10.1016/1350-4533(96)00006-9]

Crawford, B., Miller, K., Shenoy, P., Rao, R., 2005. Real-Time Classification of Electromyographic Signals for Robotic Control. Proceedings of the National Conference on Artificial Intelligence. Pittsburgh, Pennsylvania, USA, **2**:523-528.

Doorenbosch, C.A.M., Harlaar, J., 2004. Accuracy of a practicable EMG to force model for knee muscles. *Neuroscience Letters*, **368**(1):78-81. [doi:10.1016/j.neulet.2004.06.055]

Englehart, K., Hudgins, B., Parker, P.A., Stevenson, M., 1999. Classification of the myoelectric signal using time-frequency based representations. *Med. Eng. Phys.*, **21**(6-7):431-438. [doi:10.1016/S1350-4533(99)00066-1]

Georgakis, A., Stergioulas, L.K., Giakas, G., 2003. Fatigue analysis of the surface EMG signal in isometric constant force contractions using the averaged instantaneous frequency. *IEEE Trans. Biomed. Eng.*, **50**(2):262-265. [doi:10.1109/TBME.2002.807641]

Gupta, V., Suryanarayanan, S., Reddy, N.P., 1997. Fractal analysis of surface EMG signals from the biceps. *International Journal of Medical Informatics*, **45**(3):185-192. [doi:10.1016/S1386-5056(97)00029-4]

Hu, X., Wang, Z.Z., Ren, X.M., 2005. Classification of surface EMG signal with fractal dimension. *J. Zhejiang Univ. SCI.*, **6B**(8):844-848. [doi:10.1631/jzus.2005.B0844]

Huang, Y., Englehart, K.B., Hudgins, B., Chan, A.D.C., 2005. A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses. *IEEE Trans. Biomed. Eng.*, **52**(11):1801-1811. [doi:10.1109/TBME.2005.856295]

Jang, G.C., Cheng, C.K., Lai, J.S., Kuo, T.S., 1994. Using Time-Frequency Analysis Technique in the Classification of Surface EMG Signals. Proceeding of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Baltimore, Maryland, **16**:1242-1243.

Kang, W.J., Shiu, J.R., Cheng, C.K., Lai, J.S., Tsao, H.W., Kuo, T.S., 1995. The application of cepstral coefficients and maximum likelihood method in EMG pattern recognition. *IEEE Trans. Biomed. Eng.*, **42**(8):777-785. [doi:10.1109/10.398638]

Kang, W.J., Cheng, C.K., Lai, J.S., Shiu, J.R., Kuo, T.S., 1996. A comparative analysis of various EMG pattern recognition methods. *Med. Eng. Phys.*, **18**(5):390-395. [doi:10.1016/1350-4533(95)00065-8]

Kim, K.H., Kim, H.K., Kim, J.S., Son, W., Lee, S.Y., 2006. A biosignal-based human interface controlling a power-wheelchair for people with motor disabilities. *ETRI Journal*, **28**(1):111-114.

Park, J.H., Stelmach, G.E., 2006. Effect of combined variation of force amplitude and rate of force development on the modulation characteristics of muscle activation during rapid isometric aiming force production. *Exp. Brain Res.*, **168**(3):337-347. [doi:10.1007/s00221-005-0099-6]

Pincus, S.M., 1991. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA*, **88**(6):2297-2301. [doi:10.1073/pnas.88.6.2297]

Pincus, S.M., 1995. Approximate entropy (ApEn) as a complexity measure. *Chaos*, **5**(1):110-117. [doi:10.1063/1.166092]

Pincus, S.M., Goldberger, A.L., 1994. Physiological time-series analysis: what does regularity quantify? *Am. J. Physiol.*, **266**(4 35-4):H1643-H1656.

Sparto, P.J., Parnianpour, M., Barria, E.A., Jagadeesh, J.M., 2000. Wavelet and short-time fourier transform analysis of electromyography for detection of back muscle fatigue. *IEEE Trans. Rehabil. Eng.*, **8**(3):433-436. [doi:10.1109/86.867887]

Wolf, A., Swift, J.B., Swinney, H.L., Vastano, J.A., 1985. Determining Lyapunov exponents from a time-series. *Physica D*, **16**(3):285-317. [doi:10.1016/0167-2789(85)90011-9]

Zardoshti-Kermani, M., Wheeler, B.C., Badie, K., Hashemi, R.M., 1995. EMG feature evaluation for movement control of upper extremity prostheses. *IEEE Trans. Rehabil. Eng.*, **3**(4):324-333. [doi:10.1109/86.481972]

Zijdewind, I., de Groot, M.C.H., Kernell, D., 1998. Task-related variations in motoneuronal drive to a human intrinsic hand muscle. *Neuroscience Letters*, **242**(3):139-142. [doi:10.1016/S0304-3940(98)00049-4]