

# Exploring new methods for measuring, analyzing, and visualizing facial expressions

Lyke Esselink, Marloes Oomen, Floris Roelofsen

University of Amsterdam

{l.d.esselink, m.oomen2, f.roelofsen}@uva.nl

## Abstract

We explore new methods for measuring, analyzing, and visualizing facial expressions and demonstrate the utility of these methods in a case study on polar questions in Sign Language of the Netherlands.

**Keywords:** Facial expressions, Depth-sensing cameras, Clustering, Visualization

## 1 Introduction

This paper aims to make a methodological contribution to experimental research on facial expressions. As sign language researchers, our primary motivation for exploring new methods for investigating facial expressions is that, while facial expressions play a crucial role in sign language communication, current methods to measure and analyze them have, as we will argue below, substantial limitations. That said, we believe that the new methods we present here potentially have much broader applicability. Besides in sign language research, they may also be used in other domains where facial expressions are of interest, such as the analysis of multi-modal communication (cf. Steen et al., 2018), emotion recognition (cf. Jack et al., 2012; Li and Deng, 2022), and the diagnosis of certain medical conditions (cf. Hartmann et al., 2022).

We intend to make three methodological contributions, related to three important stages of scientific investigation: data collection, data analysis, and data visualization. More specifically, we explore the use of *depth-sensing cameras* to track features of a person's face (such as brow raises, eye squints, and mouth shapes), the use of *machine learning techniques* – in particular, clustering algorithms – to discover prominent patterns in such data, and the use of realistic *humanoid avatars* to visualize the observed patterns. To provide a concrete demonstration of the utility of these methods, we apply them in a case study on polar questions in Sign Language of the Netherlands (NGT).

The paper is structured as follows. Section 2 first considers methods currently used in sign language research to measure facial expressions, and discusses some of their limitations.<sup>1</sup> Subsequently, a new method making use of depth-sensing cameras is introduced. Section 3 presents two ways of analyzing data obtained with depth-sensing cameras, and Section 4 introduces a technique for visualizing the patterns found in such data. Section 5 discusses limitations of the proposed methods, and finally, Section 6 concludes.

---

<sup>1</sup>Sections 2.1 and 2.2 overlap to a large extent with Sections 2.1 and 2.2 in Esselink et al. (2023).

## 2 Data collection

### 2.1 Traditional methods

Research on facial expressions in sign languages is generally based on video data. Such data, however, is two-dimensional and therefore never fully captures the actual physical reality that it represents, which is three-dimensional. Furthermore, important details are sometimes not visible on video footage because of a limited frame rate, limited resolution, motion blur, or occlusion (e.g. a hand in front of the face). Ideally, researchers would be able to base their analysis on data that captures the poses and movements of a signer, in particular their facial expressions, in a format that stays closer to the original, with less inherent transformation (3D to 2D), compression (frame rate, resolution), and noise (blur, occlusion).

Analysis of video data starts with annotation. This process is notoriously laborious, especially when facial expressions are concerned. Even when done with great care, manual annotation has some inescapable limitations. It is inherently *subjective* (two annotators may disagree as to whether an eyebrow is raised or neutral), *not robustly reproducible* (a single annotator may label an eyebrow as raised one day, and the same eyebrow as neutral six months later), and inherently *categorical* (an eyebrow can be labeled as raised or neutral, perhaps ‘half raised’, but not ‘raised to degree 0.35’) while in reality eyebrow raise and other facial features are quantitative/continuous variables, not categorical ones — so in the annotation phase the data is further ‘compressed’, losing part of the original information. Ideally, researchers would have a method to annotate facial expressions that is objective, reproducible, quantitative rather than categorical (meaningful categories may be identified in a later stage of analysis, but should not be imposed on us from the start), and less laborious.

### 2.2 Recent advances

Kimmelman et al. (2020) and Kuznetsova et al. (2021, 2022) have recently explored a new approach to address the limitations of manual annotation of facial expressions. They use face recognition software (OpenFace) to automatically detect a signer’s eyebrows and eye-corners, and compute a degree of eyebrow raise/lowering in terms of the distance between these. This method to extract degrees of eyebrow raise/lowering from video data is automatic, objective, and quantitative. However, it still has some limitations.

First, it is *not robustly reproducible*, because the measurements of OpenFace are affected by the distance and angle between the camera and the signer’s face, which are impossible to keep constant within and across recordings. Second, measurements of relevant facial features like brow raise are *indirect*. OpenFace detects facial landmarks. Features like brow raise have to be derived from distances between certain landmarks, but this is not without issues because these distances partly depend on camera angle and distance (see Kuznetsova et al., 2021, for discussion). Third, the method as developed so far is restricted to two facial features: eyebrow raise/lowering and head rotation, though in future work it may be extended to other relevant facial features as well. Most importantly, the proposed method still takes 2D *video data* as its starting point. This is what OpenFace takes as its input. So, while this body of work makes an important first step in addressing the limitations of manual annotations, it does not address the issues of inherent transformation, compression and noise associated with video data.

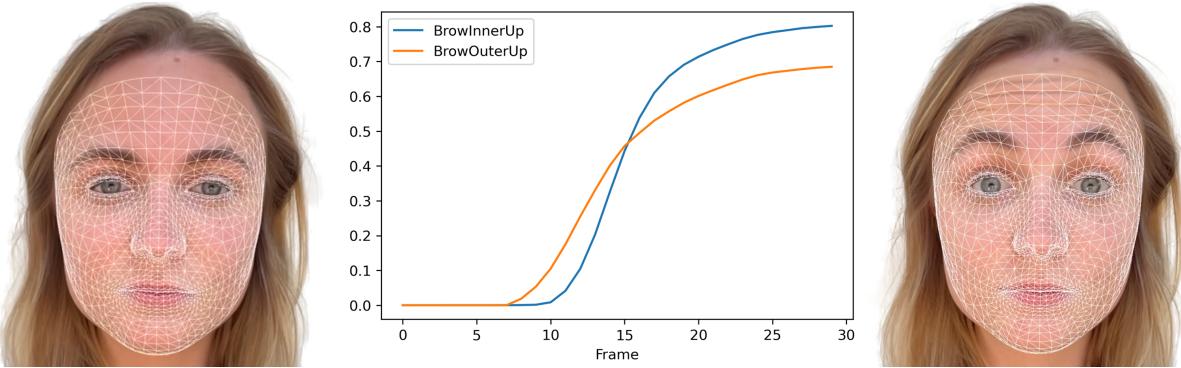


Figure 1: Example measurements of eyebrow blendshapes

### 2.3 Depth-sensing cameras

We explore a way to overcome these issues, at least partly, by using a depth-sensing camera in addition to ordinary video cameras for data collection. Specifically, we make use of a TrueDepth camera built into an iPhone 13 in combination with the free Live Link Face application for iOS by Epic Games (2022a). This hardware/software combination can be used to measure 52 facial features, called ARKit *blendshapes*. Not all 52 blendshapes are relevant for the study of facial expressions in sign languages. Moreover, many features concerning the shape of the mouth are heavily affected by mouthing, and thus cannot straightforwardly form the basis for an investigation of facial expressions. For our purposes, we selected 9 relevant blendshapes: BROWINNERUP, BROWOUTERUP, BROWDOWN, EYEWIDE, EYESQUINT, CHEEKSQUINT, NOSESNEER, MOUTHSHRUG, and MOUTHFROWN. Detailed motivation for selecting these particular blendshapes is given in Esselink (2023). Some of the selected blendshapes are originally measured with separate values for the left and right side of the face. We average these measurements to obtain a single value, as they are generally very similar. Blendshape values range from 0 to 1, and are captured at a high frame rate of 60 fps. For illustration, the plot in Figure 1 shows how the values of BROWINNERUP and BROWOUTERUP rise when the person recorded raises their eyebrows.

This method, like the recently explored methods based on OpenFace, but unlike more traditional methods, is automatic, objective, and quantitative. Moreover, unlike OpenFace, which performs landmark detection based on video input, this method bypasses the main issues associated with video data, and directly measures facial features that are of interest for sign language research as opposed to landmark coordinates, which first have to be translated into blendshape values, something which, as mentioned above, cannot always be done in a straightforward way, if at all.

### 2.4 Application in the domain of NGT polar questions

To provide a concrete illustration of how the proposed method may be used in practice, we present a case study on polar questions in NGT. The larger empirical investigation that this case study was part of is discussed in more detail in Oomen and Roelofsen (2023). The official NGT dictionary (Schermer and Koolhof, 2009, p.39) and the descriptive NGT grammar of Klomp (2021) both describe polar questions as being expressed with raised eyebrows and the head tilted slightly forward. Experimental work, however, has reported quite some

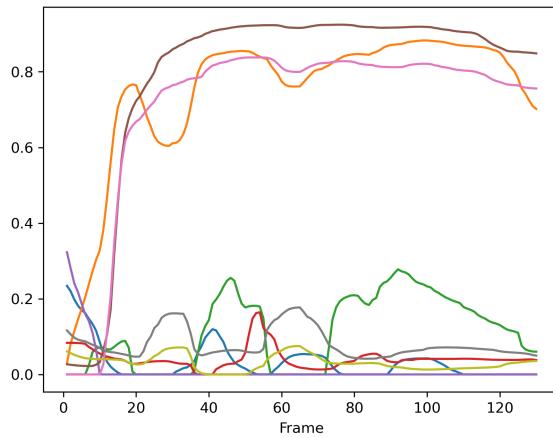


(a) NeutPos

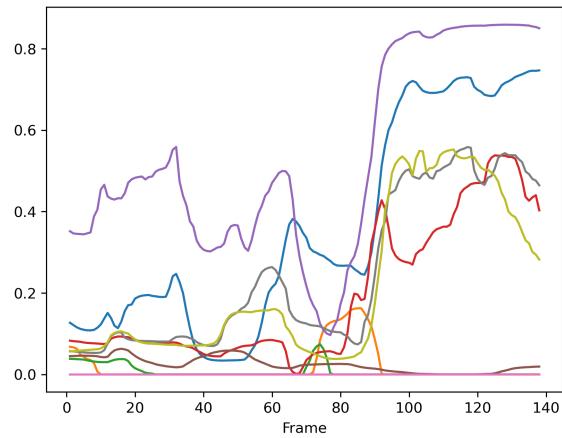


(b) PosNeg

Figure 2: Facial expressions accompanying the signs KIM – WEEKEND – HOME in questions signed by a participant in the conditions NeutPos (a) and PosNeg (b)



(a) NeutPos



(b) PosNeg



(c) Legend

Figure 3: Blendshape values accompanying the signs KIM – WEEKEND – HOME in questions signed by a participant in the conditions NeutPos (a) and PosNeg (b)

variation. For instance, besides raised eyebrows, experimental results show that lowered eyebrows also occur in polar questions in NGT (Coerts, 1992; de Vos et al., 2009). Other non-manual markers that have been claimed to be typically involved in polar question marking in sign languages in general (beyond just NGT) include ‘wide open eyes’, ‘head forward’, ‘body forward’, and ‘continuous eye contact with the addressee’ (Cecchetto, 2012; Pfau and Quer, 2010; Zeshan, 2004).

We recorded four native NGT signers who were instructed to ask questions to two confederates in a role-play setting, using a depth-sensing camera as well as an ordinary video camera. The details of the elicitation task are reported in Oomen and Roelofsen (2022). The experimental design controlled for two contextual factors which may influence the way in which a polar question is expressed: prior speaker belief and immediate contextual evidence concerning the question radical. For instance, when prompted to ask *Is the zoo open?* a participant could be given prior information by a confederate that the zoo would probably be open, but be faced with immediate contextual evidence (through role play with another confederate) that it is actually probably closed, and similarly for other combinations of prior belief and immediate contextual evidence.

For the experimental condition exemplified here we use the label ‘PosNeg’. The first part of the label, ‘Pos’, refers to the positive prior speaker belief; the second part, ‘Neg’, refers to the negative immediate contextual evidence. Experimental conditions involve a combination of positive/neutral/negative prior speaker belief and positive/neutral/negative contextual evidence, with the exception of positive-positive and negative-negative, as it is not natural to ask a question in these cases. Since we recorded participants with a depth-sensing camera as well as an ordinary video camera, we obtained fine-grained, quantitative data on the facial expressions that are used to mark polar questions in NGT, across conditions and participants.

Figures 2 and 3 illustrate the type of data that we obtained. Figure 2 shows stills of the signs KIM, WEEKEND and HOME – as occurring in the question *‘Is Kim (not) home this weekend?’* – asked in two different conditions. As can be observed, the facial expressions accompanying the signs are strikingly different depending on the condition in which the question is asked. This impression is confirmed in Figure 3, which shows the blendshape values measured in the recordings corresponding to the stills in Figure 2. In the condition with *neutral* prior speaker belief and *positive* immediate contextual evidence (Figures 2a and 3a), we see high values of BROWINNERUP, BROWOUTERUP, and EYEWIDE. On the other hand, in the condition with *positive* prior speaker belief and *negative* immediate contextual evidence (Figures 2b and 3b), we see high values of BROWDOWN and EYESQUINT. Moreover, we observe peaks for CHEEKSQUINT, NOSESNEER, and MOUTHSHRUG during the sign HOME.

### 3 Data analysis

The data obtained in this way is very rich and lends itself to various types of quantitative analyses. We explore two of these here: comparisons of patterns found across conditions and the application of machine-learning clustering algorithms to identify the most prototypical facial expressions used to mark polar questions in NGT.

#### 3.1 Quantitative comparison across conditions

For each condition (NeutPos, PosNeg, etc.) we computed the average progression of blendshape values across all participants and scenarios. This allows us to compare conditions.

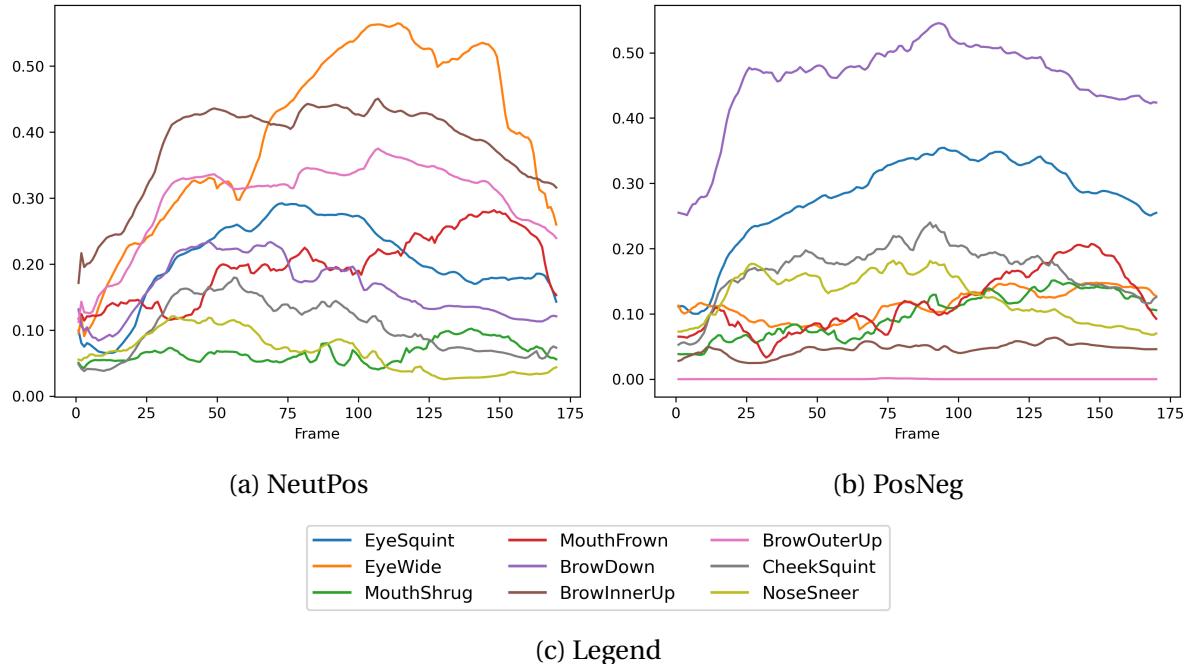


Figure 4: Mean blendshape values for condition NeutPos (a) and PosNeg (b)

Note that, in a similar way, one could draw comparisons across participants, sentence types with different syntactic structures, etc.

Figure 4a shows that, on average in the NeutPos condition, the blendshapes EYEWIDE, BROWINNERUP, and BROWOUTERUP are activated the most. Generally, the blendshapes related to the eyebrows stay relatively constant throughout a question in this condition, whereas the eyes tend to widen near the end. We see that in some cases, a question in this condition may end with MOUTHSHRUG, although this is not necessarily the case. Moreover, we see that BROWDOWN and EYESQUINT may also occur in this condition, albeit to a lesser extent than EYEWIDE, BROWINNERUP, and BROWOUTERUP. The stills in Figure 2a are in line with the general trends in Figure 4a.

In contrast to the NeutPos condition, we see that on average, the most active blendshapes in the PosNeg condition are BROWDOWN and EYESQUINT (Figure 4b). The values of these two features quickly rise at the beginning of the question, and stay relatively constant throughout. Again, MOUTHSHRUG may occur at the end of a question in this condition. Although EYEWIDE and BROWINNERUP do see some activation, their mean values are very low. This indicates that, in the PosNeg condition, participants strongly favor marking polar questions with furrowed eyebrows and squinted eyes over raised eyebrows and wide-opened eyes. Again, these trends are in line with the stills in Figure 2b.

### 3.2 Clustering to identify the most prototypical facial expressions

Another question that can be addressed based on this data is: What are the most prototypical facial expressions used to mark polar questions in NGT? To answer this question, we employ a clustering algorithm, which divides a dataset into naturally occurring ‘clusters’ of data-points. In our case, similar facial expressions will be grouped together to form one cluster. The primary clusters then correspond to the most prototypical facial expressions for polar

question marking in NGT.

Every facial expression can be represented as a vector in a nine-dimensional space, where each element of the vector corresponds to the value of one of the blendshapes (for instance, [1, 1, 0, 0, 0, 0, 0, 0, 0] would represent a facial expression with maximally raised eyebrows and opened eyes, while all other blendshapes are not activated at all). In this case-study we use HDBScan, a density-based clustering algorithm (see Esselink, 2023 for discussion of other clustering algorithms). This means that the algorithm forms clusters on the basis of the density at which datapoints occur in the nine-dimensional space. If an area is densely populated, the datapoints in that area form a cluster. Datapoints in sparsely populated regions are not assigned to any cluster.

Before we feed our data to a clustering algorithm, it needs to be prepared for processing. Utilizing the raw dataset can cause the clustering algorithm to overly emphasize facial expressions from individual participants, often occurring in adjacent frames within single recordings. These facial expressions are most similar to one another, resulting in clusters that lack generalizability across the dataset. As we aim to find the most prototypical facial expressions across participants in varying experimental conditions, pre-processing of the data is required. For this project, four pre-processing steps were taken: 1) normalization; 2) removal of transition frames; 3) downsampling; and 4) categorization. We provide a brief overview of each of these steps below; for a comprehensive explanation, we refer to Esselink (2023).

As a first step, it is necessary to normalize the data. As mentioned in Section 2.3, the Live Link Face app measures blendshapes with values between 0 and 1. However, not every person is able to reach a measured value of 1 for each blendshape. For instance, the person in Figure 1 is fully raising their eyebrows, but the measured blendshapes only reach a maximum value of 0.85 and 0.75 respectively, whereas another person might have higher or lower maximum values. A facial expression where both participant A and participant B are raising their eyebrows to the fullest extent should be regarded equal, even if the actual recorded values vary. Therefore, we adjust the data so that for each participant, each blendshape is normalized on a scale of 0 to 1, based on the highest recorded measurement across all recordings of that blendshape for that participant.<sup>2</sup> The normalized data is then used as input for the clustering algorithm. An example of this is visualized in Figure 5, showing raw and normalized measurements of two blendshapes for participant A (Figure 5a) and participant B (Figure 5b). For further details of the normalization process, see Esselink (2023).

Second, we remove so-called ‘transition frames’. These are the frames in a recording in which the participant transitions from one facial expression to another. Given that our focus is on facial expressions themselves, and not transitions between facial expressions, we can simplify our dataset by removing these frames. For details on how we carry this out, see Esselink (2023). Figure 5c shows the values measured for the blendshapes EYEWIDE, BROWINNERUP, and MOUTHFROWN for a randomly selected recording. The lines in this plot represent all blendshape measurements, whereas the dots represent selected frames. As we can see, only frames that are relatively stable compared to neighboring frames are retained.

Next, we downsample the dataset by removing every third frame for every recording. While the general patterns of the recorded data are preserved, the downsampled dataset is significantly smaller, at only two-thirds of the size of the original dataset. This reduces the computational complexity while retaining information.

<sup>2</sup>A recommendation for future research is to make one separate ‘calibration’ recording for each participant, in which they are instructed to engage each facial feature to the fullest of their abilities. This recording can then be used for normalization.

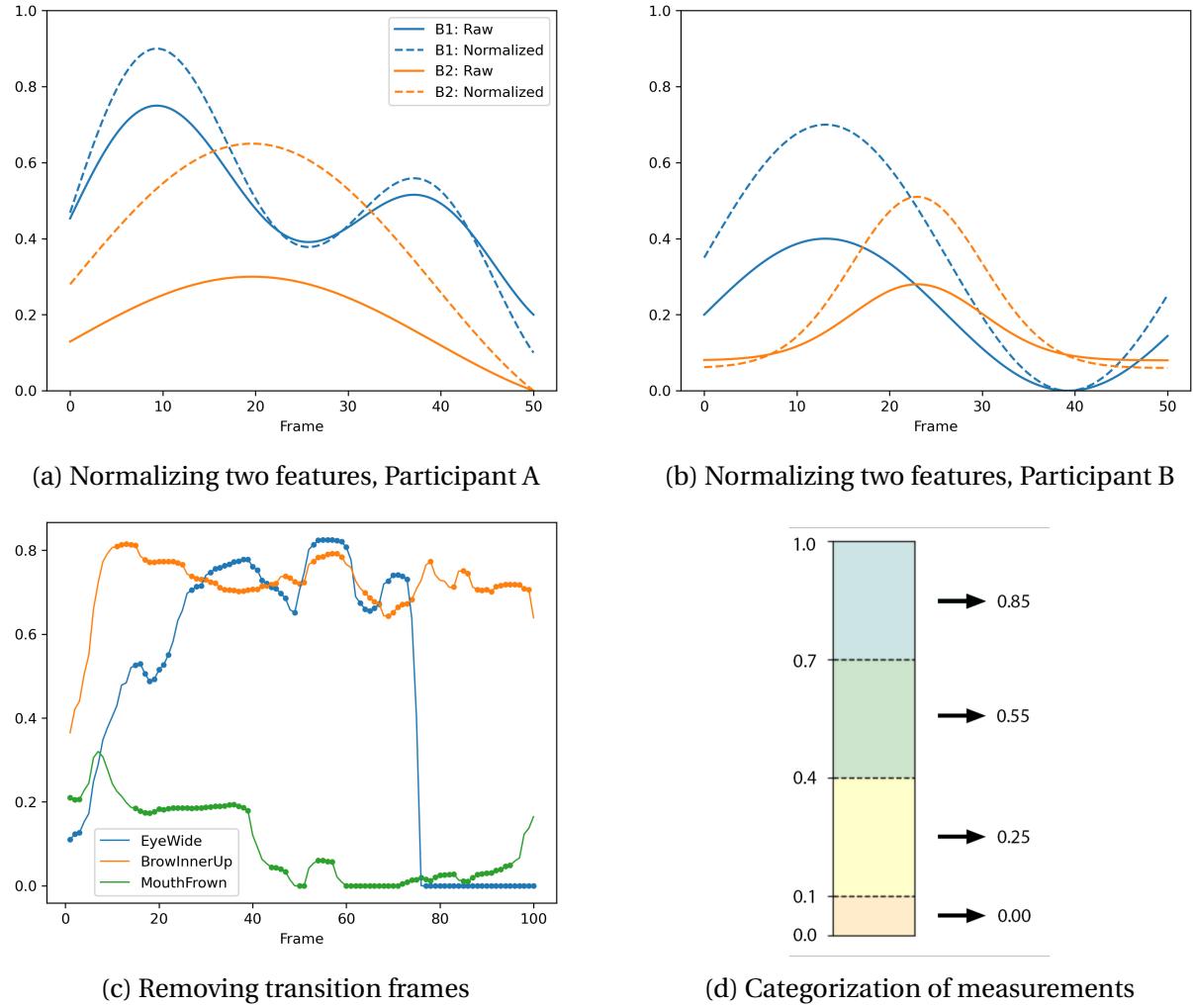


Figure 5: Pre-processing data

Finally, we restructure the data into categorical bins. Although the continuous measurements obtained by the depth-sensing camera allow for a highly detailed analysis, they also lead to clusters that are over-fitted to single participants. We group the data into four categories, as seen in Figure 5d. Datapoints with the lowest measurements, between 0.00 and 0.10, are given an ‘inactive’ value of 0.00; datapoints with slightly higher measurements, between 0.11 and 0.40, are given a ‘low activation’ value of 0.25; datapoints with even higher measurements, between 0.41 and 0.70, are given a ‘medium activation’ value of 0.55; and finally, datapoints with the highest measurements, between 0.71 and 1.00, are given a ‘high activation’ value of 0.85. Categorization in this manner preserves some of the quantitative information in our dataset, while making it possible for the clustering algorithm to identify clusters of facial expressions across participants.

## 4 Visualization of results

The analysis yields three main clusters. The facial expressions corresponding to the averages of these clusters are shown in Figure 6. Cluster A is marked by high values of BROWINNERUP

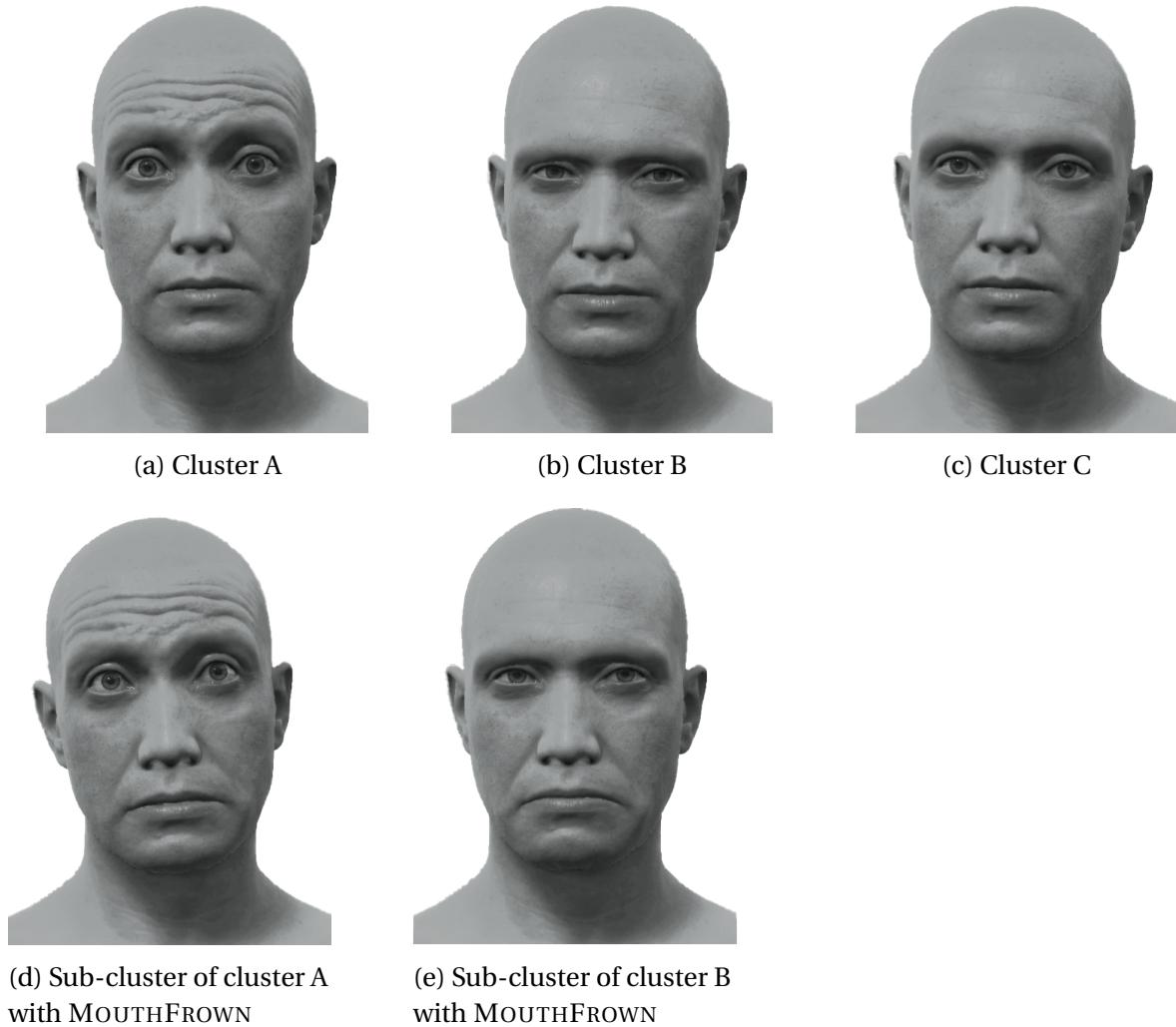


Figure 6: Visualizations of the main facial expression clusters in (a)-(c), and examples of sub-clusters of Cluster A and Cluster B with MOUTHFROWN in (d) and (e), respectively

(0.75), BROWOUTERUP (0.67), and EYEWIDE (0.82); Cluster B by high values of BROWDOWN (0.60) and moderate values for EYESQUINT (0.39); and Cluster C by low values ( $\leq 0.20$ ) for all blendshapes, thus representing relatively neutral facial expressions. A covers 19% of the data, B 39%, and C 27% (15% of the datapoints are not classified as belonging to any of these main clusters). Sub-clusters of Cluster A and B mainly differ in the value of MOUTHFROWN.

It is possible to visualize facial expressions recorded in Live Link Face through the use of a MetaHuman avatar in Unreal Engine, developed by Epic Games (2022a,b,c).<sup>3</sup> All we need for this is Unreal Engine and a computer capable of running this software, a MetaHuman blueprint (readily available for download online), and a CSV file containing a single vector of values for each of the ARKit blendshapes. A complete tutorial of setting up Live Link Face in Unreal Engine and mapping CSV files to a MetaHuman can be found in Epic Games (2023a,b). The format of the CSV file should be identical to the format of the CSV files

<sup>3</sup>Vink (2022) also extensively discusses the use of MetaHuman avatars for sign language research. He, however, focuses on the use of such avatars for the purpose of text-to-sign translation, while we focus here on their use for the purpose of visualizing hypotheses and findings in scientific research.

extracted from the Live Link Face application.

Visualizing facial expressions in this manner can be extremely useful for research purposes. First, visualizations with a MetaHuman avatar provide readers with an immediate, nuanced understanding of the facial expression at hand, much more so than a textual description or a cartoon-like drawing. Further, compared to video stills, this method has the advantages that visualizations are anonymous, can take on any appearance, and are easily manipulated on a computer to illustrate the exact facial expression that a researcher may wish to convey. This saves the time required to find a suitable video frame for each facial expression one wants to visualize, and the images will never suffer from motion blur or occlusion. Moreover, this method could become a standardized format for the visualization of facial expressions. Authors could provide CSV files with blendshape values as auxiliary materials accompanying their publications, which would make it much easier to compare facial expressions described across different publications.

It is important to note that there is a somewhat steep learning curve for visualizing facial expressions using this method, especially if experience with software like Unreal Engine needs to be acquired first. However, there are many detailed (video) tutorials available online, alongside the official documentation.

## 5 Limitations

The new methods we have explored here are promising in that they facilitate more objective and quantitative investigation of facial expressions. However, they do have certain limitations. We focus here on limitations concerning the proposed method for data collection.

First, the angle and distance between the camera and the subject's face matter. In a preliminary study, we found that camera angle had a small yet significant effect on measurements (Esselink et al., 2023). We also found that the distance between the camera and the subject's face has a significant effect on the measurements, but smaller than the effect of angle. The results on the effect of distance have not been published yet.

Second, this method relies on proprietary software, implying a lack of control. The process involves an automatic but non-transparent mapping from raw measurements to blendshape values. The upside is that researchers do not need to implement this complex step themselves, meaning that the computation step will always involve exactly the same algorithm. This results in consistent, comparable output across experiments. However, downsides include lack of access to the raw data and the inability to control any parameters of the algorithm. Further, users do not receive detailed information about changes between versions of the internal software.

Finally, data collection is affected by (temporary) occlusions of the face. If a full occlusion of the face occurs, measurements resume fairly quickly – usually in less than five to ten frames. However, it takes a few more frames before the measurements are fully accurate again. This problem can be circumvented quite straightforwardly in practice: the initial frames directly following occlusion can be discarded and, if desired, missing measurements can be estimated through interpolation. Nevertheless, this does lead to information loss. A final note regarding occlusion: at this point we don't know to what extent glasses affect measurements. We expect that this can be problematic for certain blendshapes (e.g., EYEWIDE, EYESQUINT), although we have not tested this. All data in the case-study reported here came from participants without glasses.

## 6 Conclusion and future work

This paper has explored new methods for measuring, analyzing, and visualizing facial expressions. These methods are applicable in various research domains, including research on sign language, multi-modal communication, and emotions; but also in various practical domains, including the diagnosis of certain medical conditions, the development of embodied conversational agents, game characters, and virtual environments for training certain types of social or professional interaction. Traditional methods for measuring and analyzing facial expressions, at least those used in sign language research, suffer from subjectivity, reproducibility issues, and information loss due to transformation and compression. Recent methods making use of OpenFace address some of these issues but remain constrained in scope and reliant on 2D video data. We demonstrated the potential of capturing fine-grained 3D data by means of depth-sensing cameras. Our method directly measures relevant facial features, bypassing many of the limitations associated with both traditional methods involving manual annotation and recent methods based on OpenFace. We also discussed some machine learning techniques that can be used to analyze this type of data, and the benefits of visualizing the results of such analyses using MetaHuman avatars.

One important task for future work is to determine the extent to which blendshape measurements correspond to manual annotations of facial expressions. This issue will need to be explored using a parallel dataset of blendshape measurements and manually annotated video recordings.

## Acknowledgments

This work is part of the project *Questions in Sign Language* (grant number VI.C.201.014, PI Roelofsen) financed by the Dutch Science Foundation (NWO). We thank Tobias de Ronde, Marijke Scheffener, and the participants in the experiment reported here for their invaluable contributions to the data collection process. Furthermore, we thank Rob Bellemans, Martha Larson, Eric Nalisnick, and Daniël Vink for very helpful discussions about the data analysis and visualization techniques presented here.

## References

- Cecchetto, C. (2012). Sentence types. In *Sign Language*, pages 292–315. De Gruyter Mouton.
- Coerts, J. (1992). *Nonmanual Grammatical Markers: An Analysis of Interrogatives, Negations and Topicalisations in Sign Language of the Netherlands*. PhD thesis, Universiteit van Amsterdam.
- de Vos, C., van der Kooij, E., and Crasborn, O. (2009). Mixed signals: Combining linguistic and affective functions of eyebrows in questions in Sign Language of the Netherlands. *Language and speech*, 52(2-3):315–339.
- Epic Games (2022a). Live Link Face (Version 1.2.1). <https://apps.apple.com/us/app/live-link-face/id1495370836>.
- Epic Games (2022b). METAHUMAN - High-fidelity digital humans made easy. <https://www.unrealengine.com/en-US/metahuman>.

- Epic Games (2022c). Unreal engine. <https://www.unrealengine.com>.
- Epic Games (2023a). MetaHuman Documentation: Animating MetaHumans with Live Link in Unreal Engine. <https://dev.epicgames.com/documentation/en-us/metahuman/animating-metahumans-with-livelink-in-unreal-engine>. Accessed: 30/09/2023.
- Epic Games (2023b). Unreal Engine 5.1 Documentation: Recording Facial Animation from an iOS Device. <https://docs.unrealengine.com/5.1/en-US/recording-face-animation-on-ios-device-in-unreal-engine/#rawfacialrecordings>. Accessed: 30/09/2023.
- Esselink, L. (2023). Computer vision and machine learning for the analysis of non-manual markers in biased polar questions in Sign Language of the Netherlands. Master thesis, University of Amsterdam.
- Esselink, L., Oomen, M., and Roelofsen, F. (2023). Truedepth measurements of facial expressions: Sensitivity to the angle between camera and face. In *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, pages 1–5.
- Hartmann, T. J., Hartmann, J. B. J., Friebel-Hoffmann, U., Lato, C., Janni, W., and Lato, K. (2022). Novel method for three-dimensional facial expression recognition using self-normalizing neural networks and mobile devices. *Geburtshilfe und Frauenheilkunde*, 82(09):955–969.
- Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., and Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 109(19):7241–7244.
- Kimmelman, V., Imashev, A., Mukushev, M., and Sandygulova, A. (2020). Eyebrow position in grammatical and emotional expressions in Kazakh-Russian Sign Language: A quantitative study. *PloS one*, 15(6).
- Klomp, U. (2021). *A descriptive grammar of Sign Language of the Netherlands*. LOT.
- Kuznetsova, A., Imashev, A., Mukushev, M., Sandygulova, A., and Kimmelman, V. (2021). Using computer vision to analyze non-manual marking of questions in KRSL. In *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*, pages 49–59.
- Kuznetsova, A., Imashev, A., Mukushev, M., Sandygulova, A., and Kimmelman, V. (2022). Functional data analysis of non-manual marking of questions in Kazakh-Russian Sign Language. In *Proceedings of the 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*.
- Li, S. and Deng, W. (2022). Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*, 13(3):1195–1215.
- Oomen, M. and Roelofsen, F. (2022). Biased polar questions in Sign Language of the Netherlands - Methods description. [https://uvaauas.figshare.com/articles/preprint/Biased\\_polar\\_questions\\_in\\_Sign\\_Language\\_of\\_the\\_Netherlands\\_-\\_Methods\\_description/21701954](https://uvaauas.figshare.com/articles/preprint/Biased_polar_questions_in_Sign_Language_of_the_Netherlands_-_Methods_description/21701954).

- Oomen, M. and Roelofsen, F. (2023). Biased polar question forms in Sign Language of the Netherlands (NGT): Two functions of headshake. Forthcoming in the Proceedings of FEAST: Formal and Experimental Advances in Sign Language Theory.
- Pfau, R. and Quer, J. (2010). Nonmanuals: their grammatical and prosodic roles. In Brentari, D., editor, *Sign Languages*, pages 381–402. Cambridge University Press.
- Schermer, T. and Koolhof, C. (2009). *Basiswoordenboek Nederlandse Gebarentaal*. Van Dale.
- Steen, F. F., Hougaard, A., Joo, J., Olza, I., Cánovas, C. P., Pleshakova, A., Ray, S., Uhrig, P., Valenzuela, J., Woźny, J., and Turner, M. (2018). Toward an infrastructure for data-driven multimodal communication research. *Linguistics Vanguard*, 4(1):20170041.
- Vink, D. (2022). A modern approach to realistic facial animations in signing avatars – a qualitative assessment of the MetaHuman avatar. Master thesis, University of Amsterdam.
- Zeshan, U. (2004). Interrogative constructions in signed languages: Crosslinguistic perspectives. *Language*, pages 7–39.