COMPARING FACE-TRACKING ACTION UNITS WITH EMG DURING SPEECH

Hastiossadat Nozadi [*1], Emma Irwin [1], Jamie Cheung [1], Yadong Liu [1], Bryan Gick [1]
[1]Department of Linguistics, University of British Columbia, Vancouver, British Columbia, Canada

1 Introduction

Facial expressions have typically been detected either using electromyography (EMG) to detect muscle movements or simply by visual examination of the face [1]. Distal EMG is a device attached to the surface of a participant's face that measures the electrical signals caused by muscle activation. Face-tracking software like OpenFace 2.2 is particularly gaining popularity for video face-tracking without needing EMG [2].

OpenFace 2.2 is an open-source tool that detects movements of facial landmarks through images and videos of facial movement [3], triangulating virtual dots to record movement patterns called action units (AUs), which may be associated with particular facial muscle activations. For example, the "lip corner puller" action unit (AU12) in OpenFace 2.2 [2] has been associated with activation of the Zygomaticus Major muscle (ZM), while the "lip tightener" action unit (AU23) has been associated with activation of the Orbicularis Oris (OO) muscle [3]. These muscles are activated while smiling and are related to specific actions in the Facial Action Coding System (FACS) [4], which identifies muscle movements without labelling them as facial expressions.

2 Methods

The present study uses the data collected from a previous study [5]. The data from this study involved four speakers and was collected using video and EMG data. The four speakers were equipped with mini surface EMG sensors attached to the upper lip and cheek, which tracked OO and ZM activity, and their productions were digitally recorded. The researchers used a sampling rate of 19.6Hz while the OpenFace 2.2 data uses video limited to 30 fps making quantitative comparison challenging within our study. We automatically clipped the videos to our desired length surrounding the productions of interest with the moment of closure at 5 seconds.

This was done so that clips of the productions were roughly 10 seconds long. The stimuli and sentences are the same as those provided by Liu et al. through the corpus data [5]. We used the data from four speakers who read 15 different sentences aloud under two different facial conditions, smiling and neutral, focusing on stop productions [5].

The data from the four English speakers was processed and analyzed. As mentioned above, the clipped videos were input into OpenFace 2.2 and analyzed for AUs 12 and 23 regression. Two seconds surrounding the productions of interest were extracted for four seconds. The OpenFace 2.2 timestamps were adjusted to align with the corpus data from the Liu et al. study [5].

Additionally, data from each participant was compiled, and averages for the appropriate AUs were calculated. These averages were graphed with Matplot and Numpy libraries using smoothed figures (z-score of ZM, AU12 regression, AU23 regression, and z-score of OO, against time) and compared. This was utilized to detect any similarities or differences in the EMG and OpenFace 2.2 data between the facial expressions two seconds before and after the targeted segment determined by EMG data. The ZM's z-scores from the EMG readings were contrasted with the AU12 regression (lip corner puller) data, while the z-scores for OO readings were compared against the AU23 regression (lip tightener). The EMG and AU graphs were manually annotated. The activation onset was recorded for both muscles under both conditions and was averaged out.

3 Results

There appears to be a general trend of more gradual changes in the EMG readings compared to the OpenFace 2.2 reports with more drastic readings (Fig. 1). The overall peaks regarding the OO closure, compared to the AU23, align relatively well within the duration of the stop itself. The drops of the ZM also align when it comes to the time of production of the closure. Trends are relatively equal. The average EMG latency came out to around 41ms. In the smile condition, this averages 43ms, while in the neutral condition, 41ms. Notably, the EMG data for the smile condition of P213 had to be omitted as there was no clear onset of activation before the stop. The average mean latency of ZM under neutral conditions was 43ms, while under smile conditions this was not applicable as three of the four charts showed no obvious onset. The average mean latency of the OO under neutral conditions was 39ms, while under smiling conditions, it was 44ms.

The graphs for OpenFace AUs under neutral conditions show an average latency of 20ms, with an average of 15ms under neutral conditions and 24 ms under smile conditions. Further, AU12 under neutral conditions had a latency of 19ms and 30ms under smiling conditions. AU23 had an average mean latency of 13ms under neutral conditions and an average mean latency of 20ms under smiling conditions.

4 Discussion

OpenFace 2.2 was found to have similar general trends in the EMG data when comparing the z-scores. The two readings tended to peak in relatively the same positions. This was most true for OO and AU23 comparisons, where the peak of the readings consistently happened at 0s, which is also the point of closure. In the case of ZM and AU12, participant 003 was

[*] hnozad01@student.ubc.ca

found to have rather strongly opposing trends, implying little correlation. This may have been caused by lip tightening, which is associated with the AU23, being more apparent to the front on reading available to OpenFace 2.2 via the video data.

There was found to be greater latency between the EMG and the moment of closure compared to the OpenFace 2.2 reading on average when comparing onsets. This may be due to EMG picking up the initial electric pulse sent to the muscle before the movement itself, as discussed by Roberts and Gabaldon [6]. This contrasts with the data of OpenFace 2.2, which is collected based on the apparent visual change between video frames.

5 Conclusion

While OpenFace can supplement more accurate tools like distal or subdermal EMGs, it cannot replace them. OpenFace relies entirely on visual input. OpenFace is essentially ‘lip reading’ your muscle movements, and the AUs should remain as action descriptors rather than synonyms for a muscle or muscle group. The variation in trends of P003 proves that OpenFace is not as effective at assessing the engagement of specific muscles as EMG. OpenFace can accurately confirm broader movements, such as closures and smiles, at more precise timings, as seen by the latency comparison. Still, it does not seemingly act as a stand-in for reading muscle activations at the electrical level, and the AUs do not accurately map the muscles of EMG alone.

Acknowledgements

Research supported by an NSERC Discovery grant to the last author for acknowledgments.

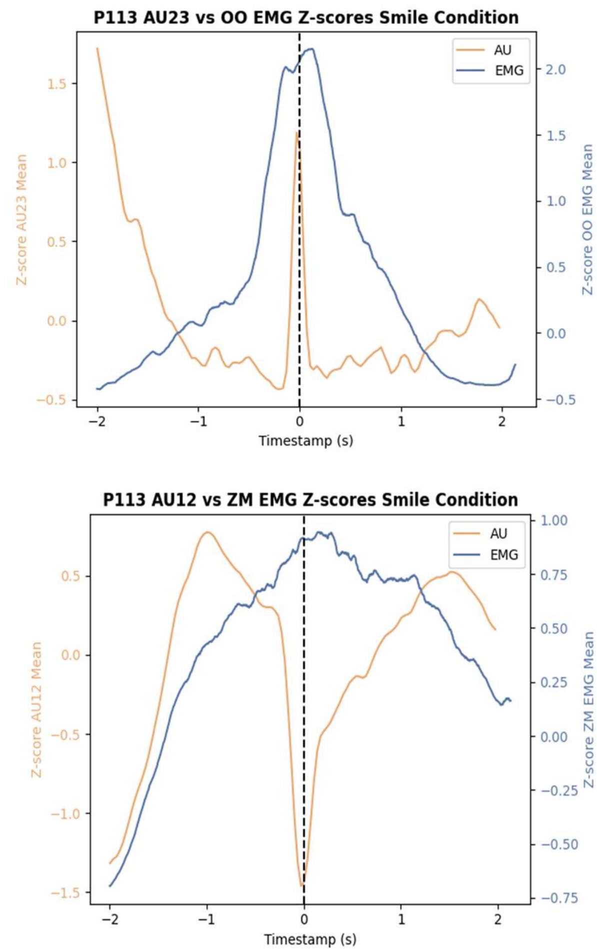

Figure 1: Above: mean OO trend (blue) and mean AU23 trend (orange). Below: mean ZM trend (blue) and mean AU12 trend (orange).

References

[1] M. Hamedi, S. Salleh, A. Noor, T. S. Tan, and I. Afizam, *Comparison of Different Time-Domain Feature Extraction Methods on Facial Gestures? EMGs*, vol. 12. 2012.

[2] K. Masai, M. Perusquia-Hernandez, M. Sugimoto, S. Kumano, and T. Kimura, “Consistent Smile Intensity Estimation from Wearable Optical Sensors,” presented at the 2022 10th International Conference on Affective Computing and Intelligent Interaction, ACII 2022, 2022.

[3] T. Baltrusaitis, A. Zadeh, Y. C. Lim and L. -P. Morency, "OpenFace 2.0: Facial Behavior Analysis Toolkit," *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, China, 2018, pp. 59-66, doi: 10.1109/FG.2018.00019.

[4] M. Perusquía-Hernández, F. Dollack, C. K. Tan, S. Namba, S. Ayabe-Kanamura, and K. Suzuki, “Facial movement synergies and Action Unit detection from distal wearable Electromyography and Computer Vision,” *ArXiv*, Aug. 2020, Accessed: Apr. 30, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Facial-movement-synergies-and-Action-Unit-detection-Perusqu%C3%ADa-Hern%C3%A1ndez-Dollack/1356fcc9cc69179c4eaef6103ac23140effc75e6

[5] Y. Liu, T. Chan, G. Purnomo, B. Gick, Talking while smiling: Suppression in an embodied model of coarticulation. In M. Tiede, D. H. Whalen & V. Gracco (Eds.) *Proceedings of the 12th ISSP*. Haskins Press: New Haven, CT. 130-133. 2021.

[6] T. J. Roberts and A. M. Gabaldón, “Interpreting muscle function from EMG: lessons learned from direct measurements of muscle force,” *Integr Comp Biol*, vol. 48, no. 2, pp. 312–320, Aug. 2008, doi: 10.1093/icb/icn056.