ELSEVIER

Journal of Neuroscience Methods 82 (1998) 207–214

JOURNAL OF NEUROSCIENCE METHODS

Using factor analysis to identify neuromuscular synergies during treadmill walking

L.A. Merkle [a,*], C.S. Layne [b], J.J. Bloomberg [a], J.J. Zhang [b]

[a] Neuroscience Laboratory, Mail Code NL/37, NASA–Johnson Space Center, Houston, TX 77058, USA

[b] Department of Health and Human Performance, University of Houston, Houston, TX 77204, USA

Received 14 May 1997; received in revised form 26 February 1998; accepted 2 March 1998

Abstract

Neuroscientists are often interested in grouping variables to facilitate understanding of a particular phenomenon. Factor analysis is a powerful statistical technique that groups variables into conceptually meaningful clusters, but remains underutilized by neuroscience researchers presumably due to its complicated concepts and procedures. This paper illustrates an application of factor analysis to identify coordinated patterns of whole-body muscle activation during treadmill walking. Ten male subjects walked on a treadmill (6.4 km/h) for 20 s during which surface electromyographic (EMG) activity was obtained from the left side sternocleidomastoid, neck extensors, erector spinae, and right side biceps femoris, rectus femoris, tibialis anterior, and medial gastrocnemius. Factor analysis revealed 65% of the variance of seven muscles sampled aligned with two orthogonal factors, labeled 'transition control' and 'loading'. These two factors describe coordinated patterns of muscular activity across body segments that would not be evident by evaluating individual muscle patterns. The results show that factor analysis can be effectively used to explore relationships among muscle patterns across all body segments to increase understanding of the complex coordination necessary for smooth and efficient locomotion. We encourage neuroscientists to consider using factor analysis to identify coordinated patterns of neuromuscular activation that would be obscured using more traditional EMG analyses.

Keywords: Coordination; Electromyography; Factor analysis; Neuromuscular synergies

1. Introduction

Factor analysis is a statistical technique used to identify a small number of groups or clusters that represent relationships among a set of interrelated variables. These correlation patterns are expressed in terms of unobservable or latent variables called 'factors'. The goal of factor analysis is to identify the not-so-observable factors from the set of observable variables (Norusis, 1994). These factors can then be interpreted and given meaning based upon the observed variables that load on them.

This useful technique patterns relationships among variables, reduces data, analyzes latent dimensions, generates factor scores, and tests hypotheses (Rummel, 1970; Gorsuch, 1983). It may be used a priori or post hoc. A priori factor analysis (confirmatory) is used for more direct input response analysis of the data. Conversely, post hoc (exploratory) helps gain insight into relationships among variables (Disch, 1989). Exploratory analysis is not always guided by a specific set of hypotheses, but can be guided by open questions about the number and kinds of factors which may be derived from a collection of variables. The present study focuses on exploratory factor analysis.

Factors account for linear relationships that exist among observed variables (Gorsuch, 1983). Factors have common (related) and unique (unrelated) compo-

* Corresponding author. Tel.: +1 281 4833361; fax: +1 281 2445734.

0165-0270/98/$19.00
PII S0165-0270(98)00054-5

nents, and are fewer in number than the total number of observed variables. Mathematically, it is similar to multiple regression. The steps involve calculation of the correlation matrix; extraction of the initial factors, application of mathematical rotations, calculation of factor scores and interpretation of results. Bartlett's test of sphericity, communalities, and Kaiser–Meyer–Olkin (KMO) measures assess the degree of relation among the set of variables and compare the magnitude of the observed correlations to the magnitudes of partial correlations. These gauge the strength of variable interrelationships (Kaiser, 1974; Norusis, 1994). Generally, they are indicators of the appropriate application of factor analysis.

The primary issues in factor analysis are extraction and rotation. The criterion for a rotation is to reach simple structure defined by Thurstone (1935), such that variables should have high loadings on one and only one factor, and zero or near-zero loadings on all other factors. Many factor extraction techniques exist, such as principal component, maximum likelihood, image, and alpha. Among these, principal component is the most popular option, because (1) each factor maximizes the variances explained from the correlation matrix; (2) a factor explains more variance compared to solutions from other approaches; (3) the amount of variance explained by each factor is equal to the corresponding eigenvalue; and (4) factors are uncorrelated with one another. Rotating factors obtains a more interpretable set of factor loadings (Nunnally, 1978; Norusis, 1994). There are many rotation options, too many to go into detail here. Varimax, quartermax and equamax rotations are orthogonal, meaning the mathematical axes are rotated at right angles, and factors are independent of one another. Of these options, varimax is very often selected because it derives uncorrelated factors and simplifies factor interpretation.

Criteria are necessary to determine how many factors to retain. Generally, factors with eigenvalues greater than one should be retained (Rummel, 1970; Gorsuch, 1983; Norusis, 1994). An eigenvalue is defined as the total variance explained by each factor. In principal components, the total variance is equal to the total number of variables (Norusis, 1994). For example, we included seven variables (muscles) in the model so that the total variance is seven. Each observed variable is initially assigned a variance of one. The eigenvalues of the variables redistribute among the factors, so factors with an eigenvalue less than one are no better than a single variable. Moreover, a common factor must account for the variability of at least one variable, so each common factor should have an eigenvalue of one or greater. Therefore, one should retain factors with eigenvalues greater than one. A scree plot (where eigenvalues are graphed in order of magnitude) often confirms this criterion. The plot shows a distinct break between the steep slope of the large factors and the gradual leveling of the rest of the factors (see Stevens (1992) for full explanation).

Factor analysis is quite popular in education and psychology. Physical educators successfully use this technique in analyzing and identifying the components of different sport skills and fitness levels. Psychologists use the technique to develop and test construct validity in their inventories. Yet, factor analysis remains unfamiliar to many scientists due to its complicated concepts and procedures. Particularly, it is rarely used in neuromuscular investigations. For this reason, this study illustrates an application of factor analysis by extracting coordinated patterns of whole-body muscle activation during treadmill walking.

Walking results from whole-body synergy or precisely timed neural activation and deactivation sequences of the musculature throughout the entire body, and a tight synergy between the movement of the head and the trunk (Bernstein, 1967; Layne et al., 1997). However, few studies have examined the relationships between upper- and lower-body activation sequences. Whole-body synergy at the neuromuscular level suggests common neural control signals among the muscles (Bernstein, 1967). These synergies function to reduce the number of controlled degrees of freedom, thereby simplifying the control of locomotion.

Despite hundreds of studies that have described muscle activation patterns during locomotion, only three used factor analysis to identify patterns of multimuscle activation during walking (Wootten et al., 1990; Davis and Vaughan, 1993; Olree and Vaughan, 1995). The findings of these studies vary. Wootten et al. (1990) used the technique to identify principal components of electromyography (EMG) which represented average EMG patterns for cluster analysis. Davis and Vaughan (1993) identified four factors labeled heelstrike, loading response, propulsion, and biphasic patterns. Olree and Vaughan (1995) identified three factors labeled loading, propulsion, and coordination. The authors agreed that the extracted factors were evidence of a few fundamental signals generated by the central nervous system to control the major muscle groups in both legs. These studies suggest that factor analysis is a useful statistical procedure in identifying interlimb neuromuscular activation patterns that contribute to human locomotion. However, only lower-body activation patterns were included in the past studies, providing insufficient information to fully understand the whole-body coordination necessary for walking. Trunk and neck musculature was added in the present investigation because these muscles play a role in head stabilization and dynamic equilibrium during locomotion. Including the neck and trunk builds upon the descriptive work by Winter and associates (Winter and Yack, 1987; Winter et al., 1990, 1993) and the pioneer application of factor analysis of Olree and Vaughan (1995), Davis and Vaughan (1993), and Wootten et al. (1990). The purpose of this study was to illustrate the application of factor analysis in the area of muscular activation patterns, by asking "what are the interrelations among muscle activation patterns that effect a coordinated outcome?".

2. Methods

2.1. Subjects

Ten males (mean age 43.5 ± 5.4 years) volunteered to participate in this study. Two subjects were American astronauts, and the other eight were Russian cosmonauts. All subjects provided written informed consent as required by the Johnson Space Center Human Research Policy and Procedures Committee.

2.2. Procedures

Data were collected as subjects walked for 20 s on a motorized treadmill (Quinton Series 90 Q 55, surface area 51 × 140 cm) at 6.4 km/h while visually fixating on an earth-fixed light-emitting diode (LED) positioned horizontally 30 cm from the eyes. The skin surface was prepared using medical alcohol preparations. Preamplifier surface electrodes (bandpass 30–300 Hz) were placed on the skin over the bellies of the left sternocleidomastoid (LSCM), left erector spinae (LES), right rectus femoris (RRF), right biceps femoris (RBF), right tibialis anterior (RTA), and right medial gastrocnemius (RGA) and secured with hypoallergenic tape. An electrode was also placed on the left side of the neck between C1 and C3 which monitored the neck extension (NE) musculature. The electrodes were covered with elastic leg wraps to prevent motion artifacts resulting from disruption of the electrode–skin surface interface. Raw EMG activity was recorded at 500 Hz, then high pass filtered (30 Hz), full wave rectified, and smoothed using a 15-ms time constant. In addition, temporal events were identified using piezoresistive force transducers (footswitches) attached to the heel and toe of both feet to record heelstrike and toe off. For each 20-s trial, 22–24 individual strides, from right heelstrike to right heelstrike, were extracted from the processed EMG on the basis of voltage records obtained from the footswitches.

2.3. Data reduction and analyses

Mean stride times were calculated by averaging the 22–24 strides within each trial relative to right foot heelstrike. EMG data for all strides were temporally normalized using cubic spline interpolation. EMG was also normalized to 100% of stride by averaging the EMG signal between right heelstrikes. The averaged waveforms were then normalized to the mean level of activation for that waveform, resulting in the mean activation within the waveform being represented as 100%. Files containing 20 data points (epochs) were created from the splined data by averaging the amplitude of each 5% interval (Yang and Winter, 1984). Since the average stride time of the group was 950 ms, each epoch represented approximately 48 ms of data. The muscle activation pattern for a gait cycle was represented by a reduced waveform comprising 20 data points. This promoted meaningful statistical analysis while retaining the basic temporal and spatial features of the activation waveforms. The reduced activation waveforms for each muscle were ensemble averaged over 10 subjects, resulting in an ensemble average for each muscle. The ensemble average of seven muscles was then reduced to a matrix of seven columns (muscle) by 20 rows (5% epochs). The reduced data for seven muscles (LSCM, NE, LES, RBF, RRF, RGA and RTA) were used in the analyses.

Procedures from the Statistics Package for the Social Sciences (SPSS) were used to conduct a factor analysis with principal component extraction and varimax rotation.

Table 1
Intercorrelations of muscle activation patterns

Muscle	NE	LSCM	LES	RBF	RRF	RGA	RTA
NE	1.00	−0.01	0.21	0.42	0.38	0.06	0.34
LSCM		1.00	0.16	0.57	−0.17	−0.38	0.52
LES			1.00	0.39	0.23	−0.13	0.67
RBF				1.00	−0.12	−0.40	0.79
RRF					1.00	−0.18	0.21
RGA						1.00	−0.57
RTA							1.00

LES, left erector spinae; LSCM, left sternocleidomastoid; NE, neck extensors; RBF, right biceps femoris; RGA, right medial gatrocnemius; RRF, right rectus femoris; RTA, right tibialis anterior.

Table 2
Initial statistics from principal components extraction

Factor	Eigenvalue	Amount of variance explained (%)	Cumulative % of variance
1	3.04	43.5	43.5
2	1.49	21.2	64.7
3	0.98	14.0	78.7
4	0.78	11.2	89.9
5	0.47	6.7	96.7
6	0.16	2.3	98.9
7	0.08	1.1	100.0
Total	7.00	100.0	100.0

3. Results

Table 1 shows the intercorrelation matrix of the variables. To ensure proper application of this technique, we inspected several indices, such as the KMO measure, communalities, and Bartlett's test of sphericity. The KMO measure is a measure of sampling adequacy. Generally, this indicator should be 0.5 or greater. The KMO measure was 0.53, indicating that the sample was adequate for a factor analysis; however, the higher this indicator, the better. Mean communality of the variables was 0.65, indicating that each muscle shares a moderate proportion of common variance with other muscles accounted for by the extracted factors. Squared multiple correlations ranged from 0.43 to 0.88. Bartlett's test of sphericity tests the hypothesis that the variance and covariance matrices are identity matrices. An identity matrix is one in which there are ones in the diagonal, and zeros in the off-diagonal. If this were the case, then the variables would be completely unrelated and a factor analysis is inappropriate for this type of data. Bartlett's test of sphericity was 62.244 ($p < 0.05$), indicating that a factor analysis was appropriate.

Two factors were extracted with approximately 65% of the variance explained (Table 2). Since the third factor had an eigenvalue of 0.98 and would possibly contribute an additional 14% of variance, a second analysis was completed, forcing the extraction of three factors. However, examination of this third extracted factor showed that only a single variable loaded on that factor (a specific, not common factor). Since factors must contain a minimum of two items to remain a common factor, the original analysis was retained with two factors. A scree plot was generated, confirming the two-factor solution.

Table 3 shows the rotated factor pattern matrix from varimax rotation. A factor loading represents the correlation between a variable and a factor. A positive loading indicates activation of a muscle. A negative loading indicates lack of neuromuscular activation. Using a standard of a factor loading equal to 0.40, five muscles had salient loadings on one and only one factor, and two were factorially complex (meaning they had salient loadings on more than one factor). To facilitate geometric representation of the factors and variables, we identified the factors as the axes, and plotted the variables according to their loadings on each factor. Figs. 1 and 2 show these representations. Variables that lie close to an axis are more highly related to that factor. Those variables that lie in between the axes are factorially complex.

Factor one was composed of the RTA, RBF, LSCM, RGA and LES. 'Transition control' was selected as the label for this factor, because these five muscles function for the control of weight transition from one lower limb to another. The RRF, RTA, NE and LES loaded on factor two. 'Loading' was the selected label for this factor because the predominant bursts of activity associated with these three muscles occur at the loading portion of the gait cycle (i.e. during weight acceptance and bearing). The RTA and LES had salient loadings on both factors (i.e. factorially complex), suggesting that these muscles play a substantial role in more than one aspect of the gait cycle.

4. Discussion

Factor analysis groups variables into underlying dimensions which can be used to explain phenomena in a

Table 3
Matrix of factor loadings after varimax rotation

Muscle	Factor 1	Factor 2
RTA	**0.868**	**0.411**
RBF	**0.862**	0.182
LSCM	**0.790**	−0.246
RGA	**−0.649**	−0.018
RRF	−0.129	**0.794**
NE	0.128	**0.753**
LES	**0.468**	**0.523**

LES. left erector spinae; LSCM, left sternocleidomastoid; NE, neck extensors; RBF. right biceps femoris; RGA, right medial gatrocnemius: RRF, right rectus femoris; RTA, right tibialis anterior.

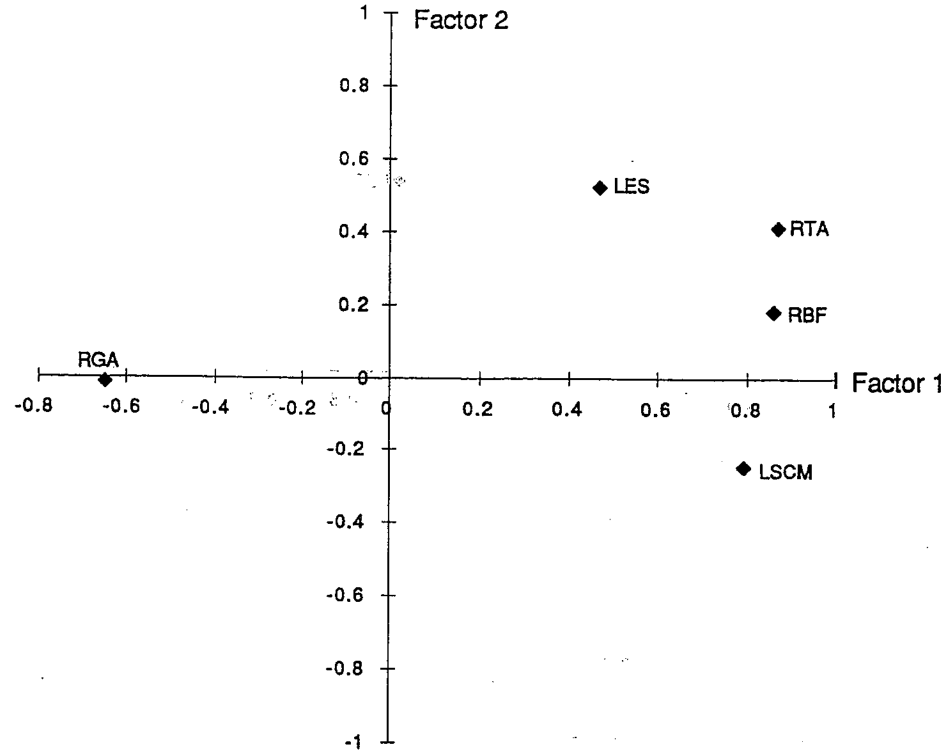

Fig. 1. Geometrical representation of variables on factor 1.

set of data. In the social sciences it is traditionally used for testing construct validity or data reduction. Yet, its application in motor control can be very valuable. In this paper, we have demonstrated how to use factor analysis to identify coordinated activation patterns across muscles during locomotion which otherwise would have remained obscured using more traditional single muscle analytic techniques. The resolved factors have offered a unique look at the neuromuscular activity among the musculature of the shank, leg, lower back and neck that supports coordinated walking. Considering the role that the neck and trunk musculature plays in maintaining head and trunk stability, it was important to also include this musculature in an analysis of locomotion neuromuscular activity patterns. Although walking is a task mainly achieved by the lower body musculature, it is the complex coordination with trunk and neck musculature that allows us to remain upright and walk smoothly and efficiently to our intended destination. Factor analysis, based upon a linear summation of orthogonal factors, is a simplified representation of the complex non-linear nature of the locomotor apparatus.

Since EMG was the source of information and is a reflection of neuromuscular control, the factors reflect how muscles are activated and deactivated in certain patterns to effect a coordinated segmental motion. Using the activation patterns of seven muscles, principal components analysis extracted two orthogonal factors aligned with 65% of the variance representing independent clusters of coordinated patterns of muscle activation during walking. If we agree with the notion of Bernstein (1967) that effective locomotion requires a whole-body synergy, then these two factors can be interpreted as representing two components of the total neuromuscular activation required to achieve this whole-body synergy. The individual factors can be interpreted as unique control signals supporting coordinated segmental motion. Within Bernstein's conception of degrees of freedom, the clustering of muscles onto factors suggests a reduction in degrees of freedom necessary to control coordinated walking.

Each factor was interpreted based upon variable loadings and the biomechanical function of the muscles within each factor. The factor structure was parsimonious, and most variables loaded on only one factor. To ease interpretation of factors, mean waveforms for each muscle are shown in Fig. 3. The RTA, RBF, LSCM, RGA and LES comprised the first factor. The RBF and RTA act in a coordinated fashion for toe clearance. The RBF contracts concentrically for knee flexion while the RTA contracts to dorsiflex the foot for right toe clearance as weight is shifted to the left foot. The high negative loading on the RGA reflects the coordinated inhibition of activity after propulsion necessary for toe clearance. The RGA activity ceases in order for the RTA to effectively dorsiflex the foot to prevent tripping. The LES, during this transition, is active to stabilize the trunk as the body's weight is shifted from the right side to the left, and to prevent

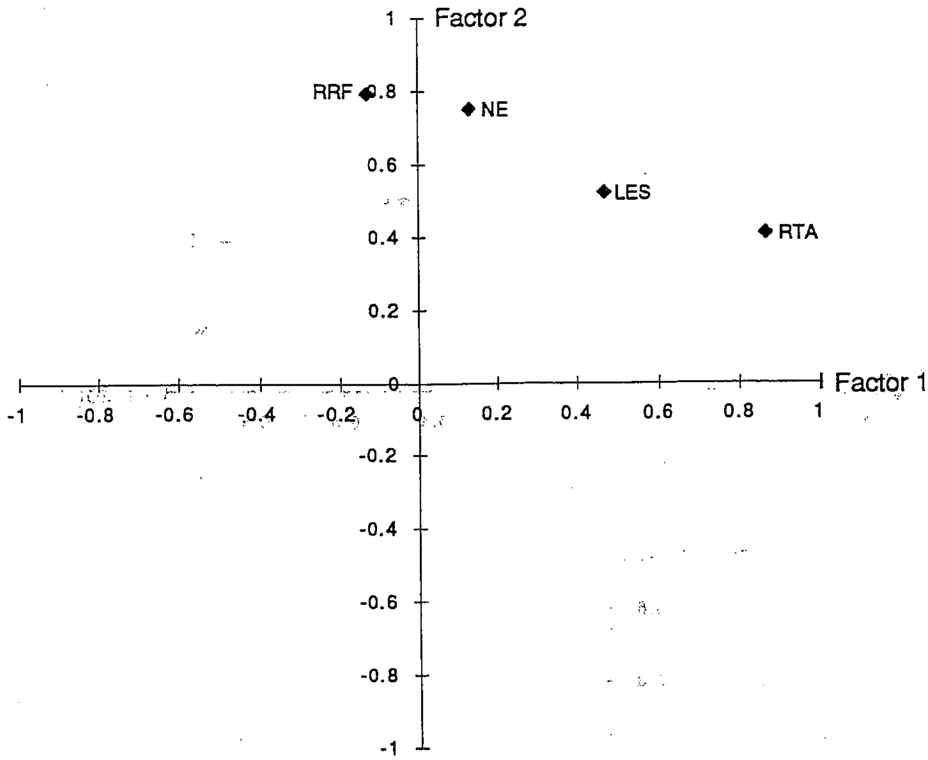

Fig. 2. Geometrical representation of variables on factor 2.

excessive forward trunk rotation. The LSCM also loads on this factor. This muscle is likely responsible for controlling excessive lateral movement of the head as weight shifts toward the support limb (Winter et al., 1990). Collectively, this group of activation patterns, including muscles from the shank, leg, trunk and neck, indicate that complex activation patterns from throughout the body are required to efficiently transfer body weight. Therefore this factor was interpreted as 'transition control'.

Factor one illustrates how factor analysis works to identify relationships which are not obvious from evaluation of individual muscle activation patterns. One would not commonly link LSCM activity with that of the leg and trunk. Yet, the LSCM shares a commonality with the other muscles that load on this factor. The activity associated with the LSCM is control of lateral head movements as weight shifts (Winter et al., 1990). Therefore this muscle must be precisely coordinated with the activity of the RBF, RGA, RTA and LES to stabilize the trunk and head, and help maintain gaze during locomotion during the transition of weight.

For factor two, the peak activity of the RRF occurs between 95 and 5% of the gait cycle (the interval around heelstrike). RRF contracts to decrease forward knee rotation at right heelstrike. The RTA controls lowering the toe to the ground. The LES again functions to control anterior and posterior trunk movement. The NE contracts to prevent excessive pitch motion of the head. These specific functions assist in control of the early stance portion of the stride cycle; hence, this factor was labeled 'loading'.

Although the LES, LSCM and NE are not directly involved in generating translation, their coordination with muscles that are directly involved in translation insures adequate trunk control and maintenance of dynamic balance. This grouping of coordinated activity might normally be overlooked using traditional EMG analysis methods. The salient loadings of the trunk and neck muscles on factors with the lower limb musculature suggest that they are tightly coupled with leg muscle activation and with each transfer of weight to fine-tune the stabilization of the head, arms and trunk, and to attenuate head accelerations (Winter et al., 1990, 1993).

The different types and number of muscles used in a factor analysis will have an effect on the outcome of the analysis. This explains why there are several differences between the factors identified in the present study and past work; different sets of muscles were used. Davis and Vaughan (1993) and Olree and Vaughan (1995) sampled 16 lower limb muscles (eight bilateral pairs) and extracted three to four factors, where we sampled seven muscles (unilateral) and extracted two factors. Our investigation also included

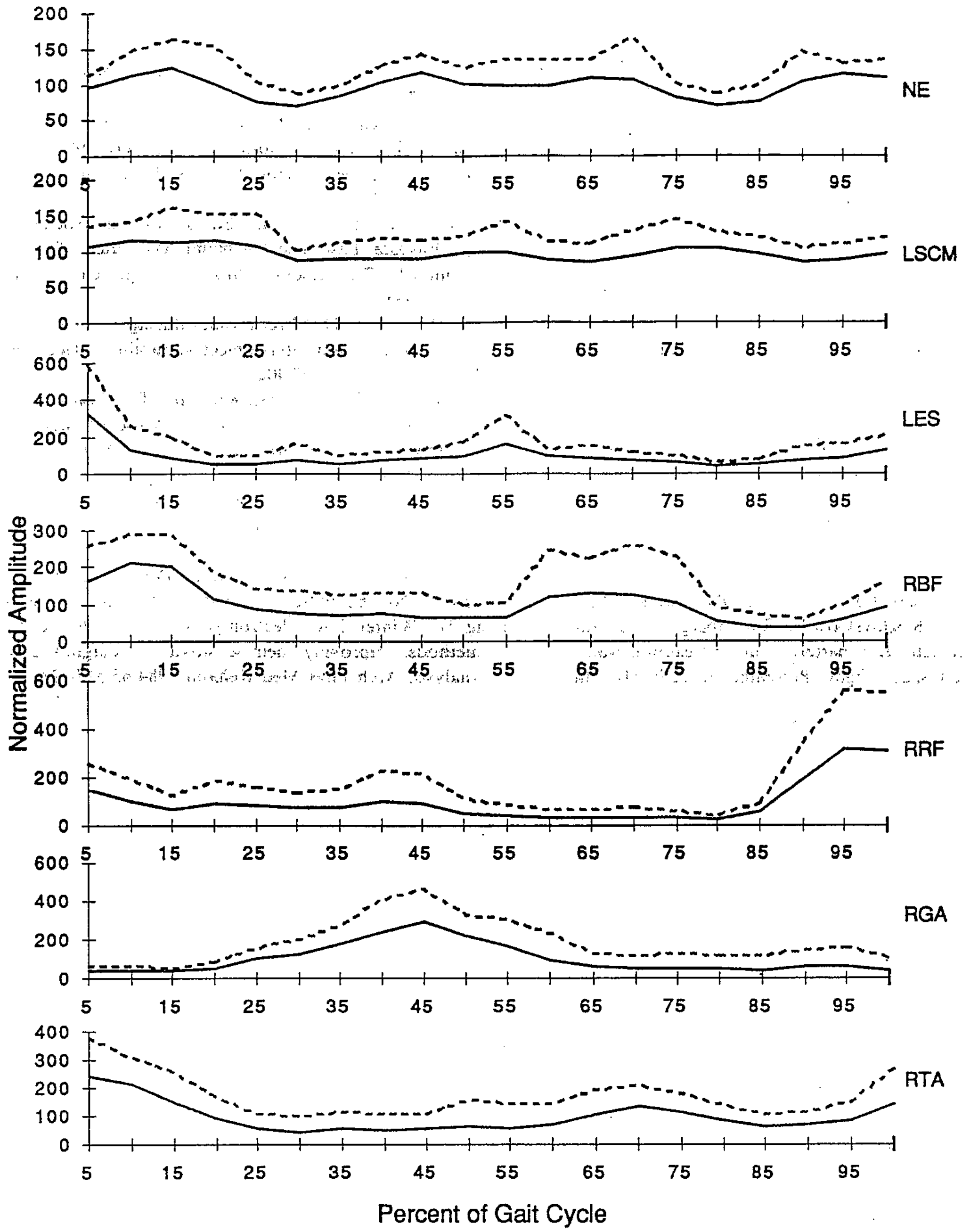

Fig. 3. Mean activation waveforms of the seven muscles sampled.

muscles from the trunk and neck. Differences in where muscles load are also the result of the different number and types of muscles sampled. Since the analysis is built upon the relations among muscles, then the use of more or less muscles in the matrix will effect the results. Predictably, different relationships were found among the set of muscles in this study than those seen previously. It should also be noted that had we analyzed EMG patterns obtained during free locomotion we may have found somewhat different loadings. Remembering that the analysis is built upon intercorrelations, different patterns associated with free versus treadmill locomotion would slightly alter the relationships among the muscles and therefore affect the loadings.

Since forces generated at one joint are continually being transferred to other joints and additional energy is being absorbed at each heelstrike, it is important to consider how other muscles throughout the body are coordinated with the lower limb musculature to ensure appropriate segmental motion. The present results demonstrate that factor analysis can be used to explore relations among muscle patterns across all body segments to more fully understand the complex coordination necessary for smooth and efficient locomotion.

Acknowledgements

This research was funded by NASA Grant 106-31-42-42. The authors thank Brian T. Peters, Shannon L. Smith and Dr. P. Vernon McDonald for their efforts during data collection.

References

Bernstein NA. The Coordination and Regulation of Movements. New York: Pergamon, 1967.

Davis BL, Vaughan CL. Phasic behavior of EMG signals during gait: use of multivariate statistics. J Electromyogr Kinesiol 1993;3:51–60.

Disch J. Selected Multivariate Statistical Techniques. Measurement Concepts in Physical Education and Exercise Science. Champaign, IL: Human Kinetics, 1989:150–180.

Gorsuch RL. Factor Analysis. 2nd ed. Hillsdale: Lawrence Erlbaum Associates, 1983.

Kaiser HF. Analysis of factorial simplicity. Psychometrika 1974;39:31–6.

Layne CS, Lange GW, Pruett CJ, McDonald PV, Merkle LA, Mulavara AP, Smith SL, Kozlovskaya IB, Bloomberg JJ. Adaptation of neuromuscular activation patterns during treadmill walking after long-duration space flight. Presented at 12th Man in Space Symp., Washington DC, 1997.

Norusis MJ. Statistics Package for the Social Sciences (SPSS) Professional Statistics 6.1. Chicago: SPSS Inc., 1994.

Nunnally JC. Psychometric Theory. 2nd ed. New York: McGraw–Hill, 1978.

Olree KS, Vaughan CL. Fundamental patterns of bilateral muscle activity in human locomotion. Biol Cybern 1995;73:409–14.

Rummel RJ. Applied Factor Analysis. Evanston, IL: Northwestern University Press, 1970.

Stevens J. Applied Multivariate Statistics for the Social Sciences. 2nd ed. Hillsdale: Lawrence Earlbaum Associates, 1992.

Thurstone LL. The Vectors of Mind. Chicago: University of Chicago Press, 1935.

Winter DA, Yack HJ. EMG profiles during normal human walking: stride-to-stride and inter-subject variability. Electroencephal Clin Neurophysiol 1987;67:402–11.

Winter DA, Ruder GK, MacKinnon CD. Control of Balance of Upper Body During Gait. Multiple Muscle Systems: Biomechanics and Movement Organization. New York: Springer-Verlag, 1990:534–541.

Winter DA, MacKinnon CD, Ruder GK, Weiman C. An integrated EMG/biomechanical model of upper body balance and posture. Progr Brain Res 1993;97:359–67.

Wootten ME, Kadaba MP, Cochran GVB. Dynamic electromyography. II. Normal patterns during gait. J Ortho Res 1990;8:259–65.

Yang JF, Winter DA. Electromyographic amplitude normalization methods: improving their sensitivity as diagnostic tools in gait analysis. Arch Phys Med Rehabil 1984;65:517–21.