Read My Ears! Horse Ear Movement Detection for Equine Affective State Assessment

Supplementary Material

This document contains the supplementary material for CVPR 2025 ABAW Workshop Paper #51 and provides further insight into the results obtained with the three different methods tested (movDet, I3D+LSTM and VideoMAE+LSTM).

8. Dataset subjects sample data

To provide further insight into the dataset used, we provide sample frames of each of the 12 videos in this supplementary material (see Figure 12).

9. Supplementary qualitative results

In this section we show the qualitative results obtained from using each method on the original full length dataset videos.

9.1. movDet

MovDet works via sampling the video at a specific frame rate, then detecting and segmenting the horse's ear region. After that optical flow between both frames ear region is calculated. Finally we threshold the average magnitude of the flow vectors to obtain a ear movement/no-movement classification (see Figure 5). We applies movDet directly to the original dataset RGB videos, obtaining time wise ear-movement classifications across each video. We condense these results into a single graph for each of the 12 videos in the dataset, where average flow gradient, groundtruth and predicted movement classification can be observed.

Figure 13 shows the qualitative results of the movDet method on the 12 dataset videos.

9.2. I3D+LSTM

For I3D+LSTM method, we adopted a window based approach to process the videos, selecting the top configuration tested from Table 3. The method was applied to 50 FPS optical flow videos of the original data, using a window size of 50 frames and a stride of 35 frames. For each window we extracted the I3D flow stream features and classified it using the best configuration model. We condense these results into a single graph for each of the 12 videos in the dataset, where both groundtruth and predicted movement detection can be observed.

Figure 14 shows the qualitative results of the movDet method on the 12 dataset videos.

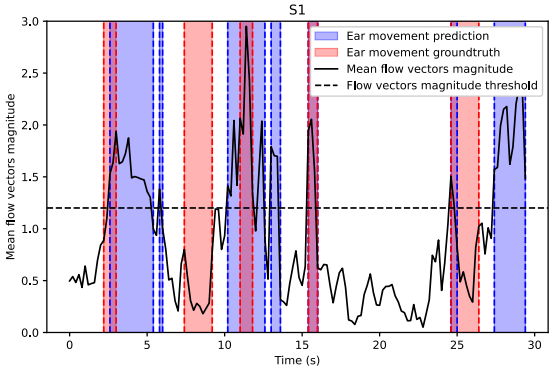
9.3. VideoMAE+LSTM

For VideoMAE+LSTM method, we adopted the same window based approach to process the videos, then selecting the top configuration tested from Table 3. In this case, the method applied to 50 FPS RGB videos of the original data, using a window size of 50 frames and a stride of 35 frames. For each window we extracted the VideoMAE features and performed classification. As before, we condense these results into a single graph for each of the 12 videos in the dataset, where both groundtruth and predicted movement detection can be observed.

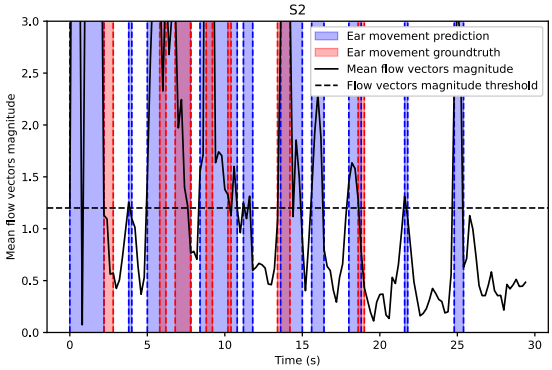
Figure 15 shows the qualitative results of the movDet method on the 12 dataset videos.

Figure 12. Sample frames for each of the 12 videos in the dataset in row-major order.

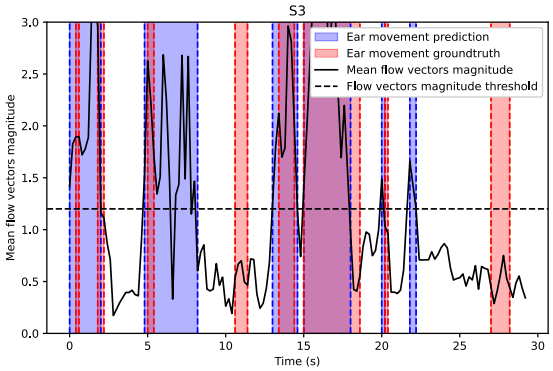
Qualitative Analysis: movDet



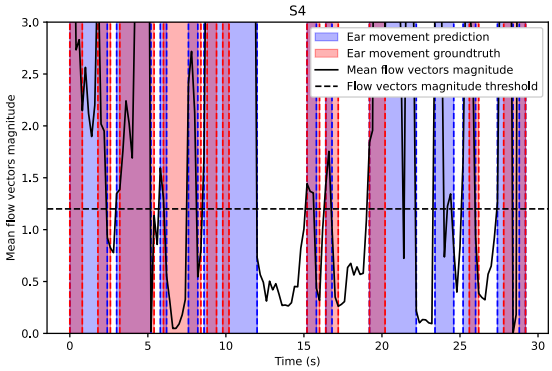
(a) S1



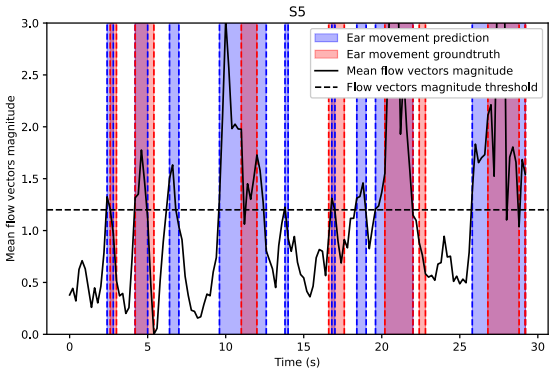
(b) S2



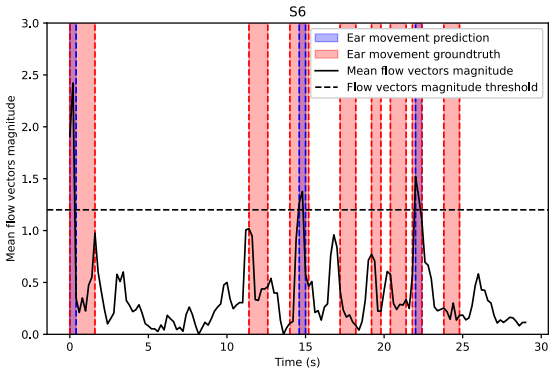
(c) S3



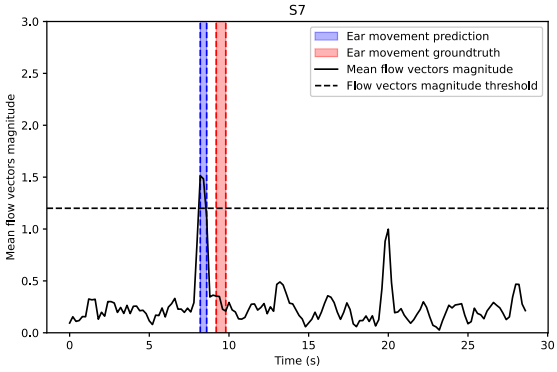
(d) S4



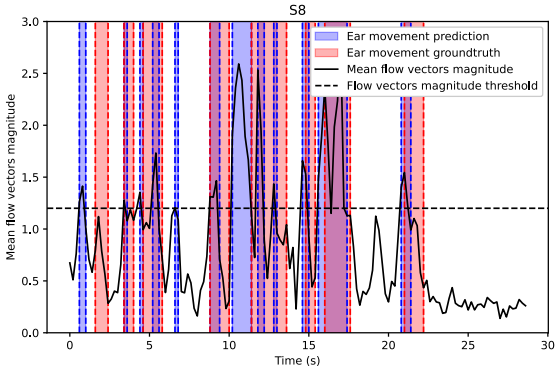
(e) S5



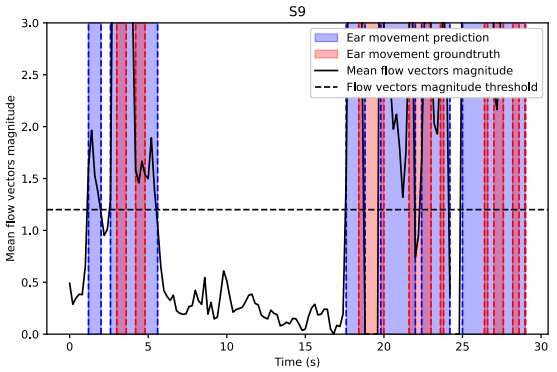
(f) S6



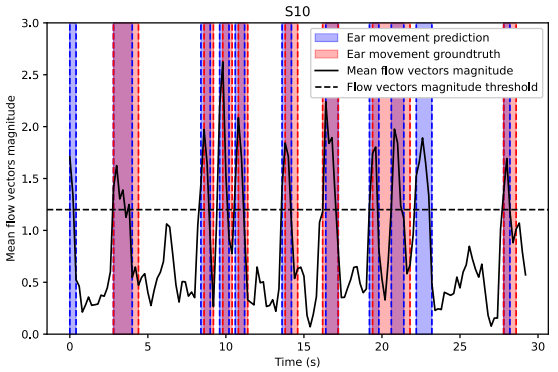
(g) S7



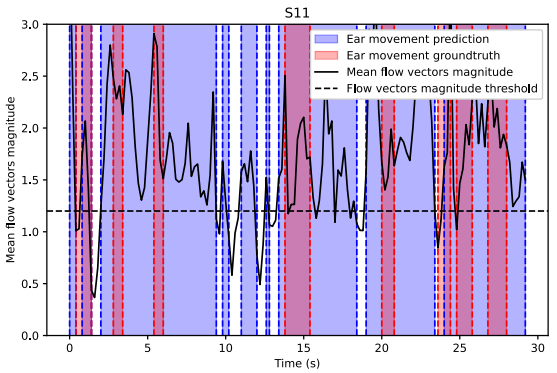
(h) S8



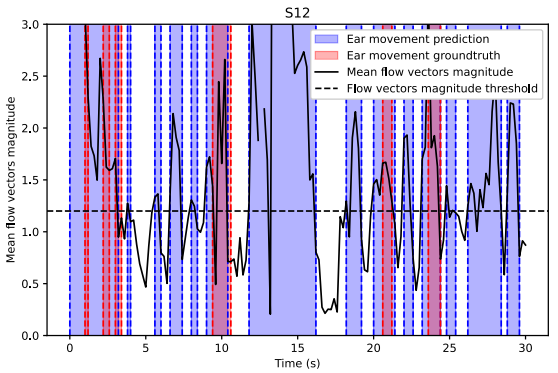
(i) S9



(j) S10



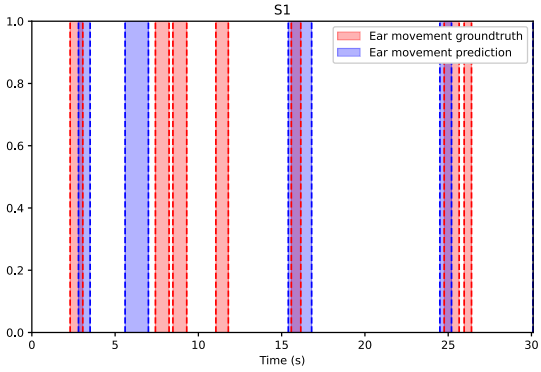
(k) S11



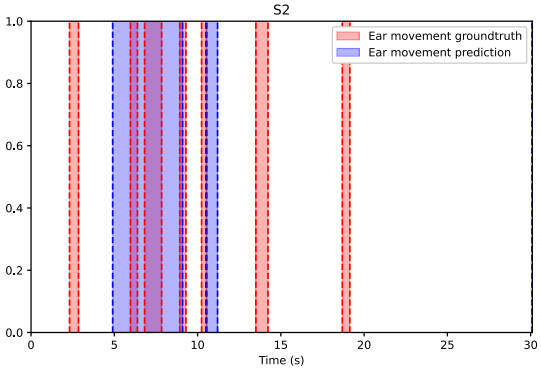
(l) S12

Figure 13. Qualitative analysis for movDet method on full-length horse videos.

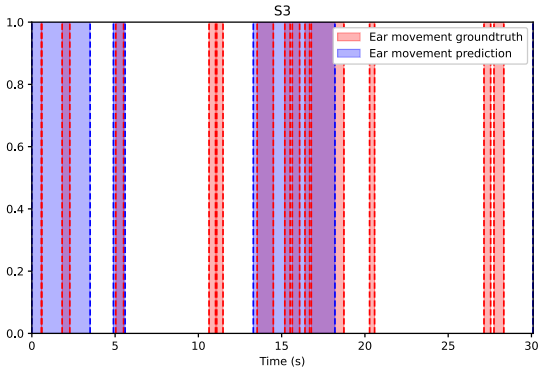
Qualitative Analysis: I3D+LSTM



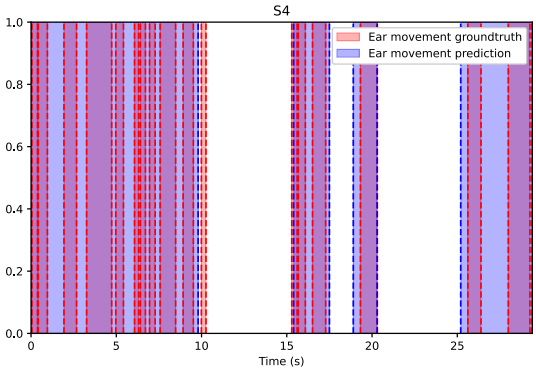
(a) S1



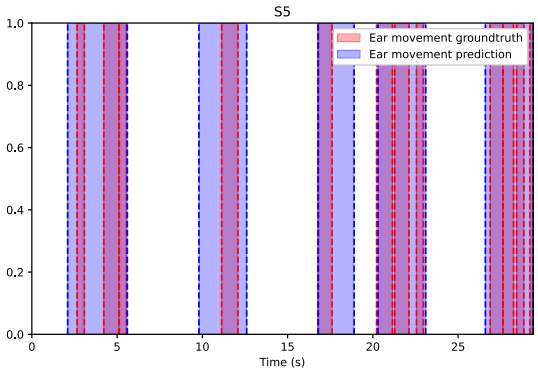
(b) S2



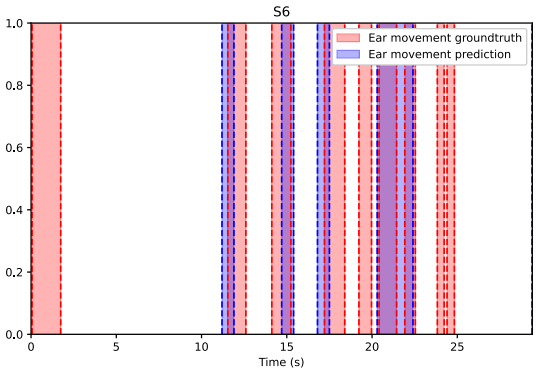
(c) S3



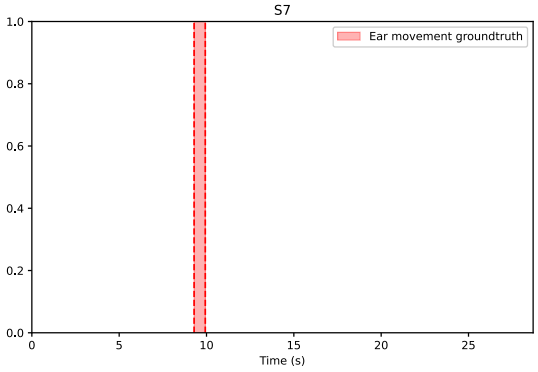
(d) S4



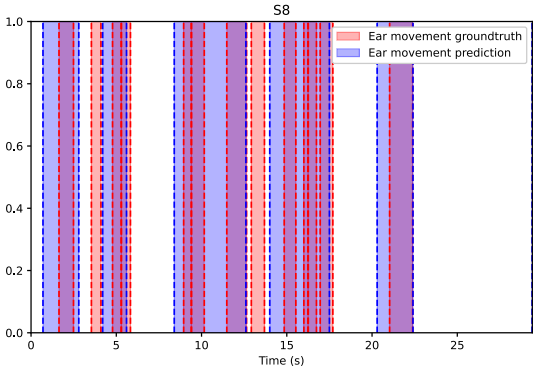
(e) S5



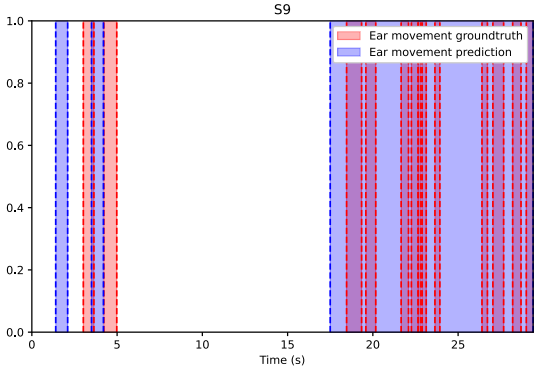
(f) S6



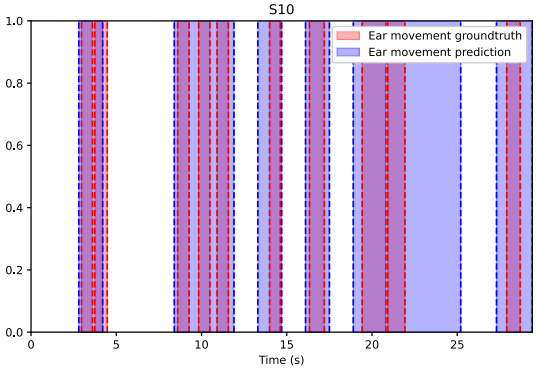
(g) S7



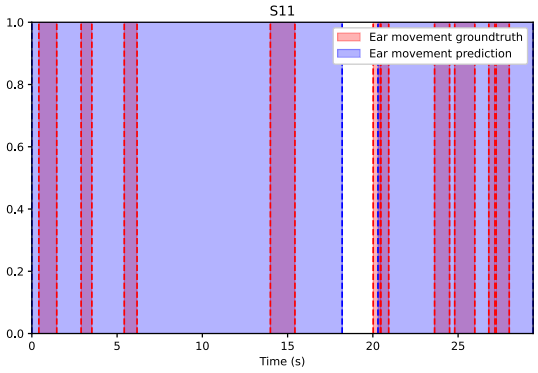
(h) S8



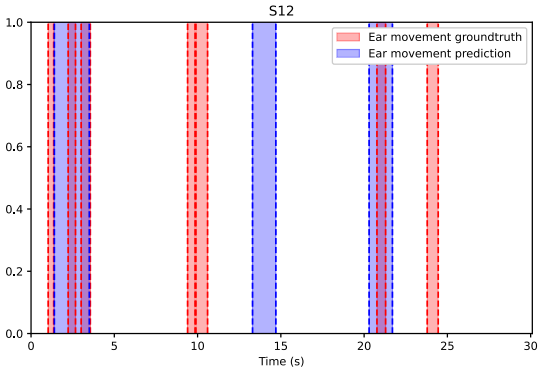
(i) S9



(j) S10



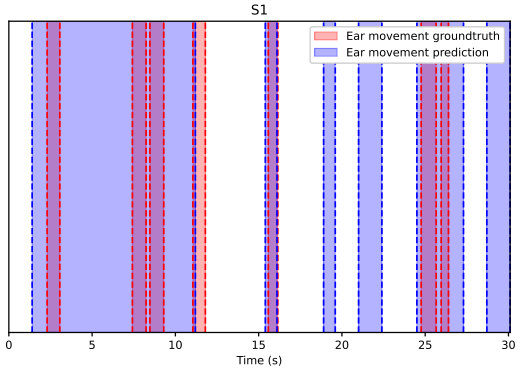
(k) S11



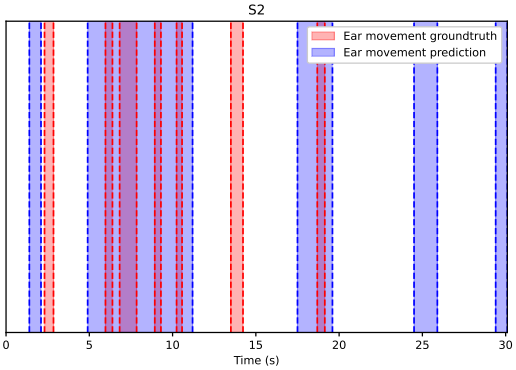
(l) S12

Figure 14. Qualitative analysis for I3D+LSTM method on full-length horse videos.

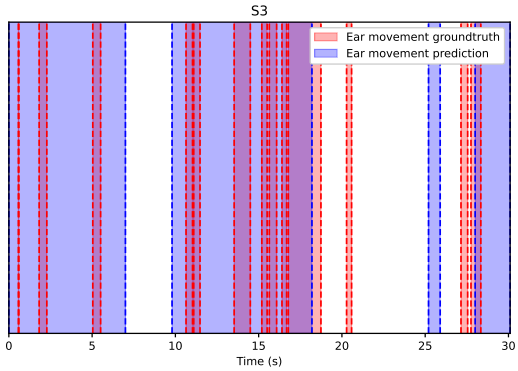
Qualitative Analysis: VideoMAE+LSTM



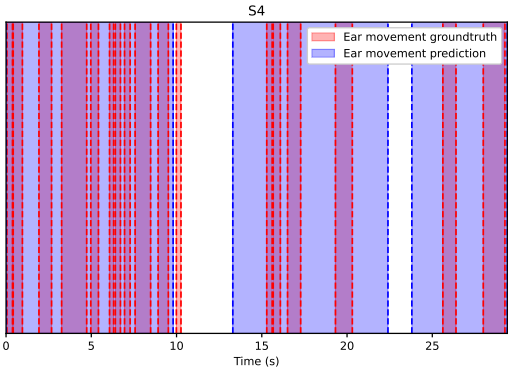
(a) S1



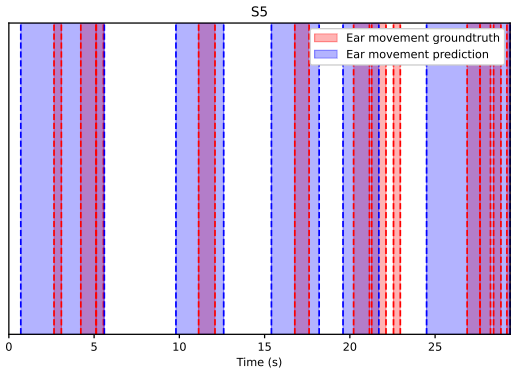
(b) S2



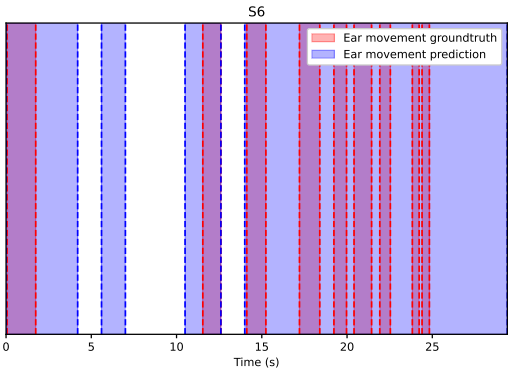
(c) S3



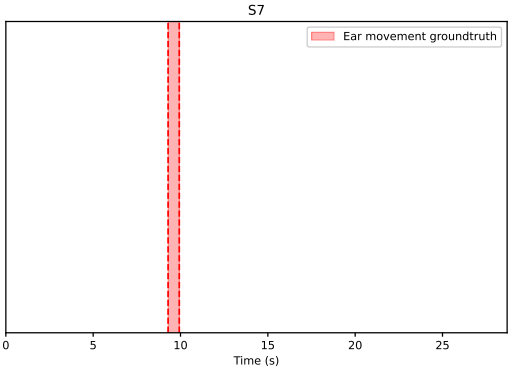
(d) S4



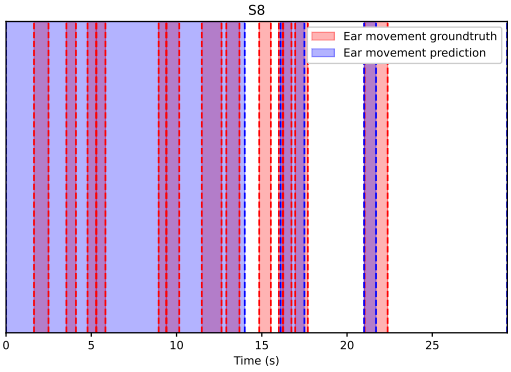
(e) S5



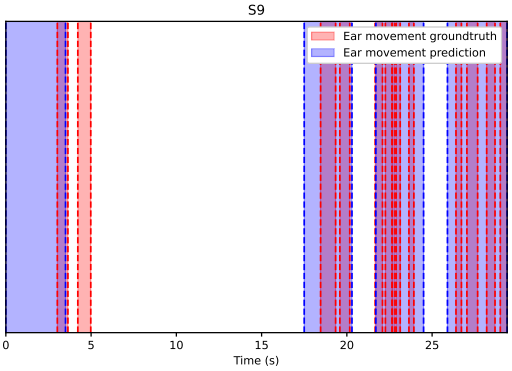
(f) S6



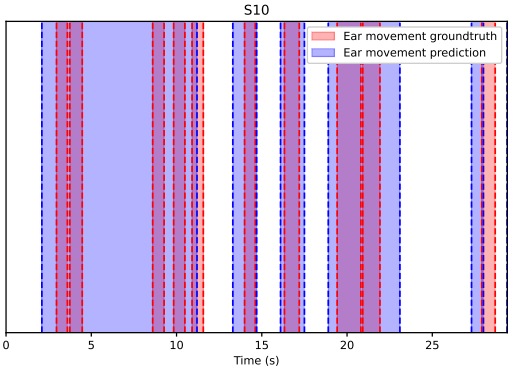
(g) S7



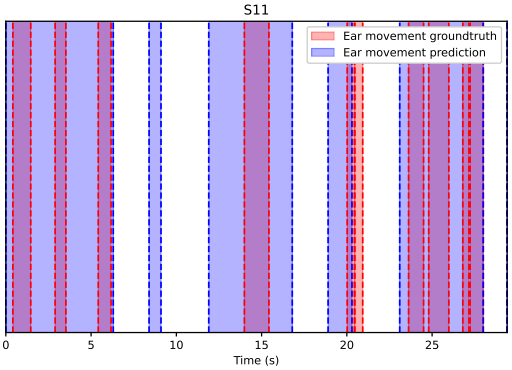
(h) S8



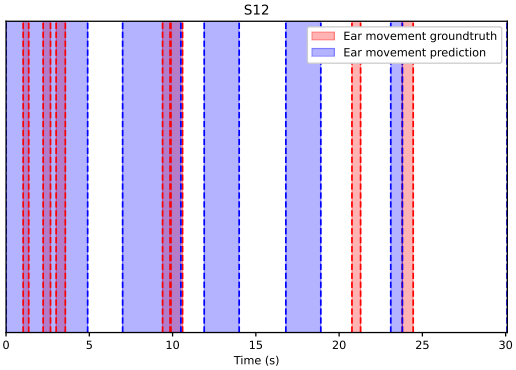
(i) S9



(j) S10



(k) S11



(l) S12

Figure 15. Qualitative analysis for VideoMAE+LSTM method on full-length horse videos.