

Extraction of spatio-temporal primitives of emotional body expressions

Lars Omlor, and Martin A. Giese

Laboratory for Action Representation and Learning, Department of Cognitive Neurology,
Hertie Institute for Clinical Brain Research, University of Tübingen, Germany
WWW home page: <http://www.uni-tuebingen.de/uni/knv/ar1/index.html>

Abstract

Experimental and computational studies suggest that complex motor behavior is based on simpler spatio-temporal primitives, or synergies. This has been demonstrated by application of dimensionality reduction techniques to signals obtained by electrophysiological and EMG recordings during the execution of limb movements. However, the existence of spatio-temporal primitives on the level of the joint angle trajectories of complex full-body movements remains less explored. Known blind source separation techniques, like PCA and ICA, tend to extract relatively large numbers of sources from such trajectories that are typically difficult to interpret. For the example of emotional human gait patterns, we present a new non-linear source separation technique that treats temporal delays of signals in an efficient manner. The method allows to approximate high-dimensional movement trajectories very accurately based on a small number of learned spatio-temporal primitives or source signals. It is demonstrated that the new method is significantly more accurate than other common techniques. Combining this method with sparse multivariate regression, we identified spatio-temporal primitives that are specific for different emotions in gait. The extracted emotion-specific features match closely features that have been shown to be critical for the perception of emotions from gait pattern in visual psychophysics studies. This suggests the existence of emotion-specific motor primitives in human gait.

Key words: blind source separation, ICA, delayed mixing, emotions, movement primitives, kinematics, gait analysis

Human full-body movements are characterized by a large number of degrees of freedom. This makes the accurate synthesis of human trajectories for applications in computer graphics and robotics a challenging problem. The analysis of motor behavior suggests the existence of simple basis components, or spatio-temporal primitives, that form building blocks for the realization of more complex motor behavior [1; 2]. Since such basic components cannot be directly observed, several studies have aimed at identifying spatio-temporal primitives by application of unsupervised learning techniques, like PCA or ICA [3; 4; 5], to data from electrophysiological and EMG recordings (e.g.[6; 7]). The same methods can be applied directly to

joint angle trajectories. However, this analysis of complex full-body movements typically results in the extraction of a relatively large number of basic components or source signals that are difficult to control and to interpret (e.g.[10]). In our study we tried to learn movement primitives of emotional gaits from joint-angle trajectories. We present a new technique for blind source separation, which is based on a nonlinear generative model that, opposed to normal PCA and ICA, can model time delays between source components and individual joint angles. Opposed to other existing algorithms for blind source separation with delays [11; 12], our method scales up to large problems, it allows dimensionality reduction, and it requires no addi-

tional sparseness assumptions. It provides a much better approximation of gait data with few basic components than other common unsupervised learning methods.

By approximating the trajectories of emotional gaits by superpositions of the extracted component signals and applying a sparsifying regression algorithm to learn a model for the mixing matrix, we extracted emotion-specific spatio-temporal features from the trajectory data. A comparison with psychophysical studies on the perception of emotional gaits reveals that the emotion-specific components derived from our kinematic analysis match features that have been described as fundamental for the visual recognition of emotions from gait. This indicates that the novel algorithm is suitable for the extraction of biologically valid movement components.

1 Trajectory data

Using a VICON motion capture system with 7 cameras, we recorded the gait trajectories from thirteen lay actors executing walking with four basic emotional styles (happy, angry, sad and fear), and normal non-emotional walking. Each trajectory was executed three times by each actor, resulting in a data set with 195 gait trajectories. Approximating the marker trajectories with a hierarchical kinematic body model (skeleton) with 17 joints, we computed joint angle trajectories. Rotations between adjacent segments were described by Euler angles, defining flexion, abduction and rotations about the connecting joint. Data for the unsupervised learning procedure included only the flexion angles of the hip, knee, elbow, shoulder and the clavicle, since these showed the most reproducible variation.

2 Blind source separation

To establish a benchmark, we first applied three established methods for the estimation of source signals to our trajectory data: PCA, fast ICA and Bayesian ICA [13] with a positivity constraint for the elements of the mixing matrix. These methods required at least 5 sources for reconstructions of the original trajectories, in order to explain at least 90 % of the variation of the data. We then performed separate ICAs for the individual joints, resulting in separate sets of source variables for each individual joint. By computing the cross-correlation functions between different sources, we

found that sources derived from different joints were often astonishingly similar and differed only by additional time delays. This finding motivated us to develop a new source separation algorithm that takes this property of the data into account by explicit modeling of these delays.

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

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

The matrix $\mathbf{A} = (\alpha_{ij})_{ij}$ is called the mixing matrix. The nonlinearity becomes obvious in the frequency domain

$$\mathcal{F}x_i(\omega) = \sum_{j=1}^n \alpha_{ij} e^{-2\pi i \tau_{ij} \omega} \mathcal{F}s_j(\omega) = \mathbf{A}(\omega) \cdot \hat{\mathbf{S}}(\omega) \quad (2)$$

where the matrix $\mathbf{A}(\omega)$ is dependent of the frequency variable, and where the vector $\hat{\mathbf{S}}(\omega)$ signifies the Fourier transform of the source signals and \mathcal{F} denotes the Fourier transform.

The model is specified by the linear mixing coefficients α_{ij} and the time delays τ_{ij} between source signals and trajectory components. The problem of blind source separation with time delays has been treated only rarely in the literature (e.g. [11; 12; 14]). The existing algorithms were not applicable to our problem because they either require positive signals, or were not suitable for dimensionality reduction (assuming more sources than signals).

An efficient algorithm for the solution of this blind source separation problem, which scales up to higher-dimensional problems, was obtained by representing the signals in time-frequency domain using the Wigner-Ville transform [9; 8], that is defined by

$$Wf(x, \omega) := \int \mathbb{E} \left\{ f\left(x + \frac{t}{2}\right) \overline{f\left(x - \frac{t}{2}\right)} \right\} e^{-2\pi i \omega t} dt \quad (3)$$

where \mathbb{E} denotes the expected value. Applying this integral transformation to equation (1) one ob-

tains:

$$\begin{aligned}
Wx_i(\eta, \omega) &= \int \mathbb{E} \left\{ \sum_{j=1}^n \sum_{k=1}^n \alpha_{ij} \overline{\alpha_{ik}} s_j \left(\eta + \frac{t}{2} - \tau_{ij} \right) \right. \\
&\quad \left. \times \overline{s_k} \left(\eta - \frac{t}{2} - \tau_{ik} \right) \right\} e^{-2\pi i \omega t} dt \\
&\approx \sum_j^n |\alpha|_{ij}^2 Ws_j(\eta - \tau_{ij}, \omega).
\end{aligned} \tag{4}$$

The last term is derived exploiting 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 moment of equation (4) in η , defined as:

$$\begin{aligned}
\int \eta \cdot Wx_i(\eta, \omega) d\eta &= |\mathcal{F}x_i(\omega)|^2 \cdot \frac{\partial}{\partial \omega} \arg\{\mathcal{F}x_i\} \\
&= \sum_j^n |\alpha|_{ij}^2 \int \eta \cdot Ws_j(\eta - \tau_{ij}, \omega) d\eta \\
&= \sum_j^n |\alpha|_{ij}^2 \cdot |\mathcal{F}s_i|^2 \cdot \left[\frac{\partial}{\partial \omega} \arg\{\mathcal{F}s_j\} + \tau_{ij} \right]
\end{aligned}$$

Analogously, the zero-order moment can be computed, yielding the two equations:

$$|\mathcal{F}x_i|^2(\omega) = \sum_j^n |\alpha|_{ij}^2 |\mathcal{F}s_j|^2(\omega) \tag{5}$$

$$\begin{aligned}
|\mathcal{F}x_i(\omega)|^2 \cdot \frac{\partial}{\partial \omega} \arg\{\mathcal{F}x_i\} &= \\
\sum_j^n |\alpha|_{ij}^2 \cdot |\mathcal{F}s_i|^2 \cdot \left[\frac{\partial}{\partial \omega} \arg\{\mathcal{F}s_j\} + \tau_{ij} \right] &\tag{6}
\end{aligned}$$

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:

- I) First, equation (5) is solved using non-negative ICA [13]. Resulting in estimates for $|\alpha_{ij}|^2$ and $|\mathcal{F}s_j|^2$. (This step could also be realized exploiting non-negative matrix factorization.)
- II) Iteration of the following two steps:
 - (a) Given the estimates obtained in step I the only remaining unknown in equation (6) is $\left(\frac{\partial}{\partial \omega} \arg\{\mathcal{F}s_j\} + \tau_{ij} \right)$.

Thus given τ_{ij} one can compute $\arg\{\mathcal{F}s_j\}$.

The τ_{ij} are initialised as $\tau_{ij} = 0$ and are optimized iteratively in step IIb.

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

$$[\widehat{\vec{\tau}}_j, \widehat{\mathbf{A}}] = \operatorname{argmin}_{[\vec{\tau}_j, \mathbf{A}]} \|x_j - \mathbf{A} \cdot \mathbf{S}(\vec{\tau}_j)\| \tag{7}$$

This minimization is accomplished following [15], assuming uncorrelatedness of the sources and independence of the time delays.

Step (2) is repeated till the delays become stationary (usually after about ten iterations).

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

$$\mathbf{a}_j \approx \mathbf{a}_0 + \mathbf{C} \cdot \mathbf{e}_j \tag{8}$$

where \mathbf{a}_0 is a vector containing all 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 $\gamma > 0$) with quadratic programming, which specifies an additional L_1 regularization term for the matrix \mathbf{C} :

$$E(\mathbf{C}) = \sum_j \|\mathbf{a}_j - \mathbf{a}_0 - \mathbf{C} \cdot \mathbf{e}_j\|^2 + \gamma \sum_{ij} |C_{ij}| \tag{9}$$

3 Results

Figure 1 presents the approximation accuracy (explained variance of the whole data set) as a function of the number of extracted sources for 5 different blind source separation methods: PCA, fast ICA, probabilistic ICA with a positivity constraint for the elements of the mixing matrix, and our new method with and without a positivity constraint for the weights α_{ij} . The new algorithm reaches an accuracy of 97% with only three source signals (even in presence of a positivity constraint), while the other methods require

at least six sources to achieve the same level of accuracy. Therefore it is sufficient to extract only three sources to describe the data almost perfectly (95% of the variance). Physiologically these sources might correspond to "synergies", in the sense of motor control [1]. To remove ambiguities between the linear weights and the delays (since for example: $\sin(x + \pi) = -\sin(x)$) we also restricted the mixing matrix to be positive. This leads only to a slight drop in accuracy but improves the interpretability of the results. In addition the positivity of the weights makes them interpretable as positive neural signals.

For additional validation we animated an artificial body model (avatar) with the approximated joint angle trajectories using three estimated sources. Animations based on approximations using the source variables derived by the described new algorithm are almost undistinguishable from animations with the original motion capture data. However, animations using trajectory approximation with three sources derived by PCA or normal ICA show strong artifacts. This provides an additional validation of the new method and demonstrates its potential for applications in computer graphics. Movies from animations with the different methods can be downloaded from the Web page: <http://www.uni-tuebingen.de/uni/knv/ar1/ar1-research.html>

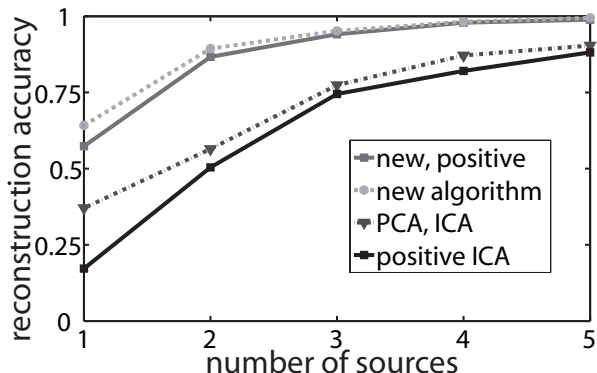


Fig. 1. Comparison of different blind source separation algorithms. Explained variance is shown for different numbers of extracted sources.

Figure 2 shows the estimated delays corresponding to the first source (τ_{i1}) for all joint angles, repetitions and actors. The estimated time delays are quite reproducible and show a characteristic profile over the different joints. The variation of the delays across emotions and actors is relatively small. This result reflects the high degree of temporal coordination of walking, which is independent from the specific emotional style. This shows

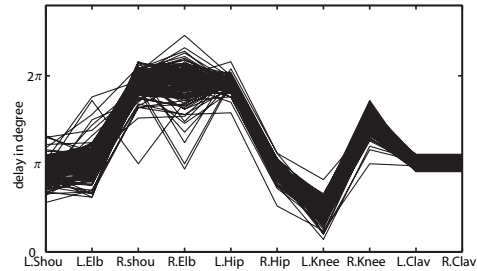


Fig. 2. Estimated Time Delays for the first source for the ten different joints.

that the estimated delays are physiologically at least not implausible.

Another way to validate the biological plausibility of the extracted spatio-temporal components is to compare the non-zero elements of the sparsified regression matrix \mathbf{C} with results from psychophysical experiments on the perception of emotional gaits. These experiments show that perception of emotions depends on specific changes of the joint angle amplitudes of different joints relative to the pattern of neutral walking. For example, angry walking is characterized by an increase of many joint angle amplitudes. Figure 3 illustrates

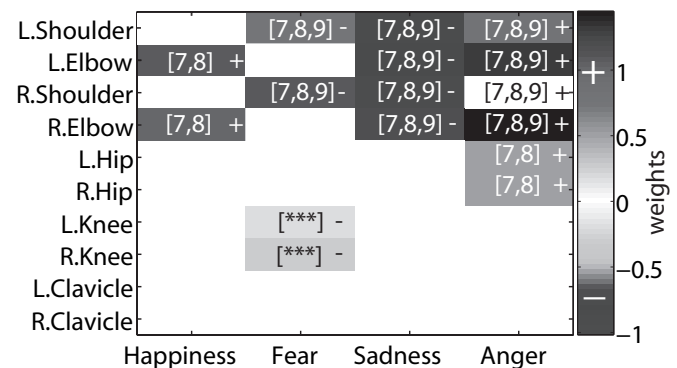


Fig. 3. Elements of the weight matrix \mathbf{C} , encoding emotion-specific deviations from neutral walking, for different degrees of freedom. Numbers indicate references describing psychophysical experiments that have reported the same critical components for visual emotion recognition.

the nonzero elements of the regression matrix \mathbf{C} as gray level plots. The signs indicate if the corresponding emotion-specific feature is related to an increase or a decrease of the amplitudes of the corresponding joints. For example, angry walking is characterized by increases of the amplitudes of many joints, while sad walking is characterized

by decreased arm movements. Interestingly, these emotion-specific kinematic features match closely dynamic features that have been described in psychophysical studies on the perception of emotional gaits, which have extracted salient features from perceptual ratings. The numbers in Figure 3 indicate psychophysical references that have reported the same feature. The only feature that has not been described in these studies is a decrease of the knee angle amplitude for fearful walking, compared to neutral walking [***]. Interestingly, we have consistently found this feature in our own psychophysical experiments on the perception of emotional gaits. These results provide evidence that the visual perception of emotions from body movements might exploit salient kinematic features that are associated with classes of emotional movements.

4 Conclusions

The proposed new algorithm accomplishes highly accurate approximation of gait trajectories with very few extracted source components. Selective modulation of the extracted primitives allows to simulate different emotional styles, and the required modulation reflects specific changes in selected joints that are consistent with features that are important for the visual perception of emotional gaits. This supports the biological relevance of the extracted sources and emotion-specific kinematic components.

Future work will try to extend this method for non-periodic movements. The high accuracy of the method also motivates an application in the context of character animation, making the method potentially suitable for learning-based movement synthesis achieving high degrees of realism.

Acknowledgements: This work was supported by HFSP, DFG and the Volkswagenstiftung. We thank C.L. Roether for help with the trajectory acquisition and the psychological interpretation of the data, and W. Ilg for support with the motion capturing.

References

[1] Flash, T., Hochner, B.: Motor primitives in vertebrates and invertebrates. *Curr Opin Neurobiol.* **15(6)** (2005) 660–666
 [2] Schaal, S., Peters, J., Nakanishi, J., Ijspeert, A.: Learning movement primitives. Inter-

national Symposium on Robotics Research (ISRR2003) (2003)
 [3] A. Hyvärinen, E.O.: A fast fixed-point algorithm for independent component analysis. *Neural Computation* **9** (1997) 1483–1492
 [4] Cichocki, A., Amari, S.: Adaptive blind signal and image processing. John Wiley, Chichester (2002.)
 [5] Bell, A.J., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. *Neural Computation* **7** (1995) 1129–1159
 [6] Ivanenko, Y., Poppele, R., Lacquaniti, F.: Five basic muscle activation patterns account for muscle activity during human locomotion. *J Physiol.* **556(Pt 1)** (2004) 267–282
 [7] d’Avella, A., Bizzi, E.: Shared and specific muscle synergies in natural motor behaviors. *Proc Natl Acad Sci U S A* **102(8)** (2005) 3076–3081
 [8] Amin, M., Zhang, Y.: Direction finding based on spatial time-frequency distribution matrices. *Digital Signal Processing* **10** (2000) 325–339
 [9] Matz, G., Hlawatsch, F.: Wigner distributions (nearly) everywhere: Time-frequency analysis of signals, systems, random processes, signal spaces, and frames. *Signal Processing* **83** (2003) 1355–1378
 [10] Safonova, A., Hodgins, J., Pollard, N.: Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. *ACM Trans. Graph.* **23(3)** (2004) 514–521
 [11] Bofill, P.: Underdetermined blind separation of delayed sound sources in the frequency domain. *Neurocomputing* **Vol. 55** (2003.) 627–641
 [12] Yeredor, A.: Time-delay estimation in mixtures. *Acoustics, Speech, and Signal Processing* **5** (2003) 237–240
 [13] Hojen-Sorensen, P., Winther, O., Hansen, L.: Mean field approaches to independent component analysis. *Neural Computation* **14** (2002) 889–918
 [14] Torkkola, K.: Blind separation of delayed sources based on information maximization. *ICASSP’96* (1996) 3509–3512
 [15] Swindelhurst, A.: Time delay and spatial signature estimation using known asynchronous signals. *IEEE Trans. on Sig. Proc.* **ASSP-33, no. 6** (1998) 1461–1470