Unsupervised learning of spatio-temporal primitives of emotional gait

Lars Omlor, Martin A. Giese
Laboratory for Action Representation and Learning/Department of Cognitive Neurology,
Hertie Institute for Clinical Brain Research, University of Tübingen, Germany

Abstract

Experimental and computational studies suggest that complex motor behavior is based on simpler spatio-temporal primitives. This has been demonstrated by application of dimensionality reduction techniques to signals from electrophysiological and EMG recordings during execution of limb movements. However, the existence of such primitives on the level of the trajectories of complex human full-body movements remains less explored. Known blind source separation techniques, like PCA and ICA, tend to extract large numbers of components from such trajectories, which are difficult to interpret. For the analysis of emotional human gait patterns, we present a new non-linear source separation technique, realizing a more appropriate modeling of temporal delays. The proposed method allows the accurate modeling of high-dimensional movement trajectories with very few source components, and is significantly more accurate than other techniques. Combining this method with sparse regression, we identified primitives for the encoding of individual emotions in gait that match features that are important for the perception of emotional body expressions in psychological studies. This suggests the existence of emotion-specific motor primitives in human gait.

For the analysis of electrophysiological and EMG data known blind source separation techniques like PCA or ICA have been successfully applied for the extraction of basic motor components (e.g.[1,2]). These studies support the hypothesis of the existence of a limited set of movement primitives, which forms the basis for the realization of more complex motor behaviors.
In our study we tried to exploit similar unsupervised learning techniques for the identification of movement primitives that are relevant for the expression of emotions in gait. Psychophysical studies suggest that the perception of emotions in gait might be based on specific emotion-specific dynamic features. We demonstrate that such features can be learned immediately from kinematic data.
Common source separation techniques typically require relatively large numbers of sources in order to appropriately approximate complex trajectories. The number of sources can be drastically reduced without a loss of approximation quality if temporal delays between the trajectories of different degrees of freedom are taken into account. We have developed a new blind source separation algorithm which allows a suitable modeling of such time delays. Related existing techniques require non positive sources, necessitate additional sparseness assumptions[3], or do not allow a dimension reduction[5].

Trajectory data: Movement trajectories were recorded from four lay actors executing walking with four basic emotional expressions (happy, angry, sad and fear) and neutral walking using a VICON motion capture system. Using a kinematic body model with 17 joints, we computed joint angles from the marker positions. As data for the unsupervised learning procedure we used only the flexion angles of the hip, knee, elbow, shoulder and the clavicle, since these showed the most reproducible variation.

Blind source separation: We have compared different methods for blind source separation for our data set: PCA, fast ICA and bayesian ICA with a positivity constraint for the elements of the mixing matrix[4]. These method required at least 5 sources for an accurate reconstruction (unexplained variance smaller than 90 %) of the original trajectories.
We also analyzed the data on a joint-by-joint basis, performing separate ICA's for the individual joints. Computing the autocorrelation functions between the sources, we found that the sources extracted from separate joints were extremely similar in terms of their shapes, but differed from each other by time delays. This motivated the development a new algorithm that allows for an appropriate modeling of this inherent structure of the data.

Signifying by x_i the i-th component of the approximated trajectory and by s_j the j-th unknown source signal, the data is modeled by the following *nonlinear* generative model:

$$x_i(t) = \sum_{j=1}^{n} \alpha_{ij} s_j(t - \tau_{ij}) \tag{1}$$

The model is specified by the linear mixing coefficients α_{ij} and the time τ_{ij} delays between source signals and trajectory components. The problem of blind source separation with time delays has only been rarely been treated in the literature (e.g.[3,5,6]).
An efficient algorithm for the solution of this problem that scales up to high-dimensional problems was obtained by transforming the signals in time-frequency domain using the Wigner-Ville transform, that is defined by

$$Wf(x,\omega) := \int \mathrm{E}\left\{ f(x + \frac{t}{2}) \overline{f(x - \frac{t}{2})} \right\} e^{-2\pi \mathrm{i} \omega t} dt$$

Applying this integral transformation to equation (1) one obtains:

$$\begin{aligned} Wx_i(\eta,\omega) &= \int \mathrm{E}\Big\{ \sum_{j,k=1}^{n} \alpha_{ij} \overline{\alpha_{ik}} s_j(\eta + \frac{t}{2} - \tau_{ij}) \overline{s_k}(\eta - \frac{t}{2} - \tau_{ik}) \Big\} e^{-2\pi \mathrm{i} \omega t} dt \\ &= \sum_{j}^{n} |\alpha|_{ij}^2 W s_j(\eta - \tau_{ij}, \omega) \end{aligned} \tag{2}$$

The last equality sign above is due to the (approximate) independence of the sources. With the additional assumption that the data coincides with the mean of its distribution ($x_j \approx E(x_j)$) one can compute the first and the zeros order moment from equation (2) resulting in the two equations:

$$|\mathcal{F}x_i|^2(\omega) = \sum_{j}^{n} |\alpha|_{ij}^2 |\mathcal{F}s_j|^2(\omega) \tag{3}$$

$$|\mathcal{F}x(\omega)|^2 \cdot \frac{\partial}{\partial \omega} \arg\{\mathcal{F}x\} = \sum_{j}^{n} |\alpha|_{ij}^2 \cdot |\mathcal{F}s_i|^2 \cdot \left[\frac{\partial}{\partial \omega} \arg\{\mathcal{F}s_j\} + \tau_{ij} \right] \tag{4}$$

Here $\mathcal{F}$ denotes the Fourier transform. From these equations the unknowns can be estimated. To recover the unknown sources s_j, mixing coefficients α_{ij} and time delays τ_{ij} we used the following two step algorithm:

1. First, equation (3) is solved using non-negative ICA[7]. (This step could also be realized exploiting non-negative matrix factorization.)

2. Iteration of the following two steps:

   (a) Equation (4) is solved numerically for $\frac{\partial}{\partial \omega} \arg\{\mathcal{F}s_j\}$, and by integration $\mathcal{F}s_j$ is obtained with initialization $\tau_{ij} = 0$.

   (b) The mixing matrix and the delays are obtained by solving the following optimization problem (with $\mathbf{S}(\vec{\tau_j}) = (s_k(t_i - \tau_{jk}))_{i,k},\ \mathbf{A} = (\alpha_{ij})_{ij}$):

   $$[\widehat{\vec{\tau_j}}, \widehat{\mathbf{A}}] = \underset{[\vec{\tau_j}, \mathbf{A}]}{\operatorname{argmin}} \| x_j - \mathbf{A} \cdot \mathbf{S}(\vec{\tau_j}) \|$$

   This minimization is accomplished following[8], assuming uncorrelatedness for the sources and independence of the time delays.

To construct a mapping between the linear weights $\mathbf{A}$ and the emotional expression we considered the following multi-linear regression model

$$\mathbf{a}_j \approx \mathbf{a}_0 + \mathbf{C} \cdot \mathbf{e}_j$$

where $\mathbf{a}_0$ is a vector with the weights for neutral walking, and $\mathbf{a}_j$ the weight vector for emotion j. $\mathbf{e}_j$ is the j-th unit vector. The columns of the matrix $\mathbf{C}$ encode the deviations in weight space between emotion j and neutral walking. To obtain sparsified solutions for this matrix, we solved the regression problem by minimizing the following cost function (with $\gamma > 0$):

$$E(\mathbf{C}) = \sum_{j} \| \mathbf{a}_j - \mathbf{a}_0 + \mathbf{C} \cdot \mathbf{e}_j \|^2 + \gamma \sum_{ij} |C_{ij}|$$

Results and Discussion We have compared several blind source separation methods including PCA, fast ICA and our new method. In addition, we tested two methods with a positivity constraint for the elements of the mixing matrix. The first was a probabilistic ICA[7] and the second our algorithm with the additional constraint $\alpha_{ij} \geq 0$.
The results for this comparison are shown in Figure 1 where the approximation accuracy is plotted against the corresponding number of sources. The comparison between the tested algorithms reveals that methods (PCA and ICA) with purely linear superposition of the source signals, without specific treatment of time delays, result in approximations with limited accuracy, explaining about 90% of the variance of the data with 5 sources.

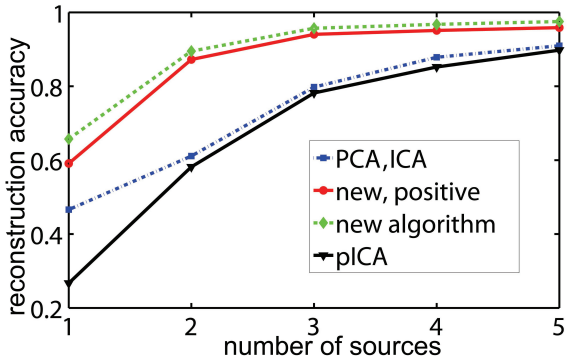

Figure 1: Comparison of different blind source separation algorithms. Explained variance is shown for different numbers of extracted sources.

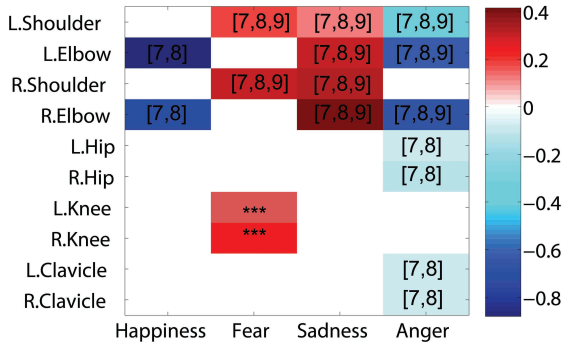

Figure 2: Elements of the weight matrix **C**, encoding emotion-specific deviations from neutral walking, for different degrees of freedom. Numbers indicate references describing psychophysical experiments that have reported the same critical components for visual emotion recognition.

The proposed new algorithm reaches the same level of accuracy with only two sources. Superpositions with more than two sources approximate the data almost perfectly, explaining more than 97% of the variance. The inclusion of a positivity constraint for the weights in the new algorithm did not change the results very much. For an additional verification of our results, we used the approximated trajectories for the animation of an avatar with 13 segments and 10 joints. Animations with trajectories based on 3 sources with the proposed method look very natural.

To test whether our algorithm extracts components that are biologically meaningful, we compared the elements of the regression matrix **C** with results from psychophysical experiments on the perception of emotional gaits. These experiments show that the perception depends on specific changes of individual degrees of freedom relative to the pattern of neutral walking. We found excellent consistency between the features extracted by our learning algorithm and features reported in these behavioral studies, e.g. increased step length for angry walking, or decreased movements for sad walking. The numbers in Fig. 2 indicate the features and references of behavioral recognition studies that reported consistent features. The only one feature that has not been reported in these psychological experiments was an decreased flexion of the knee angles for angry walking ($***$ in Figure 2).

We conclude that the proposed new method accomplishes more accurate approximations of emotional gait trajectories with fewer sources than other common blind source separation techniques. In addition, we have shown that the learned sparsified model for emotional gaits extracts features that match components that are important for the visual perception of emotional walks. Our results provide evidence for the existence of emotion-specific movement primitives, and suggest that spatio-temporal features that are critical for the visual perception of emotional body expressions match highly informative components of the relevant motor patterns.

Bibliography

1. Ivanenko YP, Poppele RE, L. F. *J Physiol.* **556(Pt 1)**, 267–82 (2004 Apr 1).
2. d'Avella A, B. E. *Proc Natl Acad Sci U S A* **102(8)**, 3076–81 (2005 Feb 22).
3. Bofill, P. *Neurocomputing* **Vol. 55**, 627–641 (2003.).
4. A. Cichocki, S. A. *John Wiley, Chichester* (April 2002.).
5. Yeredor, A. *Acoustics, Speech, and Signal Processing* **5**, 237–40 (April 2003).
6. Torkkola, K. *ICASSP'96* , 3509–3512 (1996).
7. Højen-Sørensen, P., Winther, O., and Hansen, L. *Neural Computation* **14**, 889–918 (2002).
8. Swindelhurst, A. *IEEE Trans. on Sig. Proc.* **ASSP-33,no. 6**, 1461–1470 (February, 1998.).
9. Montepare, J. M., S. B. G. A. *J. Nonverb. Behav.* **Vol 11 (1)**, 33–42 (March 1987.).
10. de Meijer, M. *Journal of nonverbal behaviour* **Vol 13 (4)**, 247–268 (December 1989.).
11. Wallbott, H. G. *European Journal of Social Psychology,* **Vol 28**, 879–896 (1998.).

Acknowledgements We thank Claire Roether for her help with the data acquisition and the psychological interpretation. Supported by HFSP, Volkswagenstiftung, DFG. Additional support by MPI for Biological Cybernetics Tübingen.