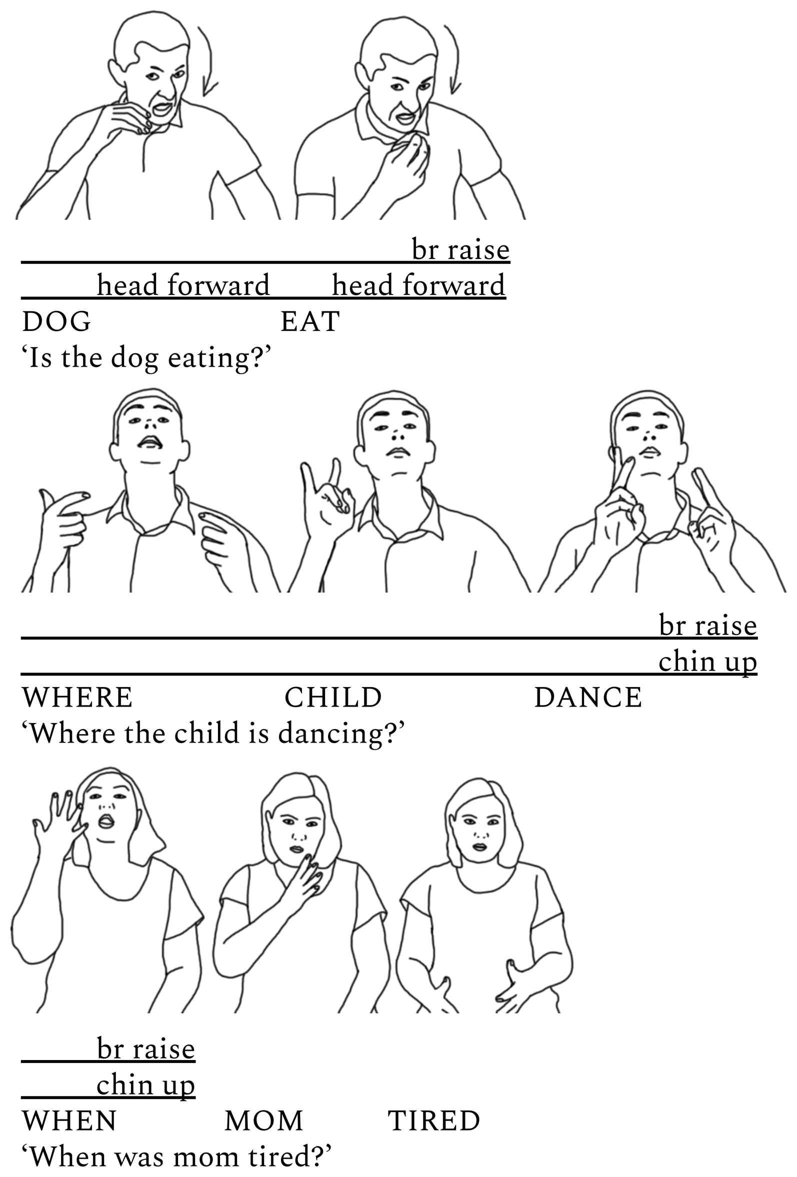
Functional Data Analysis of Non-manual Marking of Questions in Kazakh-Russian Sign Language

Anna Kuznetsova, Alfarabi Imashev, Medet Mukushev, Anara Sandygulova, Vadim Kimmelman

1. Data

- 10 simple sentences with a subject and an intransitive verb, each in three forms – statement, polar question and wh-question
- Produced by 9 native KRSL signers, 5 deaf signers and 4 hearing children of deaf adults (CODAs) currently working as KRSL interpreters
- Dataset was initially collected for an NLP task
- The sentence were originally created in written Russian, and translated to KRSL by a native hearing KRSL signer with neutral emotion; the translations were recorded to be used as stimuli
- 270 videoclips in total



2. Face landmarks extraction

Face landmarks were extracted from the videoclips using OpenFace.
Output of the Openface (Fig.1):

- 3d coordinates 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)
- a confidence score from 0 to 1 for the whole frame

Only 103 frames from 12 videos had a low confidence score (< 0.8).
Eyebrow raise was calculated as a distance between the eyebrow points and the eye line (Fig. 2).

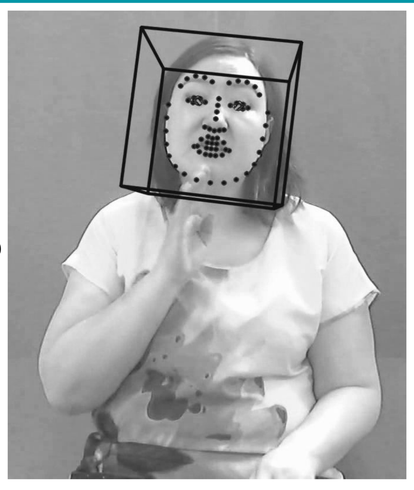


Fig. 1

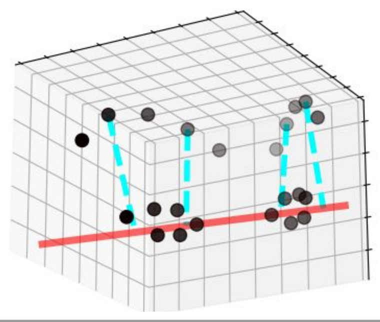


Fig. 2

3. Correction Model

OpenFace model has a rotation bias in 3d face landmarks detection (Fig. 3). Our solution is to use a machine learning model to predict the biased distances and subtract them from the OpenFace output.

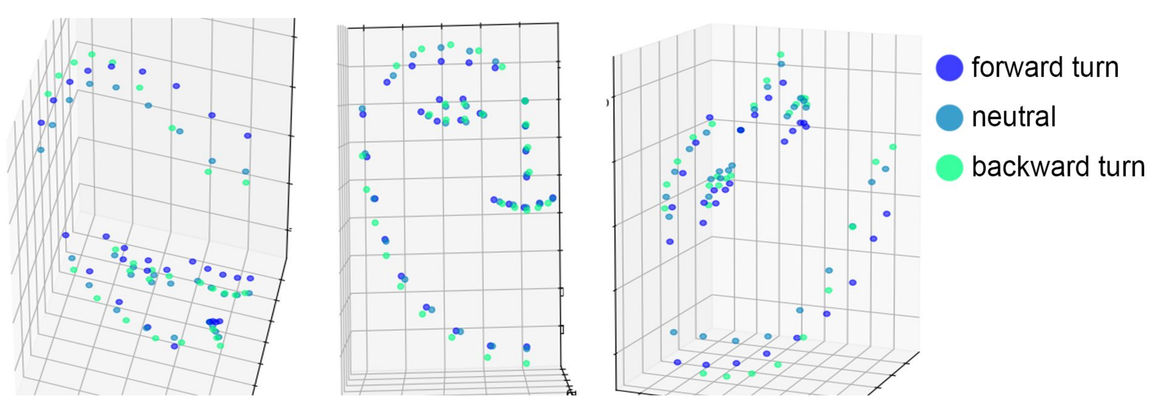


Fig. 3. The behaviour of keypoints with different head turns.

Training data:

- statements, specifically the manually selected videos where no eyebrow movement is present (63 sentences in total, 4414 frames),
- cross-validation on 4 folds (test size – 25%, 1104 frames, train size – 75%, 3310 frames)

Input features:

- the rotation angles of the head in three dimensions (pose Rx, pose Ry, pose Ry 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 (personal) sentence and signer features

Target was the vertical eyebrow distance.
Scores:

Model	MSE	MSE (no personal features)
Linear regression (Kuznetsova et al., 2021)	1.4	4
Multilayer Perceptron	0.38	3.2

6. Discussion

- FDA and Sign Languages
  - FDA helps analyse sentences with different durations and different number of signs with landmark registration
  - fPCA provides the mean to use the whole sentence contour, rather than some handpicked features in statistical analysis
  - PCs are interpretable and easy to visualise
- Applying to Naturalistic Data
  - Materials for this study were collected for NLP tasks in a constrained way and with a small number of signers => We encourage to try this approach on more naturalistic data
- Data Manipulation
  - Correction model is not ideal, it is better to change the OpenFace model itself, which we couldn't do this time
- Availability of the Code
  - https://github.com/kuzanna2016/non-manuals-2021.

4. Functional Data Analysis

FDA provides the means to analyze continuous functional data like classic statistical methods analyze scalars. In Gubian et al. (2009) FDA was introduced as a tool to analyze dynamic transitions in speech signals. We want to show that we can apply FDA to non-manuals because they are also dynamic features.

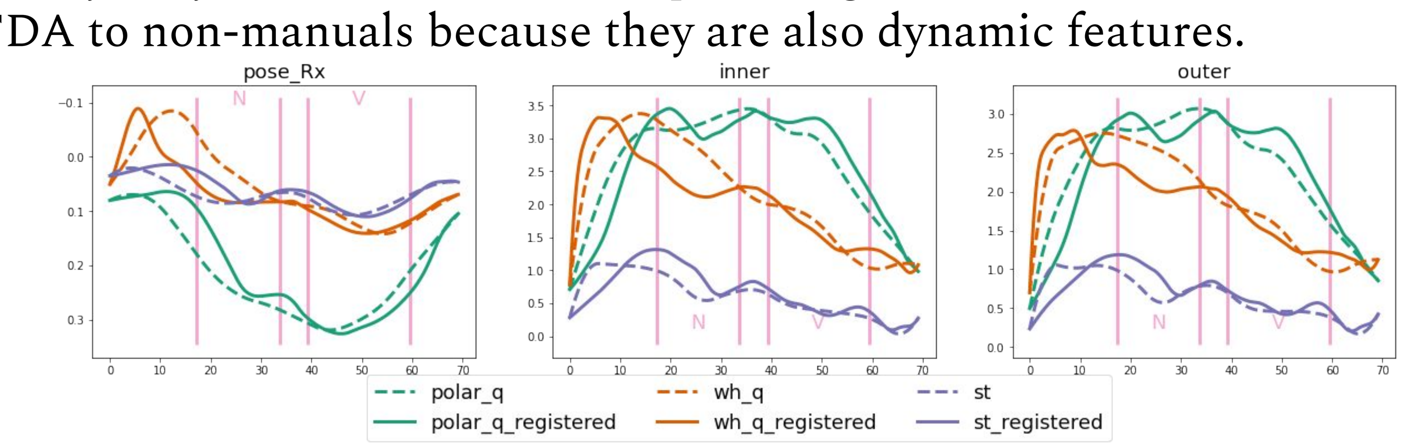


Fig. 4. The mean curves of the sentence types before and after landmark registration. The contours become more clear after landmark registration, see e.g. the two pronounced peaks for head rotation in polar questions and the shift of the peak before the noun for wh-questions.

Analysis steps:

1. Transform time measurements into function form using B-splines and standard least-squares interpolation with a regularization term.
2. Normalize functions so that all observations have the same duration
3. Align functions on the landmarks - the start and end frames of the hand signs (Fig. 4).
4. Perform fPCA (Fig. 5). fPCA 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. The first four principal components explain 93-96% of the variance.

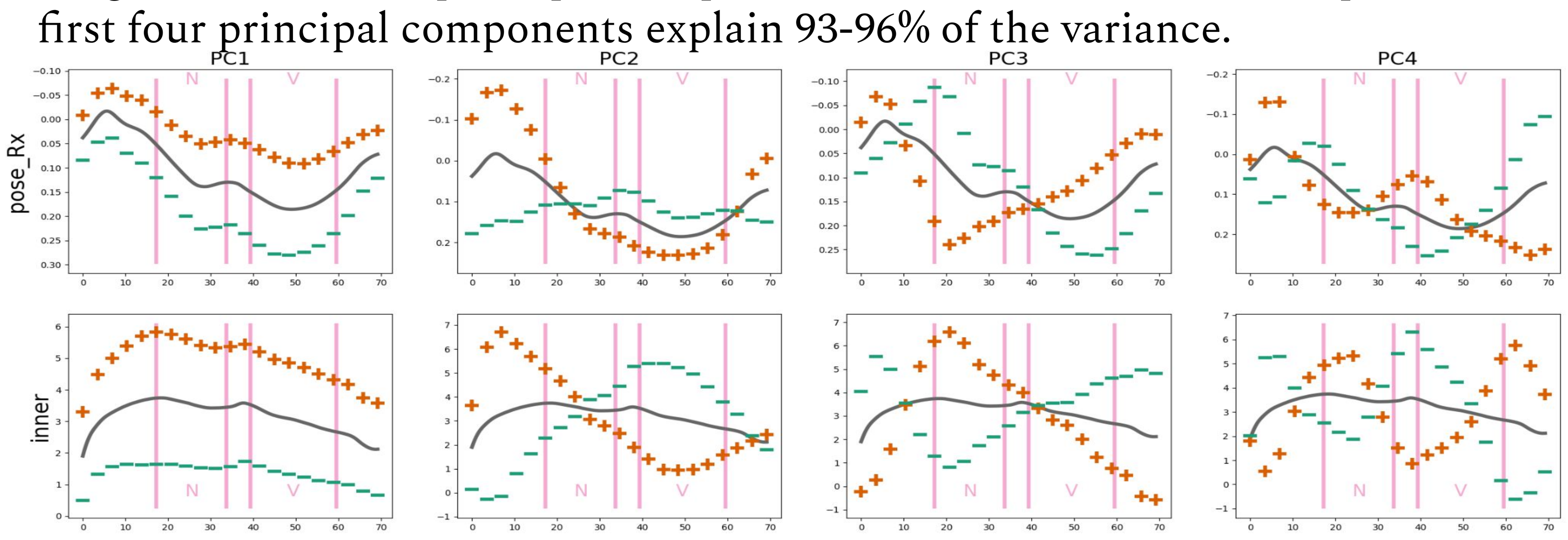


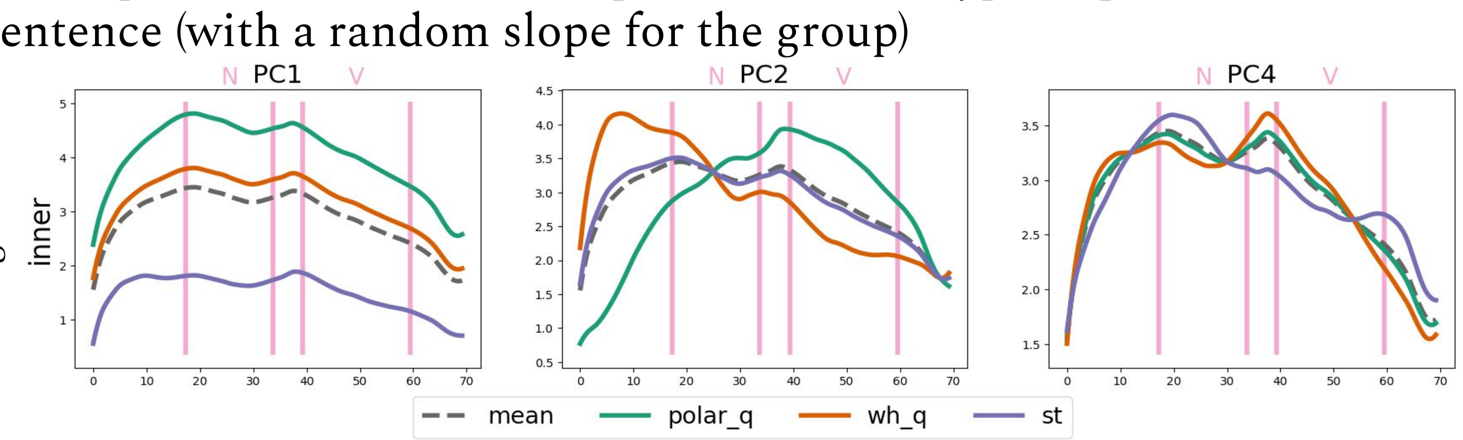
Fig. 5. The perturbation graphs for the top 4 principal components. The solid curve is the mean of the dataset. Lines with the '+' sign are the curves where the principal component was added to the mean and lines with the '-' sign are the curves where the principal component was subtracted from the mean. The weight of the principal component is equal to the standard deviation of the dataset weights for that principal component.

5. Statistical Analysis Results

Mixed-effects multivariate linear regression:

- The fixed predictor variables:
  - sentence type (categorical, three levels: statement, polar question, wh-question)
  - group (categorical, deaf vs. hearing)
  - all the interactions between the two predictors
- The random variables:
  - participant (with a random slope for sentence type or part of sentence)
  - sentence (with a random slope for the group)

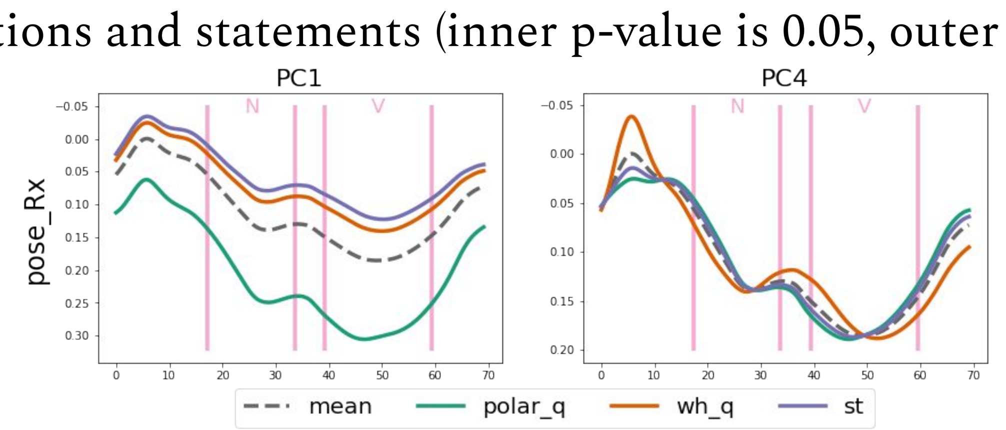
Fig. 6



Significant PCs for the eyebrow movement (Fig. 6):

- PC1 - amplitude:
  - statements and polar questions (p-value < 0.001)
  - statements and wh-questions (p-value is 0.0498)
- PC2 - raise on the verb vs raise before the noun
  - polar and wh-questions (p-value < 0.001)
- PC4 - pronounced raise before the noun, and a raise before the verb
  - wh-questions and statements (inner p-value is 0.05, outer p-value is 0.0273)

Fig. 7



Significant PCs for the head movement (Fig. 7):

- PC1 - a deep forward tilt on the sentence peaking at the noun and verb
  - polar questions and wh-questions (p-value < 0.00291)
  - polar questions and statements (p-value is 0.0016)
- PC4 - a pronounced backward tilt at the beginning of the sentence on the wh-sign, and a nod between the noun and the verb
  - wh-questions and statements (p-value is 0.00229)
  - wh-questions and polar questions (p-value is 0.02667)

Significant PCs for Deaf/hearing differences (Fig. 8):

- PC1 for eyebrow movement (inner p-value is 0.02764 and outer is 0.03632)
- Hearing signers tend to have higher eyebrow raise than the deaf signers

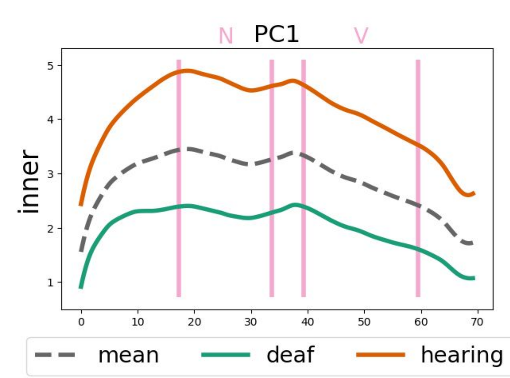


Fig. 8