Estimation of Velocity and Acceleration

We discuss here how we estimate the velocities and acceleration profiles of the joints of interest using the extracted keypoints from OpenPose or VideoPose3D. A smoothened derivative is computed on the 2D / 3D joint position time series using convolution with a biphasic filter [1; 2]. We choose a smoothened derivative with controllable smoothing parameters to be able to control the velocity and acceleration profiles. We find that using a 101-point filter achieves a lowpass filtering of about 2 Hz, giving a sufficiently smooth and physiologically plausible movement acceleration profile[3; 4].

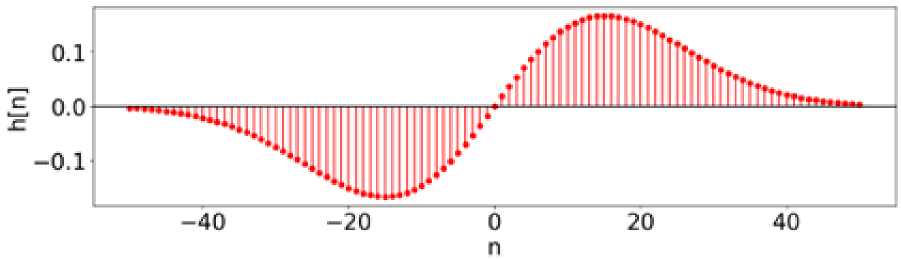

Figure 1: Plot of impulse response a 101 point bi-phasic filter for velocity and acceleration $\tau_1 = \tau_2 = 15$ and $d_1 = d_2 = 2$. The sampling rate is 100 Hz.

The biphasic filter is defined by

$$h[n] = \frac{1}{\tau_1\sqrt{2\pi}} e^{-\frac{(n-d_1)^2}{2\tau_1^2}} - \frac{1}{\tau_2\sqrt{2\pi}} e^{-\frac{(n-d_2)^2}{2\tau_2^2}} \tag{1}$$

We convolve the biphasic filter for differentiation of the position coordinates along each axis (x, y, z) for each joint to obtain the velocity for that joint. We also compute the magnitude of the velocity vector from the individual components.

$$\begin{aligned} v_x[n] &= h[n] * p_x[n] \\ v_y[n] &= h[n] * p_y[n] \\ v_z[n] &= h[n] * p_z[n] \\ v[n] &= \sqrt{v_x^2[n] + v_y^2[n] + v_z^2[n]} \end{aligned} \tag{2}$$

where $p_x[n]$,$p_y[n]$ and $p_z[n]$ are the position coordinates of the corresponding joint along x, y and z axes respectively and $*$ denotes the convolution operation. To prevent any artefacts from the differentiation operation, the position of each keypoint is extrapolated for 102 timesteps (one more than the filter length) before and after the time series by the starting and ending values respectively. This is under the assumption that these are the rest positions of the keypoint.

We use the same biphasic filter for differentiation of the velocity along each axis to obtain the acceleration along that axis as well as compute the magnitude of the acceleration vector.

$$
\begin{aligned}
a_x[n] &= h[n] * v_x[n] \\
a_y[n] &= h[n] * v_y[n] \\
a_z[n] &= h[n] * v_z[n] \\
a[n] &= \sqrt{a_x^2[n] + a_y^2[n] + a_z^2[n]}
\end{aligned} \tag{3}
$$

References

[1] D. J. Hermes, "Vowel-onset detection," *The Journal of the Acoustical Society of America*, vol. 87, no. 2, pp. 866–873, 1990.

[2] P. Rao, T. P. Vinutha, and M. A. Rohit, "Structural Segmentation of Alap in Dhrupad Vocal Concerts," *Transactions of International Society for Music Information Retrieval (TISMIR)*, vol. 3, pp. 137–152, Sep 2020.

[3] W. Pouw, J. de Wit, S. Bögels, M. Rasenberg, B. Milivojevic, and A. Ozyurek, "Semantically related gestures move alike: Towards a distributional semantics of gesture kinematics," in *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Body, Motion and Behavior: 12th International Conference, DHM 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part I*, pp. 269–287, Springer, 2021.

[4] L. Pearson and W. Pouw, "Gesture–vocal coupling in karnatak music performance: A neuro–bodily distributed aesthetic entanglement," *Annals of the New York Academy of Sciences*, vol. 1515, no. 1, pp. 219–236, 2022.