

Functional Data Analysis of Non-manual Marking of Questions in Kazakh-Russian Sign Language

Anna Kuznetsova¹, Alfarabi Imashev², Medet Mukushev²,
Anara Sandygulova², Vadim Kimmelman³

¹Charles University, ²Nazarbayev University, ³University of Bergen
kuzannagood@gmail.com, vadim.kimmelman@uib.no
{alfarabi.imashev, mmukushev, anara.sandygulova}@nu.edu.kz

Abstract

This paper is a continuation of Kuznetsova et al. (2021), which described non-manual markers of polar and *wh*-questions in comparison with statements in an NLP dataset of Kazakh-Russian Sign Language (KRSL) using Computer Vision. One of the limitations of the previous work was the distortion of the 3D face landmarks when the head was rotated. The proposed solution was to train a simple linear regression model to predict the distortion and then subtract it from the original output. We improve this technique with a multilayer perceptron. Another limitation that we intend to address in this paper is the discrete analysis of the continuous movement of non-manuals. In Kuznetsova et al. (2021) we averaged the value of the non-manual over its scope for statistical analysis. To preserve information on the shape of the movement, in this study we use a statistical tool that is often used in speech research, Functional Data Analysis, specifically Functional PCA.

Keywords: non-manuals, Functional Data Analysis, Computer Vision

1. Introduction

In sign languages, besides hand signs, multiple non-manual markers are employed, such as body and head movements, movements of facial features and direction of the eye gaze (Pfau and Quer, 2010). These features can be linguistically significant, for instance, it is frequent for different types of questions to be marked only with non-manuals, leaving the manual signs and their order the same as in statements (Cecchetto, 2012).

In Kuznetsova et al. (2021) we provided the first description of some non-manual markers in Kazakh-Russian Sign Language (KRSL) based on a dataset that was collected for an NLP task. The material for that study was taken from Kimmelman et al. (2020) and comprised of video recordings of statements and questions produced by nine native signers of KRSL.

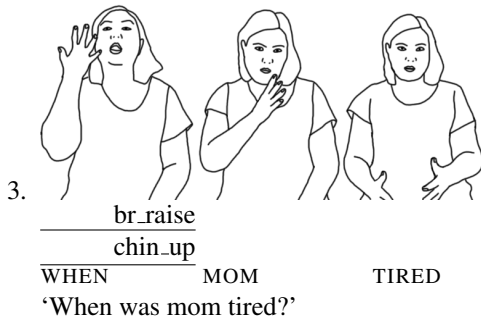
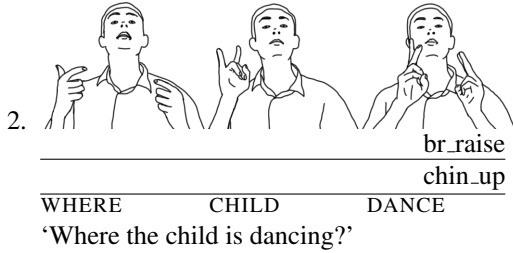
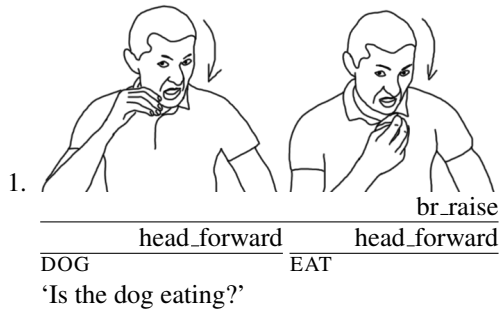
Research on sign language is usually not automated, meaning that linguists need to manually annotate material and make their observations subjectively. We tried to test whether this can be overcome with state-of-the-art Computer Vision tools in Kuznetsova et al. (2021). Using OpenFace (Baltrušaitis et al., 2018; Baltrušaitis et al., 2013; Zadeh et al., 2017) we were able to extract face landmarks in 3-dimensional space and use them to measure eyebrow movement and head rotation angle. However, we faced the model bias, which distorted the positions of the facial landmarks with the change of the head rotation angle (see Section 2.2). Our solution was to train a simple linear regression model to predict this bias and then subtract it from the initial results of the OpenFace. We achieved relatively stable data and statistically analyzed it using a mixed-effects multivariate linear regression model. However, our analysis was not on continuous data of the movements but on discrete points that represented the mean value of the feature

over the duration of the movement. The results suggest that in our KRSL dataset polar questions are marked by eyebrow raise on the whole sentence, and consecutive forward head tilts on the subject and verb (see example 1). On the other hand, *wh*-questions are marked by backward head tilts on the *wh*-sign, and by eyebrow raise on the *wh*-sign that can spread over the whole sentence (see examples 2, 3).

Based on these prior results, the goals of this study are the following. Firstly, we will try to improve on our bias detection model. Secondly, we will use Functional Data Analysis to analyze continuous movement of the eyebrows and head. We hope that this work will be helpful to linguists who also want to study non-manual movements in other languages because we believe that our approach can be extended to other datasets. We share the code with a step-by-step instruction on <https://github.com/kuzanna2016/non-manuals-2021>.

2. Data Extraction and Correction

For the current study, we used the same video clips and annotations as in Kuznetsova et al. (2021). The data contains recordings of 10 simple sentences with a subject and an intransitive verb, each in three forms – statement, polar question and *wh*-question (for example, the signed versions of “The dog is eating.”, “Is the dog eating?” and “Where is the dog eating?”). At the beginning of the *wh*-questions, there is also a *wh*-sign. The sentences were produced by nine native KRSL signers, 5 deaf signers and 4 are hearing children of deaf adults (CODAs) currently working as KRSL interpreters. In total we have 270 videoclips.



2.1. Face Landmark Extraction

We firstly needed to extract face landmarks from the video clips. We use the same method as in Kuznetsova et al. (2021) – OpenFace (Baltrušaitis et al., 2018; Baltrušaitis et al., 2013; Zadeh et al., 2017). OpenFace outputs face landmarks location in 3d space in millimetres, the location of the head with respect to the camera in millimetres, the head rotation in radians around three axes, which can be interpreted as pitch (Rx), yaw (Ry), and roll (Rz) and a confidence score from 0 to 1 for the whole frame. Only 103 frames from 12 videos had a low confidence score (< 0.8); we did not use those frames and filled in the neighbouring frames' values. The next step in the analysis is to calculate the eyebrow distances. In Kuznetsova et al. (2021) the distance between the eyebrow points and the eye line was used. The main reason for that was that this distance is the most intuitively interpretable as the eyebrow movement is mostly vertical. We also tried other distances – distance to the upper nose point (27),¹ distance to a horizontal plane, but they did not work as well, so we will not discuss them. For distance calculations we used the following eyebrow points: outer left eyebrow – 18, inner left eyebrow – 20, inner right eyebrow – 23, outer right eyebrow – 25.

¹The numbers correspond to the numbers used in OpenPose's output files.

2.2. Correction Model

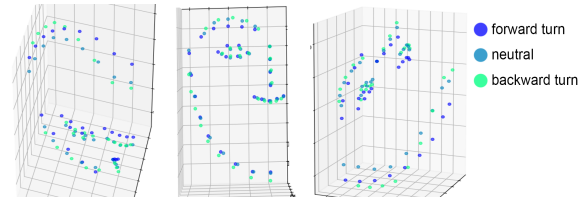


Figure 1: The behaviour of keypoints with different head turns on the test video.

As already stated in Kuznetsova et al. (2021), we found out that the OpenFace model has a rotation bias in 3d face landmarks detection. This means that the location of the points distorts with the head rotation: for instance, the eyebrows become more rounded in the backward tilt and more flat in the forward tilt (we examined this behaviour in the test video, where we recorded the head movement from the low to high pitch without the eyebrow movement, see Figure 1). We tried to eliminate this distortion using different geometrical techniques, but in the end we decided to switch to machine learning tools. The model should learn the bias distortion from the frames without eyebrow movement, then this bias can be predicted for all the frames and later subtracted from the initial distance. In Kuznetsova et al. (2021) we used a simple linear regression model to predict this bias. The training data was from the statements, specifically the manually selected videos where no eyebrow movement is present (63 sentences in total, 4414 frames). Our choice of the model was based on the observation that the distortion seems to be linear and consistent across signers – Pearson correlation coefficient between vertical head angle and the eyebrow distance to the eye line in sentences with no eyebrow raise is -0.39 for the inner distance and -0.4 for the outer distance.

This time we tried to improve the bias prediction by using a more advanced model, specifically multilayer perceptron. We believe it is sufficient for our task: it is not a deep model, can handle a moderate number of samples without overfitting and it can also capture some nonlinear dependencies. We performed hyperparameter search using cross-validation on 4 folds (test size – 25%, 1104 frames, train size – 75%, 3310 frames). The input features were the rotation angles of the head in three dimensions (pose_Rx, pose_Ry, pose_Rz in OpenFace), the cosine of the head rotation angles, the location of the head (pose_Tx, pose_Ty, pose_Tz in OpenFace), the one-hot encoded sentence and signer features. As previously mentioned in Kuznetsova et al. (2021) the big increase in quality is mostly attributed to the addition of the signer features, as the model learns individual parameters of the face of the signer. This set of features thus makes the model only applicable to our dataset and we encourage the researchers to retrain their models if they want to use our method.

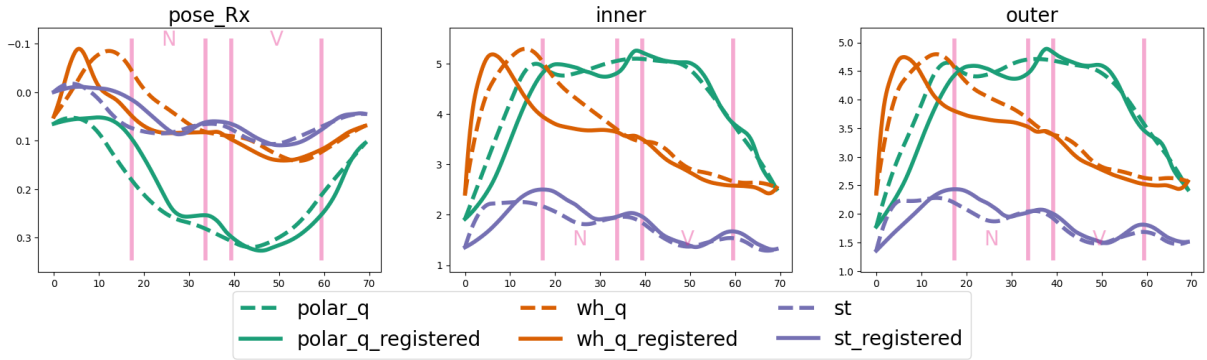


Figure 2: The mean curves of the sentence types before and after landmark registration. The red lines represent the boundaries of the hand signs.

For the experiment we used the *sklearn* library for Python (Pedregosa et al., 2018). The baseline model is the linear regression model from Kuznetsova et al. (2021) with L2 regularization (*sklearn.linear_model.Ridge*) and the examined model is Multi-layer Perceptron regression (*sklearn.neural_network.MLPRegressor*). The inner and outer eyebrow distances were predicted simultaneously.

The best result was achieved by the *MLPRegressor* with hidden layer size 40 – combined MSE score for inner and outer distances was 0.38, which improved on the baseline score of 1.45 for inner eyebrows and 1.36 for outer eyebrows. The best score of the model without the sentence and speaker features was 3.2 (the model had hidden layer sizes 45 and 40), which is also an improvement from the baseline score of 4 but is still significantly worse than the model with individual features.

As before, we used the trained model to predict the “default” eyebrow distance for all frames and then subtracted it from the originally computed distance.

3. Functional Data Analysis

Eyebrow movement and head rotation angle are dynamic features, therefore we want to analyse them as continuous data rather than discrete, as we did in Kuznetsova et al. (2021). In Gubian et al. (2009) Functional Data Analysis (FDA) was introduced as a tool to analyze dynamic transitions in speech signals. FDA provides the means to analyze continuous functional data like classic statistical methods analyze scalars (Ramsay and Silverman, 2002). Our main focus will be on functional principal component analysis (fPCA) – a tool that converts functional data into a scalar representation with minimum information loss. Our analysis is described by the following algorithm. Firstly, time measurements need to be transformed into function form. This can be done by using basis functions like B-splines and standard least-squares interpolation with a regularization term. Functions are normal-

ized so that all observations have the same duration – to compare them across time. It is also possible to align functions on the landmarks – so that events in all observations coincide in time. In our case, the landmarks are the start and end frames of the hand signs. After the data preparation, fPCA, which finds a representation of the data with a smaller dimension size saving the variation. Principal components can afterwards be interpreted and analyzed with classical statistical methods, like mixed-effect multivariate linear regression. fPCA eliminates the problem of manually picking the scalar features from the dataset – in Kuznetsova et al. (2021) it was the mean across the manual signs and with fPCA we will be able to take into consideration the whole contour. In our analysis we use the *scikit-fda* library for Python (Carreño et al., 2022).

3.1. Data Preprocessing

The first step to FDA is to turn raw data points into continuous functional data. This is done by the combination of the set of functions. In our case, the most applicable set of functions is B-splines (de Boor, 1978) as the data is not periodic and can vary in shape greatly. Our data is quite noisy, therefore we do not want the function to approximate our data ideally, we want a smooth representation. This can be done by adjusting the numbers of the functions in the combination – the number of “hills” by the regularization term and by the order of the B-spline. When fitting the curves to the data we can compute the fitting error and try to minimize it when choosing the hyperparameters, however, visual inspection is still a valuable step. Based on both methods, we set the number of basis functions at 14 and the order of functions at 3, because it smoothes the data enough, saving the important features.

We want to align our functions on the start and end of the hand signs because we need to determine which constituent is marked by the non-manual and because we have different numbers of signs: there is an additional wh-word sign at the beginning of the wh-questions. We extracted the boundaries of the signs

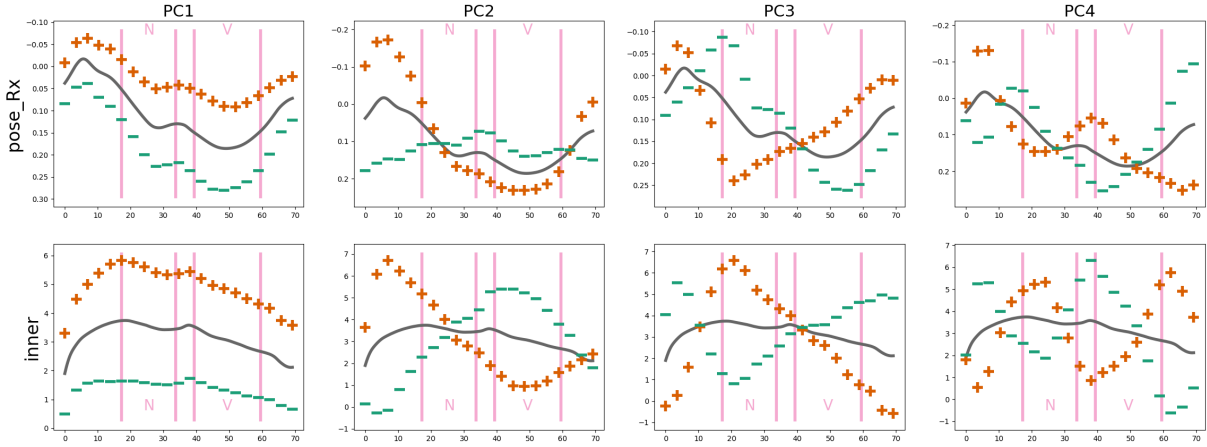


Figure 3: The perturbation graphs for the top 4 principal components. The solid curve is the mean of the dataset. Lines with the ‘+’ sign are the curves where the principal component was added to the mean and lines with the ‘-’ sign are the curves where the principal component was subtracted from the mean. The weight of the principal component is equal to the standard deviation of the dataset weights for that principal component.

from the manual annotation and we aligned them to the mean of those boundaries across all sentences (17.27 – the start of the noun, 33.71 – the end of the noun, 39.33 – the start of the verb, 59.46 – the end of the verb). The importance of landmark registration is described in the document entitled Time normalisation and landmark registration in the additional material from Gubian et al. (2015)². In the analysis of formant curves the authors claim that although overall non-registered results go in the same direction with the registered results, the effectiveness of the obtained principal components (see Section 3.2 on fPCA) decreased. The principal components from non-registered data described less variance and tried to incorporate the boundaries information which can be explicitly done with landmark registration.

The effect of the landmark registration can be seen in Figure 2 where the mean of each sentence type is plotted before and after registration. The peaks of the wh-questions have been moved to the left, which reflects the position of the wh-sign at the beginning of the sentence, while polar questions and statements have been slightly moved to the right as the mean positions of the hand signs are influenced by the wh-questions and are skewed to the right. Moreover, the peaks in all sentence types became more pronounced as they became more aligned. Moreover, it is clear from the figure that inner and outer eyebrow movement do not differ much, so we will not discuss outer eyebrow movement separately.

3.2. Functional PCA

With registered and smoothed data we can perform fPCA. One of the applications of PCA is dimension-

ality reduction. PCA provides principal components (usually vectors) and their weights for each data point so that the sum of the dataset mean and the weighted sum of the principal components will reconstruct the data point. For data point x_i the formula

$$mean + s_i^1 * PC_1 + s_i^2 * PC_2 .. + s_i^n * PC_n,$$

where s_i^n – is the score of the n th principal component for that data point and PC_n is the n th principal component, will produce the best approximation of x_i . Principal components are ranked from the most informative to least, so the first principal component will capture the biggest variance in the dataset. This feature is the reason why PCA is used in dimensionality reduction: using only some of the first principle components the data can be expressed with some percent of the saved variance. Functional PCA has the same output but principal components are in function form (Jolliffe and Jackson, 1993). Functional principal components are modifying functions that work like the regular principal components. To reconstruct a function from the dataset we need to add functional principal components multiplied by their weights to the mean curve. We performed fPCA independently on our three features. The first four principal components explain 93-96% of the variance (Table 1).

	PC1	PC2	PC3	PC4	Total
head rotation	69%	14%	6%	4%	93%
inner brows	83%	7%	4%	2%	96%

Table 1: The explained variance ratios of the principal components.

Functional principal components are modifiers of the mean curve; therefore the best way to look at them and interpret them are perturbation graphs (Figure 3). The

²<https://github.com/uasolo/FDA-DH/blob/master/paper/TimeRegistration.pdf>

perturbations are defined as variations over the mean: we add (lines with ‘+’ sign) and subtract (lines with the ‘-’ sign) each principal component from the mean curve (the solid line) with the weight equal to the standard deviation of the dataset weights for that principal component. We can interpret these lines as the borderline cases of the principal component modification.

In Figure 3 we can see that PC1 mainly alters the amplitude of the movement and to some extent the bulge of the curve both in eyebrow and head rotation cases. Next we will explore the eyebrow movement components. PC2 seems to distinguish between curves that have the eyebrow raise before the noun and the curves which have the eyebrow raise on the verb. PC3 acts as a separator between curves with one main raise on the noun and gradual decline to the end of the sentence and curves with slight raise before the noun and a plunge on the noun. The last component is more complicated with more than one peak, it will be harder to interpret it correctly. Still, PC4 either has a raise on the noun and a slightly lower raise at the end of the verb or two raises: one before the noun and one before the verb. As for the head movement, PC2 distinguishes between an almost flat movement with a small bump between noun and verb and a raise before the noun with a deep plunge on noun and verb. PC3 has either a raise before the noun and a plunge until the end of the sentence or a raise on the noun and a decline towards the verb with a small hump between the noun and the verb. Finally, PC4 has very subtle differences and the least amplitude of the changes: it separates the high rise before the noun from the small rise on the noun and a big hump between the noun and the verb and a more smooth hump there.

3.3. Statistical Analysis

In the previous section we obtained valuable discrete features for all sentences – scores of the principal components, which we can analyse with the standard statistical tools. We will repeat the analysis made in Kuznetsova et al. (2021) with some alterations. The analysis is made in R. The model that we are using is a mixed-effects multivariate linear regression (Baayen et al., 2008; Bates et al., 2015). The fixed predictor variables for the model are sentence type (categorical, three levels: statement, polar question, wh-question), group (categorical, deaf vs. hearing), and all the interactions between the two predictors. The random variables are participant (with a random slope for sentence type or part of sentence), and sentence (with a random slope for the group). We also use the *lme4* package (Chung et al., 2015) with the help of the *blme* package (Chung et al., 2013) to achieve convergence with a small number of levels for the random effects. The significance of the group feature was calculated with the *ANOVA* function from the *car* package (Fox and Weisberg, 2019). We have three levels in our sentence type feature, therefore we would need to test three hypotheses and account for the multiple comparison problem. In Kuznetsova et al.

(2021) we overcame this problem with the orthogonal contrast: we compared statements with wh-questions and the mean of the statements and wh-questions with the polar questions. The features were the distances and the concept of the mean of the distances is intuitive, however, when the features are principal components, the mean of the principal components is more complicated. That is why we decided to make a more complicated analysis with a pairwise comparison. We use the *multcomp* package (Hothorn et al., 2008) to do this. We used Tukey Contrasts and the p-values were adjusted with the single-step method (Bretz et al., 2016). We made separate models for the inner eyebrow distance, for the outer eyebrow distance and for the vertical head rotation angle, and for each principal component, producing a total of 15 models. The result of the models is discussed in Section 4.

4. Results

4.1. Eyebrow Movement

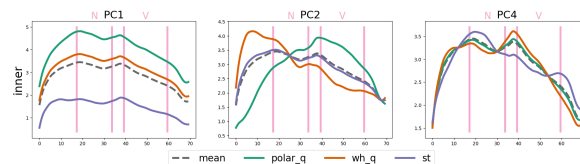


Figure 4: Mean curves of the eyebrow movement for each sentence type reconstructed with the significant principal components separately.

From visual inspection of the mean curves (Figure 2) we come to the same conclusion as in Kuznetsova et al. (2021): polar questions are marked by the eyebrow raise on the noun and verb with some nods in-between, while wh-questions are marked by the eyebrow raise on the wh-sign at the beginning of the sentence and gentle eyebrows lowering to the end of the sentence. Statements have some eyebrow movement but the amplitude is much lower and it may be the effect of the inconsistency of marking across signers. We will report only on the significant features; the full results of the statistical analysis can be found with the code.

The first principal component has a significant impact in distinguishing between polar questions and statements: in inner and outer eyebrows the p-value is < 0.001 ; and wh-questions and statements: in inner eyebrows the p-value is 0.0498. The mean PC1 score for the polar questions is 9.74 for the inner eyebrows and 6.94 for the outer eyebrows, while for the statements it is -11.65 for inner and -9.27 for outer and for the wh-questions it is 2.54 and 2.78 respectively. According to the shape of the perturbation graph (Figure 4), polar questions have a big amplitude raise and statements have a low eyebrow raise with a flatter curve, while wh-questions are close to the mean.

The second principal component is also significant, but for the distinction between the polar and wh-questions. For both the inner and outer eyebrows the p-value is < 0.001 . The mean PC2 score for the polar questions is -6.06 and -4.79, and for the wh-questions it is 4.83 and 3.78 for inner and outer eyebrows respectively. Polar questions thus have a more gentle raise to the verb and wh-questions have a sharp raise before the noun, on the wh-sign (Figure 4).

The fourth principal component also has a significant impact, but the least one. It distinguishes between the wh-questions and statements. For the inner eyebrows the p-value is 0.0501 and for the outer it is 0.0273. The mean PC4 score for the wh-questions is 1.23 for the inner eyebrows and -1.37 for the outer, and for the statements it is -1.53 and 1.28. In the Figure 4 it is a very subtle difference, statements deviate slightly from the mean curve in three positions, on the sign boundaries, while wh-questions have a more pronounced deviation in the beginning, on the wh-sign, and a raise before the verb.

Thus, we confirm the previous observations that polar questions are marked by eyebrow raise on the noun and verb, while wh-questions are marked by eyebrow raise at the beginning of the sentence on the wh-sign.

4.2. Head Movement

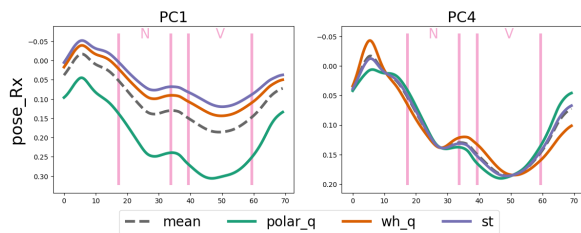


Figure 5: Mean curves of the vertical head rotation for each sentence type reconstructed with the significant principal components separately.

From visual inspection of the data, wh-questions seem to be marked with the backward tilt on the wh-sign, polar questions have a forward tilt on the noun and verb, and statements have small movements that resembles quick nods on the noun and the verb.

The statistical analysis shows that the first principle component significantly impacts the separation between wh-questions and polar questions (p-value < 0.00291) and statements and polar questions (p-value 0.0016). The mean score of the first component for wh-questions is 0.3, for polar questions -0.82 , for statements 0.47, which means that polar questions have a deep forward tilt on the sentence peaking at the noun and verb, while wh-questions and statements have a more flattened movement (Figure 5, the first column). The next significant principal component is the fourth principal component. Statement and wh-questions dif-

fer significantly (p-value is 0.00229) and so do wh-questions and polar questions (p-value is 0.02667). The mean scores of the fourth principal component are -0.1 for polar questions, 0.12 for wh-questions and -0.02 for statements. According to the perturbation graph (Figure 5, the second column), this means that wh-questions have a pronounced backward tilt at the beginning of the sentence on the wh-sign, and a nod between the noun and the verb, while statements and polar questions do not have a backward tilt in the beginning. We come to the same conclusion that the polar questions are marked with a continuous forward tilt on the noun and verb and the wh-questions are marked with a backward tilt on the wh-sign.

4.3. Deaf/hearing Differences

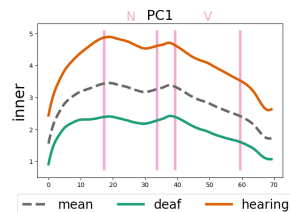


Figure 6: Mean curves of the eyebrow movement for the sentences with deaf and hearing signers reconstructed with the significant principal components separately.

In Kuznetsova et al. (2021) we did not find any statistically significant differences between the deaf and hearing signers. This time we can report that there are differences in some principal components.

The eyebrow movement has shown some significant differences in the first principal component for both the inner eyebrows and the outer eyebrows (p-values are 0.02764 and 0.03632 respectively). The first component mean scores for the inner eyebrows are -7.49 for the deaf signers and 10.31 for the hearing signers, for the outer eyebrows -8.85 and 12.18. Figure 6 shows that in the first component the hearing signers tend to have higher eyebrow raise than the deaf signers.

5. Discussion

5.1. FDA and Sign Languages

The main source of the Functional Data Analysis techniques for this study was the website hosted by Michele Gubian. In his works, Gubian explores how FDA can be applied to speech research; however, he points out that FDA can be applied to other types of uni- or multi-dimensional continuous signals. We took inspiration from this and were able to translate his approach to sign language prosody. We believe that FDA has significantly improved our analysis. Firstly, we were able to analyse sentences with different durations and different number of signs with landmark registration. Secondly, with fPCA we were able to take into account the

whole sentence contour, rather than some handpicked features. The principal components that we obtained were interpretable and it was easy to explore the visualisations. We hope that our work will increase interest in applying computer vision tools and FDA to sign language data. Section 5.2 has some practical advice to those who would like to try this approach.

5.2. Applying to Naturalistic Data

This study was made with the materials that were collected for NLP tasks in a constrained way and with a small number of signers. Moreover, almost half of the signers were hearing children of deaf adults, which means that the sample was not homogeneous, which is reflected in the differences between deaf and hearing signers. This makes our dataset far from naturalistic and we cannot guarantee that this approach will work on naturalistic data.

However, we believe that it is still possible and we encourage researchers to test it. We suggest finding materials in corpus where sign boundaries and non-manuals are already annotated. Various non-manual markers can be obtained with OpenFace, including head rotation in three axes, head movement, eye aperture, eye gaze, mouthing and eyebrow movements. We advise to obtain the frames with no non-manual markers from the same materials and same signers to use in the correction model, if the non-manuals in question can be modified by the head rotation. When using the correction model the id of the material and the id of the signer should be used as categorical features (like we used the sentence id and the signer id). The following analysis can be done with FDA or another framework, depending on the aims of the study. Lastly, we recommend inspecting frames with low confidence scores from OpenFace as they can damage the results of the correction model and the subsequent analysis. Frames with low confidence scores should not be included in the correction model training set, but they can be used in other steps if their values are filled in by the neighboring values or the mean of the neighboring values.

5.3. Data Manipulation

We understand that our approach of correcting the OpenFace results can introduce unwanted noise to the data and it would be more reliable to modify the predictor. The approach of putting a correction model on top of the predictor is indirect and subjective, as the features that we use only reflect our empirical observations, while the predictor has important internal states that can directly solve the problem. Although OpenFace is a state of the art tool the problem of general 3D reconstruction from a single camera is challenging, especially when the camera is not constrained, and the reconstructed 3D shape is not always going to be accurate and will be affected by rotation up to a point. We did not try other models that can perform 3D reconstruction of the face landmarks and did not modify the original model. We also did not retrain it on our

data because we do not have the resources to annotate it for this task. If there are other solutions, we would encourage to try them out in subsequent research.

5.4. Availability of the Code

We produced a script which captures all elements of the data preparation, including the bias detection model, Functional Data Analysis and statistical analysis for further research on non-manual markers in sign languages. The script is freely available on GitHub with a step-by-step instruction: <https://github.com/kuzanna2016/non-manuals-2021>.

6. Conclusions

In this study, we (1) re-tested and improved techniques for eyebrow distance extraction using computer vision tools and (2) introduced FDA as a tool to analyse dynamic shapes of non-manuals. We supported the conclusions about the non-manual marking of the questions in the KRSL dataset from Kuznetsova et al. (2021) with the new analysis. In the KRSL dataset the wh-questions are marked with a backward head tilt and an eyebrow raise on the wh-word while polar questions are marked with a forward head tilt and an eyebrow raise on the noun and verb. We also found a difference between the deaf and hearing signers: the hearing signers tend to have more expressive non-manuals, meaning that the manuals have a bigger amplitude and the features are more pronounced.

Furthermore, this study demonstrates that computer vision techniques can be applied for sign language linguistic research, specifically research on non-manuals. Although these tools are very useful, they also have limitations. For example, the OpenFace model distorts face landmarks when the head is rotated. We have found one solution to this problem. We train an additional model on top of the predicted results to predict the errors and then we subtract the errors from the OpenFace output.

Moreover, we used a new statistical tool for linguistic analysis: Functional Data Analysis. It was already proven to be a great tool for spoken language phonetics and this study provides evidence that it can also be used for sign language prosody. FDA provides a way to work with continuous data, to shift curves and to extract features from these curves using functional principal component analysis. The translation of continuous data into scalar points helps analyse this data with standard statistical procedures.

We hope that our research will be useful in solving the problem of quantitative analysis of sign language linguistic features.

7. Bibliographical References

- Baayen, H., Davidson, D., and Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59:390–412, 11.

- Baltrušaitis, T., Robinson, P., and Morency, L.-P. (2013). Constrained local neural fields for robust facial landmark detection in the wild. In *2013 IEEE International Conference on Computer Vision Workshops*, pages 354–361.
- Baltrušaitis, T., Zadeh, A., Lim, Y. C., and Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pages 59–66.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.
- Bretz, F., Hothorn, T., and Westfall, P. (2016). *Multiple Comparisons Using R*. Chapman and Hall/CRC.
- Carreño, C. R., Suárez, A., Torrecilla, J. L., Berrocal, M. C., Manchón, P. M., Manso, P. P., Bernabé, A. H., Fernández, D. G., Yujian Hong, and Rodríguez-Ponga Eyriès, P. M. (2022). Gaa-uam/scikit-fda: Version 0.5. <https://doi.org/10.5281/zenodo.4406983>.
- Cecchetto, C. (2012). Sentence types. In Roland Pfau, et al., editors, *Sign language: An international handbook*, pages 292–315. De Gruyter Mouton.
- Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., and Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multi-level models. *Psychometrika*, 78:685–709, 10.
- Chung, Y., Gelman, A., Rabe-Hesketh, S., Liu, J., and Dorie, V. (2015). Weakly informative prior for point estimation of covariance matrices in hierarchical models. *Journal of Educational and Behavioral Statistics*, 40:136–157, 04.
- de Boor, C. (1978). *A Practical Guide to Splines*. Springer New York.
- Fox, J. and Weisberg, S. (2019). *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, third edition.
- Gubian, M., Torreira, F., Strik, H., and Boves, L. (2009). Functional data analysis as a tool for analyzing speech dynamics - a case study on the french word c'était. In *Interspeech 2009*. ISCA, September.
- Gubian, M., Torreira, F., and Boves, L. (2015). Using functional data analysis for investigating multi-dimensional dynamic phonetic contrasts. *Journal of Phonetics*, 49:16–40, March.
- Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3):346–363, June.
- Jolliffe, I. T. and Jackson, J. E. (1993). A user's guide to principal components. *The Statistician*, 42:76–77.
- 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):e0233731, June.
- 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, Virtual, August. Association for Machine Translation in the Americas.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2018). Scikit-learn: Machine learning in python.
- Pfau, R. and Quer, J. (2010). Nonmanuals: their prosodic and grammatical roles. In Diane Brentari, editor, *Sign Languages*, pages 381–402. Cambridge University Press.
- Ramsay, J. and Silverman, B. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer New York.
- Zadeh, A., Baltrušaitis, T., and Morency, L.-P. (2017). Convolutional experts constrained local model for facial landmark detection. *ArXiv e-prints*, arXiv:1611.08657.