HAL
open science

Quantifying facial expression intensity and signal use in deaf signers

Helen Rodger, Junpeng Lao, Chloé Stoll, Anne-Raphaëlle Richoz, Olivier Pascalis, Matthew Dye, Roberto Caldara

To cite this version:

Helen Rodger, Junpeng Lao, Chloé Stoll, Anne-Raphaëlle Richoz, Olivier Pascalis, et al.. Quantifying facial expression intensity and signal use in deaf signers. Journal of Deaf Studies and Deaf Education, 2019, 24 (4), pp.346-355. 10.1093/deafed/enz023 . hal-02567187v2

HAL Id: hal-02567187

https://hal.science/hal-02567187v2

Submitted on 25 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Heliyon 7 (2021) e07018

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

The recognition of facial expressions of emotion in deaf and hearing individuals

Helen Rodger [a,*,1], Junpeng Lao [a,1], Chloé Stoll [b,1], Anne-Raphaëlle Richoz [a], Olivier Pascalis [b], Matthew Dye [c], Roberto Caldara [a]

[a] Department of Psychology, University of Fribourg, Fribourg, Switzerland
[b] Laboratoire de Psychologie et de Neurocognition (CNRS-UMR5105), Université Grenoble-Alpes, France
[c] National Technical Institute for Deaf/Rochester Institute of Technology, Rochester, New York, USA

ARTICLE INFO

Keywords:
Facial expression recognition
Dynamic versus static expression recognition
Bayesian hierarchical modelling
Psychophysics

ABSTRACT

During real-life interactions, facial expressions of emotion are perceived dynamically with multimodal sensory information. In the absence of auditory sensory channel inputs, it is unclear how facial expressions are recognised and internally represented by deaf individuals. Few studies have investigated facial expression recognition in deaf signers using dynamic stimuli, and none have included all six basic facial expressions of emotion (anger, disgust, fear, happiness, sadness, and surprise) with stimuli fully controlled for their low-level visual properties, leaving the question of whether or not a dynamic advantage for deaf observers exists unresolved. We hypothesised, in line with the *enhancement hypothesis*, that the absence of auditory sensory information might have forced the visual system to better process visual (unimodal) signals, and predicted that this greater sensitivity to visual stimuli would result in better recognition performance for dynamic compared to static stimuli, and for deaf-signers compared to hearing non-signers in the dynamic condition. To this end, we performed a series of psychophysical studies with deaf signers with early-onset severe-to-profound deafness (dB loss >70) and hearing controls to estimate their ability to recognize the six basic facial expressions of emotion. Using static, dynamic, and shuffled (randomly permuted video frames of an expression) stimuli, we found that deaf observers showed similar categorization profiles and confusions across expressions compared to hearing controls (e.g., confusing surprise with fear). In contrast to our hypothesis, we found no recognition advantage for dynamic compared to static facial expressions for deaf observers. This observation shows that the decoding of dynamic facial expression emotional signals is not superior even in the deaf *expert* visual system, suggesting the existence of optimal signals in static facial expressions of emotion at the apex. Deaf individuals match hearing individuals in the recognition of facial expressions of emotion.

1. Introduction

In everyday life, facial expression is an integral part of nonverbal communication. A wealth of personal and interpersonal information is transmitted through facial expressions (Jack and Schyns, 2015). A critical aspect of facial expression is the encoding and decoding of emotion. Indeed, successful communication relies on a common, shared representation and understanding of internal emotional states. The classical theory of universal facial expressions of emotion, dating back to Darwin's early research, assumes that a collection of distinct emotions can be commonly recognised across different cultures: the six basic emotions (i.e., anger, disgust, fear, happiness, sadness, and surprise; Ekman and Friesen, 1975, 1978). However, recent research has provided increasing evidence that the internal representations and expressions of human emotions are *not* universal. The categorisation of facial expressions of emotion has instead been shown to be biased by different cultures (see Caldara 2017, Jack et al., 2009; 2012), to differ across the human lifespan (Rodger et al., 2018; Rodger et al., 2015; Richoz et al., 2018), and is highly contextualised (e.g., Aviezer et al., 2011). Together, these findings indicate that already within healthy populations, facial expression recognition differs according to cultural, developmental and contextual influences. Few studies have examined facial expression recognition

* Corresponding author.
E-mail address: helen.rodger@unifr.ch (H. Rodger).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2021.e07018
Received 3 February 2021; Received in revised form 25 March 2021; Accepted 4 May 2021
2405-8440/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

ability in deaf populations. This is surprising, as in the absence of an auditory channel, and given the multisensory nature of facial expression recognition, it is possible that internal representations of facial expressions of emotions may differ between deaf and hearing populations.

In addition to recent advances demonstrating that facial expressions are not perceived in a universal manner, further research has shown that to better understand facial expression recognition, it is necessary to acknowledge that our internal representations of facial expressions of emotion are developed from dynamic multimodal signals (visual, auditory, and other spatial-temporal contextual information, e.g., Campanella and Belin, 2007; Collignon et al., 2008). Until recently, facial expression recognition studies have typically tested recognition using restricted sets of unimodal static facial images. As our understanding of the influence of dynamic and multimodal displays of emotion on recognition develops, previous findings based on such stimuli could potentially be limited. Recently, dynamic face stimuli have been increasingly used in research with hearing observers, as such stimuli are argued to be more ecologically valid depictions of the emotional expressions encountered in everyday life (e.g., Johnston et al., 2013; Paulmann et al., 2009; Trautmann et al., 2009). Despite this assumption, healthy adults do not always recognise facial expressions of emotion better from dynamic compared to static facial stimuli. Some studies suggest there is a dynamic advantage for expression recognition (e.g., Ambadar et al., 2005; Cunningham and Wallraven, 2009; Giard and Peronnet, 1999; Knappmeyer et al., 2003; Paulmann et al., 2009; Wehrle et al., 2000), however later studies indicate that the dynamic advantage is either minimal (e.g., Gold et al., 2013) or inexistent (e.g., Fiorentini and Viviani, 2011). A recent large cross-sectional study showed that the advantage of recognising dynamic facial expressions is emotion-specific as a function of age (Richoz et al., 2018). For example, the advantage of recognising the dynamic facial expressions of disgust and surprise is consistent across age, indicating a reliable effect of a dynamic advantage for at least these two emotions (Richoz et al., 2018).

Taken together, the research indicates that for hearing individuals, facial expressions of emotion are multimodal representations comprising mainly visual, auditory, and spatial-temporal contextual information. Furthermore, dynamic displays of emotion can aid recognition of specific emotions as a function of age. Given the dynamic multimodal nature of facial expression recognition, in the absence of an auditory channel, we hypothesise that deaf observers represent emotion differently to hearing observers. Studies pointing to differences between deaf and hearing groups have suggested that when compared with age-matched hearing controls, deaf children experience a possible developmental delay in facial expression recognition ability (Most and Michaelis, 2012; Wang et al., 2011). However, other findings indicate that differences in emotion recognition ability between deaf and hearing children are related to language ability rather than deafness per se (Dyck et al., 2004; Sidera et al., 2017). Further findings supporting this assertion have shown that deaf children with normal language ability show similar emotion recognition ability compared to age-matched hearing controls (Ziv et al., 2013; Hosie et al., 1998).

These findings indicate there is a difference between hearing and deaf individuals in facial expression recognition during childhood. Few studies have investigated the impact of dynamic displays of emotion on facial expression recognition in deaf populations. In the absence of an auditory channel, it is possible that deaf observers have developed greater sensitivity to visual stimuli. This is known as the enhancement hypothesis (Sidera et al., 2017). Therefore, in line with this hypothesis, we hypothesise that the additional perceptual information provided by dynamic stimuli would result in better recognition performance for dynamic compared to static stimuli, and deaf compared to hearing observers. Jones, Gutierrez and Ludlow (2018) found a dynamic advantage for facial expression recognition in young deaf children. Using dynamic and static facial expression stimuli Jones et al. (2018) tested deaf and hearing children between 6-12 years old in two studies. In the first, they found a dynamic advantage in deaf children as they performed better in the dynamic compared to the static condition, whereas no difference was found between conditions for hearing children. However, in the second study, in which the intensity of the facial expression was varied, no group differences were found.

In a study with adult deaf signers and hearing non-signers, using dynamic facial stimuli, which included both emotional and communicative facial expressions, contrarily to the study with children, Grossman and Kegl (2007) found that hearing participants categorize expressions more accurately than deaf participants do. Interestingly, both groups showed similar patterns of misidentification when their responses were compared in confusion matrices. As a confusion matrix displays all possible response information to a multiple-choice problem, it is a direct reflection of the internal representation of different emotion categories. This result is further surprising as it is expected that the practice of sign language could potentially influence how facial expressions of emotion are encoded and decoded, as the face is used to express language-specific grammatical signals (Aarons, 1996; Bahan, 1996; MacLaughlin, 1997; Neidle et al., 1996; Baker-Shenk, 1983; 1986; Neidle et al., 1997; Petronio and Lillo-Martin, 1997; Hoza et al., 1997). However, it is difficult to draw solid conclusions from this early dynamic study as no static stimuli were presented as a control condition, and only a subset of the six basic emotions was included (Grossman and Kegl, 2007).

Recently, using static stimuli, which varied in emotion intensity, Stoll et al. (2019) measured the quantity of signal and emotional intensity required to categorise all six basic emotions in deaf signers and hearing non-signers. Recognition performance between the two groups was comparable for all emotions with the exception of disgust, for which deaf signers needed higher levels of signal and intensity for accurate recognition. Another recent study also using static stimuli only, a smaller sample of participants than Stoll et al., and a range of tasks from the *Florida Affect Battery*, found that for two of the five tasks undertaken (the Facial Emotion Discrimination and Naming tasks), deaf participants had significantly poorer performance in recognising fear expressions (Martins et al., 2019). Disgust was not studied. The confusions made between expressions were not analysed, so it is not possible to determine whether lower performance for fear could be attributed to a difference in the confusions made between deaf and hearing participants.

In a further study, Krejtz et al. (2020) used 10-second morphed facial expressions from a neutral to one of three basic facial expression (happiness, sadness, or anger). The task was to recognize the displayed emotion as quickly as possible. Deaf participants showed a marginal, non-significant effect ($p = 0.6$) for response times in categorizing the facial expressions of emotion, while both groups reached a similar level of accuracy. However, such performance was achieved with a limited decisional space (3 out of the 6 basic facial expressions of emotion; see Ramon et al., 2019), a slow unecological unfolding of facial expressions over time (i.e., 10 s to reach the peak intensity), and stimuli varying for both the low-level visual properties and the amount of low-level physical information, as well as the absence of a comparison with static facial expressions. The limitations of this study thus leave the current question, of whether a dynamic advantage for facial expression recognition in deaf adult signers exists, unresolved. Overall, performance for static emotional expression recognition has largely been comparable between deaf and hearing adult populations. Further studies are necessary to determine whether the encoding and decoding of dynamic facial expressions is different in deaf individuals compared to their hearing counterparts for all six basic emotions, with well controlled stimuli displaying ecologically valid facial expressions.

To this aim, we quantified recognition ability for the six basic expressions of emotion (anger, disgust, fear, happiness, sadness, and surprise) in young adult deaf signers and hearing non-signers using dynamic, shuffled, and static face stimuli. We explicitly asked our observers to be as accurate as possible, since the majority of facial expression recognition studies address precision rather than latency, especially when comparing across populations (e.g., Richoz et al.,

2018; Rodger et al., 2015; Rodger et al., 2018; Wyssen et al., 2019). Accuracy is the benchmark measure to assess the recognition of facial expressions of emotion. Notably, even when only three expressions are used, no significant effects were found for latency across populations (Krejtz et al., 2020). Thus, we hypothesized that in the absence of an auditory channel and with the additional perceptual information provided by dynamic emotional stimuli, deaf signers would show higher recognition performance in the dynamic compared to the static condition, and compared to hearing observers. Importantly, the stimuli were controlled for low-level image properties, such as luminance and contrast, as well as the amount of low-level physical information (see Gold et al., 2013). We tested a relatively large sample of participants. We modelled expression recognition ability as a function of group, emotion, and stimulus category (static, dynamic, or shuffled), and aimed to provide an estimation of the magnitude of any dynamic advantage for facial expression recognition. The shuffled condition, in which randomly permuted video frames of an emotional expression are presented, was included as a control for the dynamic condition to show that the addition of spatial-temporal information alone does not influence expression recognition. Instead, the sequence of the spatial temporal information is essential, so a dynamic advantage would only be found for the dynamic condition in which an emotional expression is dynamically unfolding in sequence, from a neutral expression to full intensity. Using a Bayesian hierarchical multinomial regression model, we estimated a complete confusion matrix for both deaf and hearing participants' performance for the three stimulus categories (static, dynamic, and shuffled) to determine whether recognition performance differed across groups, conditions, and expressions.

2. Methods

The experimental material, raw data, and analysis scripts are openly available at https://osf.io/6bdn8.

2.1. Participants

A total of 45 undergraduate deaf students from the National Technical Institute for Deaf/Rochester Institute of Technology, participated in the current study. Four participants were excluded from the data analysis as they did not complete the whole experiment. The age range of the deaf participants (25 females, 20 males) was between 18 and 30 years of age ($M = 21.7$ years old; $SD = 2.4$ years). All deaf participants had severe to profound hearing loss (dB loss >70) from birth or the first three years of their life and were all native or early ASL signers (before the age of 5). Among the deaf participants, 12 used cochlear implants (4 occasionally, 8 all the time/every day) and 12 used a hearing aid (7 occasionally, 5 all the time/every day). In addition, 19 participants had a deaf family member (parents and/or siblings). 46 hearing non-signers (28 females, 18 males) from the Rochester Institute of Technology were also tested. The overall age of hearing participants was between 18 and 31 ($M = 21.41$ years old; $SD = 3.26$ years). All participants had normal or corrected-to-normal vision. Participants provided written informed consent and received $10 for their participation. The study was approved by the local Ethical Committee of the Rochester Institute of Technology (RIT, Rochester, New-York).

2.2. Stimuli

The face stimuli set was taken from Gold et al. (2013). It contains dynamic facial expressions from eight different actors (four females), each displaying the six basic emotions (i.e., anger, disgust, fear, happiness, sadness and surprise; Ekman and Friesen, 1976). Each video stimulus lasts for 1 s at a frame rate of 30 frames per second. The face stimuli start from a neutral expression and naturally progress to the maximum intensity of a specific facial expression of emotion within one second. If the fully articulated expression is reached before one second, the actors maintain the full intensity expression until the end of the video. The face stimuli were cropped at the hairline to display only internal facial features, and were presented in grayscale. The stimuli were resized to 768

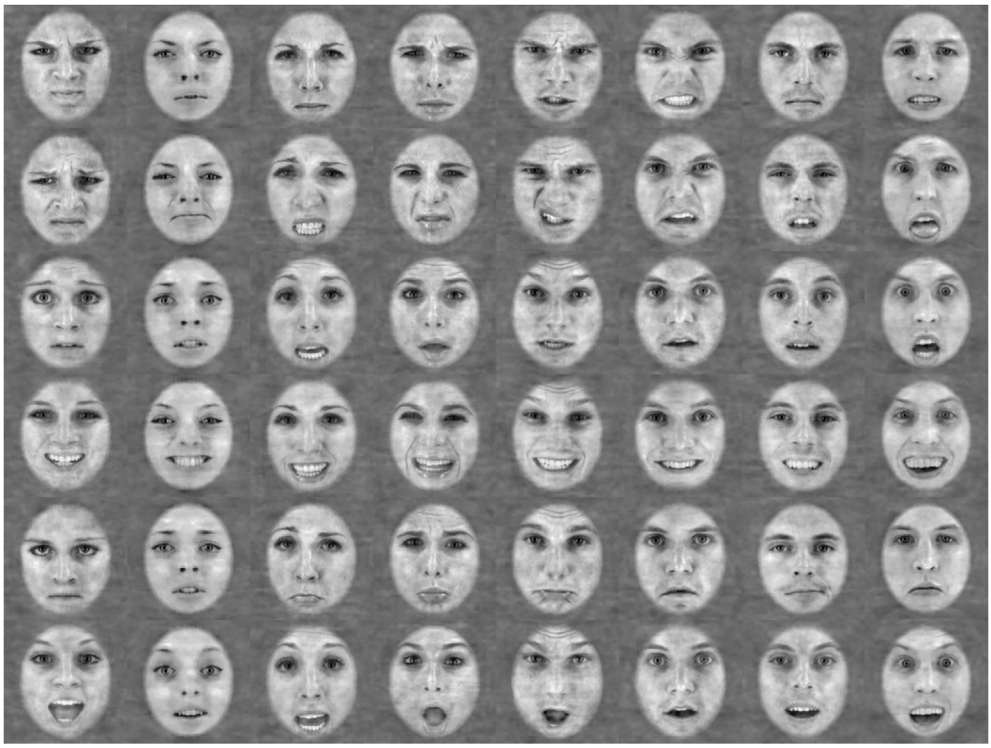

Figure 1. Apex frame of each actor (column) for the six expressions (row) at maximum intensity. Adapted with permission from Gold et al. (2013).

pixels in height and 768 pixels in width, which subtended a visual angle of 12° on the screen, at a viewing distance of 70 cm. To control for the low-level properties and the image statistics of the stimuli, we normalized the videos across all frames and all expressions using the SHINE toolbox with the default option (Willenbockel et al., 2010). Figure 1 shows the apex frames from all video stimuli after normalization.

Three different conditions were generated from the dynamic face stimuli: dynamic, static, and shuffled. In the dynamic condition, each face stimulus was presented in its original frame order. In the shuffled condition, the video was shown with the order of the frames randomly permuted. In the static condition, the apex frame of the face stimulus was presented for 1 s (i.e., 30 frames). The raw pixel intensity differences between each frame of the dynamic movies were computed. To maintain the same pixel-level intensity change between the frame changes in the static condition, we added the randomly permuted pixel intensity difference between every 2 frames to each frame at random locations in the static images. The resulting stimulus in the static condition thus contained the same pixel-level intensity change between two consecutive frames as the dynamic condition. A demo of the visual stimuli, as viewed by the participants, can be found in the supplementary videos.

Finally, the stimuli were displayed on a color liquid-crystal display (LCD) with a resolution of 1440 × 900 pixels and a refresh rate of 60 Hz, at a distance of 70 cm from the participant. The whole experiment was programmed in Matlab (Matlab, 2014B) using the Psychophysics Toolbox (Brainard, 1997).

2.3. Procedure

The experimental procedure is shown in Figure 2. A white fixation cross was presented in the centre of the screen for 500 ms at the beginning of each trial. A random stimulus was then presented in the centre of the screen for 1 s. Note that the same presentation time of 1 s was also previously used in other studies using dynamic face stimuli (Adolphs et al., 2003; Recio et al., 2013; Richoz et al., 2015). After the face stimulus was presented for 1 s, it disappeared and a response window was displayed on the screen until the participant responded. The participant's task was to categorise how the person presented on the screen was feeling by pressing the corresponding expression key on the keyboard. Participants were told that they would see faces expressing: anger, disgust, fear, happiness, sadness, or surprise. They could also press a "don't know" key if they were unsure, had not had enough time to see the expression, or did not know the answer. This option was included to reduce the noise and

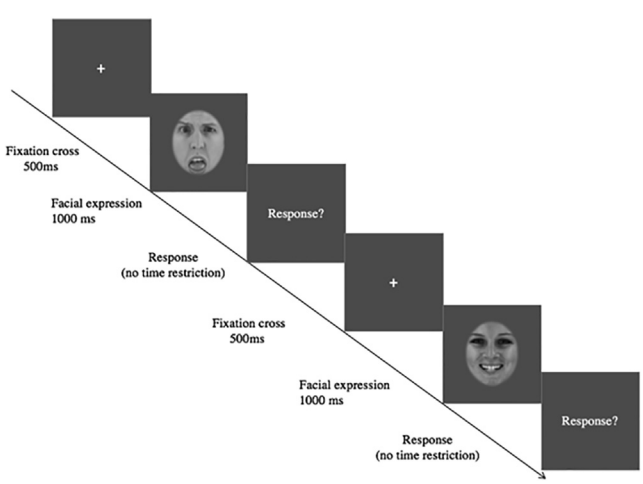

Figure 2. Schematic representation of the procedure. Each trial began with a white fixation cross that was presented for 500 ms, followed by a face presented for 1s, which expressed one of the six basic facial expressions of emotion: anger, disgust, fear, happiness, sadness, or surprise. After each trial, participants were asked to categorize the previously seen expression.

response bias produced by the lack of such a key. No feedback was provided after a response, and participants were instructed to respond as accurately as possible with no time restriction. The stimuli were blocked by conditions. Each condition (i.e., dynamic, shuffle, static) consisted of two blocks of 48 trials (eight actors, six expressions) presented twice (96 expressions for each condition), for a total of 288 trials. Participants took part in all three conditions, which were counterbalanced in random order. Before starting the testing phase, participants completed 12 practice trials for each condition. The whole experiment lasted around 30 minutes. For deaf participants, instructions were both written and signed by the experimenter.

2.4. Data analysis

The response from each participant is modelled using a Hierarchical Multinomial Regression Model. Specifically, for each participant, their response is organised in a 3 (conditions) by 6 (presented facial expressions of emotion) by 7 (possible behavioural categorisation: 6 emotions + I don't know) array. Each entry of the array is the number of responses per stimuli type and presented expression. The response array is modelled using a Multinomial distribution, with the response probabilities p_1, ..., p_7 modelled as a *SoftMax* transformation of a multivariate mixed-effect model. The full model is displayed below (*i* for each task, *j* for each expression, *g* for each group, and *k* for each participant):

Observed response

$$\left[y^1_{k,i,j}, \ldots, y^7_{k,i,j}\right] \sim \text{Multinomial}\left(n, p_{k,i,j}\right)$$

$$p_{k,i,j} = \text{SoftMax}\left(\mu_{k,i,j}\right)$$

Mixed – effect model

$$\mu_{k,i,j} = X^{kg}\beta_{g,i,j} + b_{k,i,j}$$

Group level coefficients

$$\beta_{g,i,j} \sim \text{Normal}(0,1)$$

Subject – level coefficients

$$b_{k,i,j} \sim \text{Normal}(\eta_k, 1)$$

$$\eta_k \sim \text{Normal}(0,1)$$

Hierarchical multinomial regression model of participant emotion responses

We used PyMC3 (version 3.3; Salvatier et al., 2016) to build the Hierarchical Multinomial Regression Model and performed probabilistic inference using NUTS to sample from the posterior distribution (Hoffman and Gelman, 2014). Four MCMC chains were run with 3000 samples, each with the default sampler setup. The first 2000 samples were used to tune the mass matrix and step size for NUTS. These samples were subsequently discarded, leaving a total 4000 samples for each model parameter. Model convergence was diagnosed by computing Gelman and Rubin's convergence diagnostic (R-hat, 1992), examining the effective sample size, checking whether there is any divergent sample that has been returned from the sample, and visually inspecting the mixing of the traces (Gabry et al., 2017). Using the posterior samples, we computed the estimation of behavioural response confusion matrices conditioned by group and stimuli type. We then performed contrasts between different groups and conditions to estimate the effect of groups, stimuli type, and their interactions. Data analysis was performed in Python using Jupyter Notebook. The results were displayed using Seaborn and Matplotlib.

3. Results

3.1. Descriptive results

Summary statistics by group and condition are displayed as confusion matrices in Figure 3. The percentage of accurate recognition (the diagonal of the confusion matrices) is shown in the line plots of Figure 4. We observed a similar pattern of recognition performance for the emotion categories in both groups to that which is typically

reported in the facial expression recognition literature; the highest average recognition accuracy was for happiness and the lowest was for fear (e.g., Calder et al., 2003; Rapcsak et al., 2000; Richoz et al., 2018; Rodger et al., 2015; Rodger et al., 2018; Zhao et al., 2016) (Figure 4). Moreover, both groups of observers displayed similar confusion patterns, for example, fear was easily confused with surprise, whereas surprise was at times miscategorised as happy. Importantly, the performance of the deaf observers was nearly identical in comparison to the hearing controls. The largest performance difference between the two groups was for surprise in the dynamic condition: the average accuracy of the deaf participants was 0.685 [0.628, 0.739] (bracket shows the 95% bootstrap confidence interval), whereas the average accuracy of the hearing observers was 0.781 [0.743, 0.817]. A full table reporting the mean recognition performance can be found in the analysis notebook.

3.2. Hierarchical Multinomial Regression Model

Using the posterior distribution, we computed the population level effect (i.e., fixed effect) of the mixed-effect model: the categorisation performance for the dynamic condition (Figure 5), the static condition (Figure 6), and the shuffled condition (Supplementary Figure 1). Similar to the descriptive results presented above, both groups of observers showed no significant difference in the estimated behavioural performance. The largest group difference was again for the facial expression of surprise in the dynamic condition: accuracy was estimated at 0.713 [0.630, 0.791] (brackets show the 95% Highest Posterior Density HPD Interval) for deaf participants and 0.811 [0.748, 0.868] for hearing controls, with the difference estimated at -0.0975 [-0.196, 0.00648]. This difference was not significant as the distributions overlap.

To quantify the potential advantage of dynamic facial expressions over static stimuli, we computed the contrast of dynamic – static within each group of observers. As shown in the diagonal subplots of Figure 7, there is no strong indication of a dynamic advantage except for the facial expression of surprise. Both groups recognised surprise better when it was presented dynamically: with 3.98% [-0.98, 8.938] more accuracy for deaf participants and 10.30% [5.861, 15.135] more accuracy for hearing controls. The magnitude of the dynamic advantage is stronger for hearing controls than for deaf observers: the estimation difference is 6.31% [-0.229, 13.436]. The dynamic advantage is mostly driven by the difference in confusing surprise with happiness: 1.26% [-3.483, 5.680] less in the dynamic condition for deaf participants and 7.17% [3.037, 11.358] less for hearing controls.

4. Discussion

Our results show a common pattern of recognition for facial expressions of emotion in deaf and hearing observers: both groups of observers show similar accuracy and confusion in recognising the six basic emotions. Moreover, participants responded in a similar way to both static and dynamic presentation of the stimuli. There is no clear effect of a dynamic advantage in either group, except for the expression of surprise: hearing controls recognised surprise more accurately when it was presented dynamically. The average advantage was estimated at around 10% recognition performance improvement for hearing controls. Deaf observers also showed an increase of approximately 6% accuracy when surprise is presented dynamically, however, the posterior estimation overlapped with zero, which means this estimated advantage failed to reach significance. A dynamic advantage for surprise recognition in hearing individuals has also previously been found (Richoz et al., 2018). Here, the dynamic advantage in the hearing group for surprise was

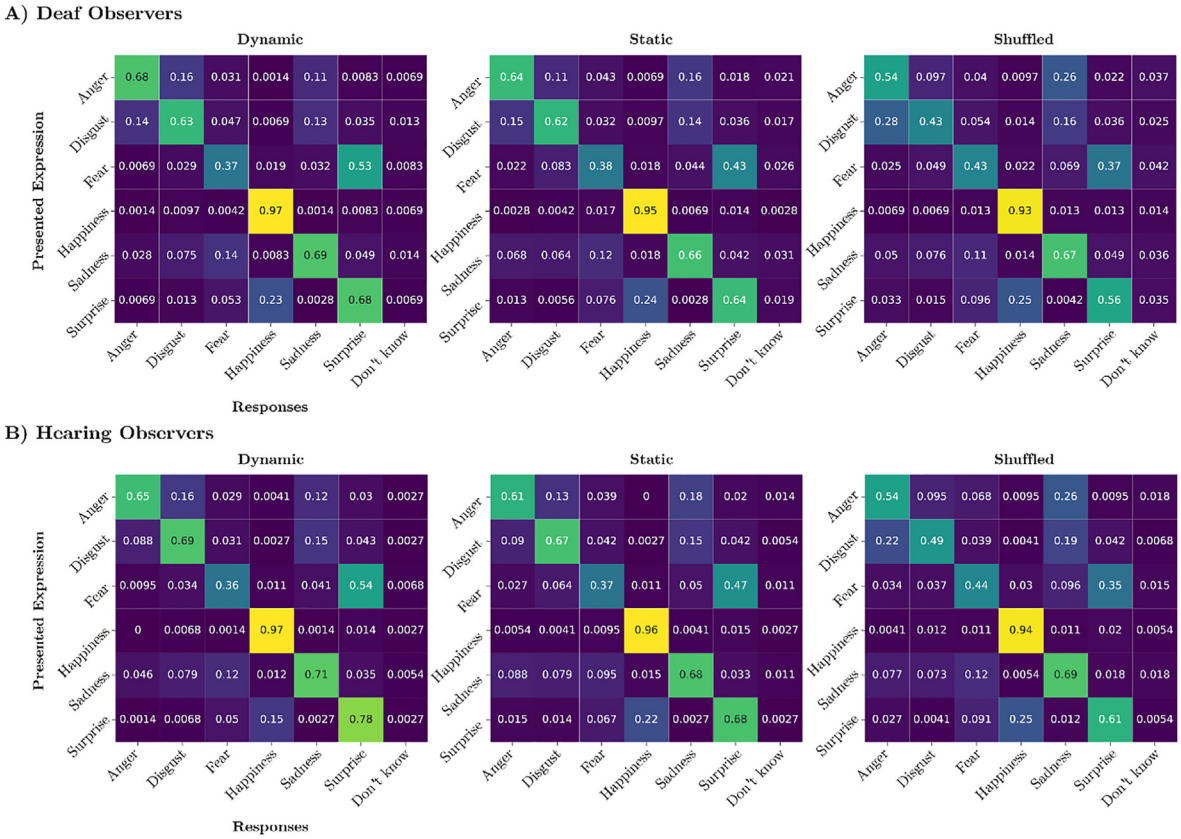

Figure 3. Confusion matrices of A) deaf and B) hearing observers for each stimuli type. The six basic facial expressions of emotion that were presented to the participants are displayed on different rows, while each column shows the average frequency of the response given by the participants. Dark blue tones indicate low frequency while blue-to-green shades indicate high frequency. The values in the main diagonal indicate the recognition performance for each expression.

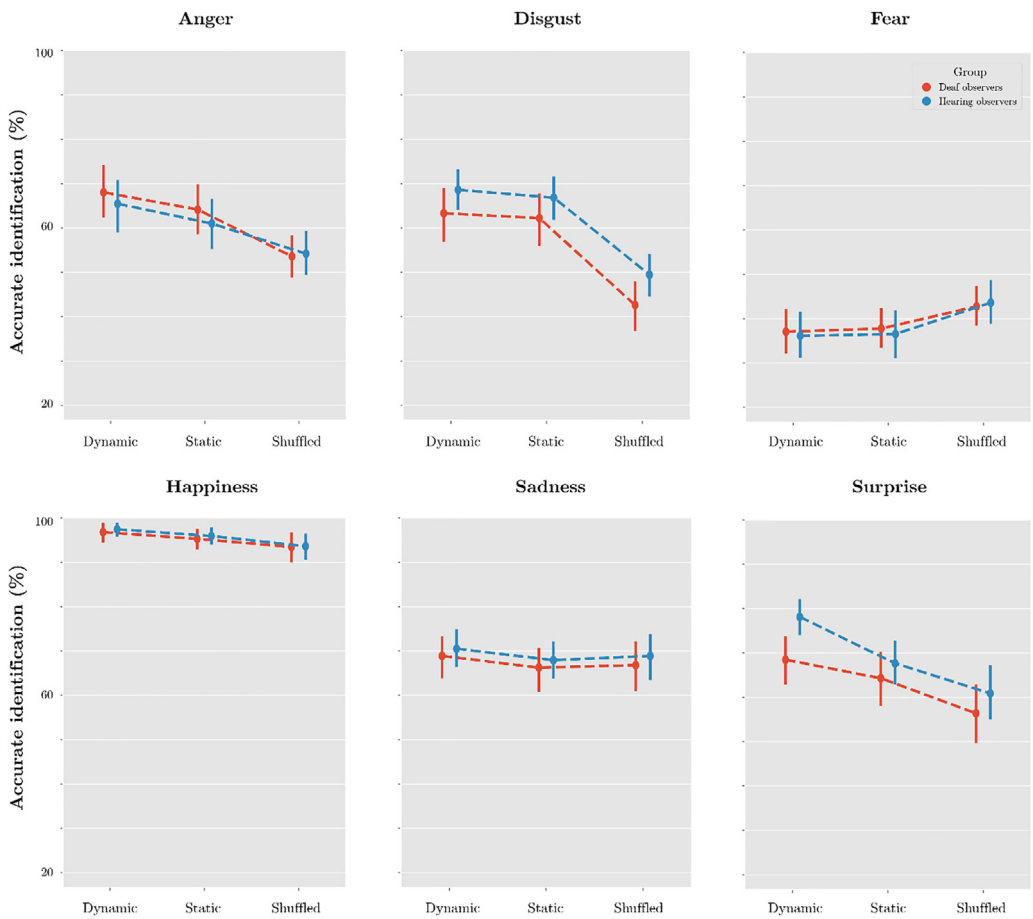

Figure 4. Mean recognition accuracy across the two groups of observers for each expression in the three different conditions. Error bars show the 95% bootstrap confidence interval for the mean. Red indicates the performance of deaf observers and blue indicates hearing observers.

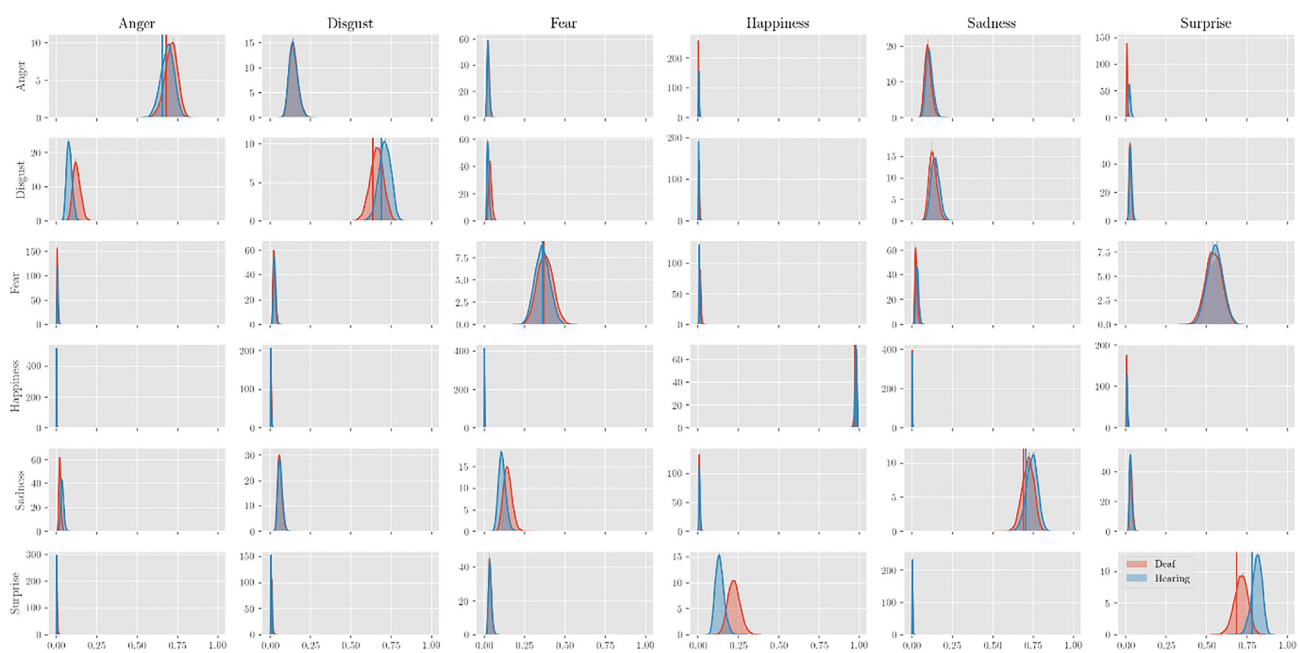

Figure 5. The marginal posterior distribution of the participants' response probability conditioned on the facial expressions of emotion in the Dynamic condition. The six basic facial expressions which were presented to the participants are displayed on different rows, while each column shows the responses given by the participants (the *Don't know* response is not shown). The subplots in the diagonal are the accurate identification, with the group average shown as vertical lines. Red indicates deaf observers' performance and blue indicates hearing observers. Overlapping distributions indicate that there are no significant differences between groups.

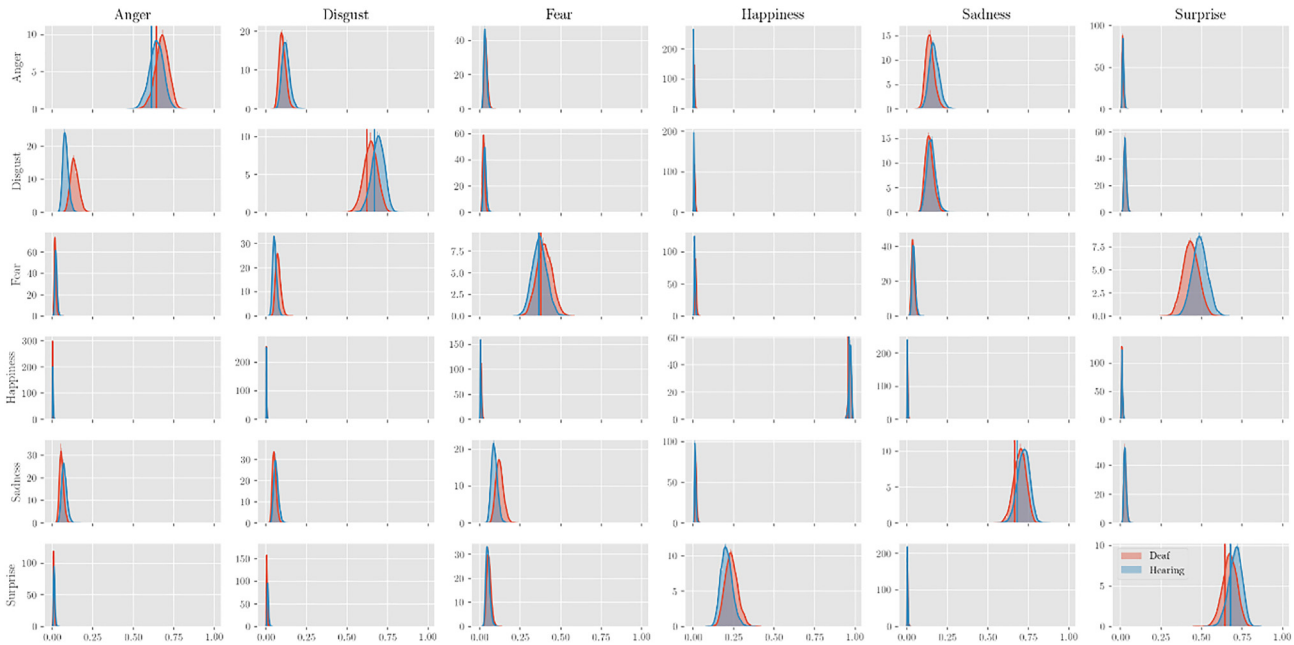

Figure 6. The marginal posterior distribution of the participants' response probability conditioned on the presented facial expressions of emotion in the Static condition.

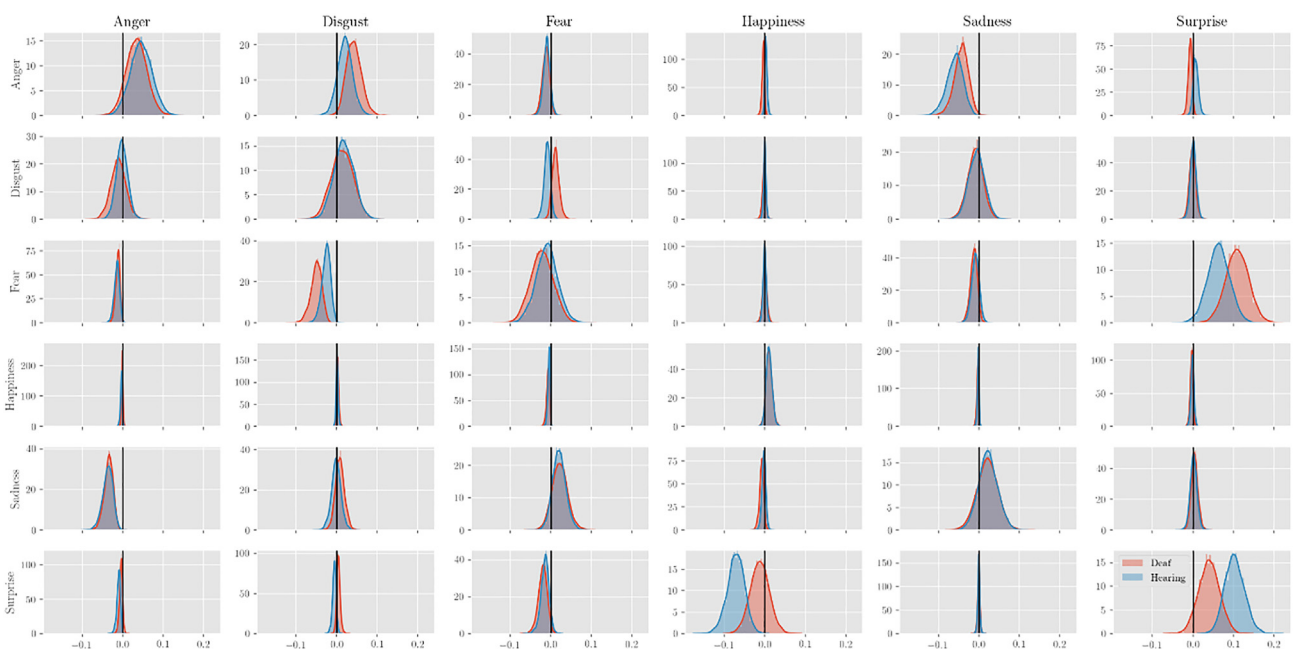

Figure 7. The marginal posterior distribution of the contrast (Dynamic – Static) of the participants' response probability. The black vertical line indicates the zero (i.e., no difference). Red color indicates deaf observers and blue color indicates the hearing observers.

explained by a lower number of miscategorisations of surprise for happiness in the dynamic compared to the static and shuffled conditions.

The results contradict our hypothesis, in which we expected deaf observers to perform better in the recognition of dynamic compared to static expressions, and compared to hearing individuals. A recent study found comparable performance between deaf and hearing observers for the recognition of static stimuli of varying intensities (Stoll et al., 2019). However, previous studies using dynamic stimuli found a dynamic advantage in deaf children when the stimuli did not vary in emotional intensity (Jones et al., 2017), and no advantage was found in deaf adults in a study testing anger and surprise expressions (Grossman and Kegl, 2007). In the present study, which tested all six basic emotions within the same paradigm, the additional perceptual information provided by a dynamic expression of emotion did not affect recognition performance as predicted. Our findings are reinforced by a recent eye-tracking study with deaf and hearing observers in which no differences between groups were found as recognition accuracy for the three dynamic emotions tested, happiness, sadness and anger, were similar (Krejtz et al., 2020). Despite finding comparable recognition performance between deaf and hearing adult observers for the six basic emotions, it is not possible to conclude at this stage that the facial processing systems of deaf and hearing adults are equivalent. While our experimental approach has shown sensitivity to reveal significant differences across different populations when they are present (e.g., Richoz et al., 2018), in the context of this study it can only

be inferred that dynamic stimuli do not provide an advantage in facial expression recognition for deaf observers. Our sample size conforms to those used in previous studies in the literature, nonetheless larger sample sizes should be used in the future to confirm our findings. Presently, this lack of a dynamic advantage for facial expression recognition in young adult deaf signers adds to our understanding of emotion recognition in this population, a question which has been little studied previously.

Krejtz et al. (2020) found a marginal, non-significant effect for response times during the categorization of morphed neutral to happy, angry and sad expressions. In the present study, we explicitly instructed our observers to favour accuracy rather than speed, in order to conform to our previous studies comparing different populations (e.g., Richoz et al., 2018; Rodger et al., 2015; Rodger et al., 2018; Wyssen et al., 2019) and the large majority of the literature in affective science. As such, response times could not be analysed here. Future studies are necessary to further clarify whether the marginal effects reported by Krejtz et al. (2020) would become significant or not when using static and dynamic stimuli of the six basic facial expressions of emotion, with stimuli controlled for ecologically valid time unfolding, low-level properties, and the amount of information displayed by the expressions.

Another factor to consider in interpreting the results is the effect of sign language on facial expression recognition. As mentioned in the introduction, previous studies comparing deaf and hearing participants have emphasised the effect of sign language experience on face processing and recognition. The practice of sign language has been thought to potentially influence how facial expressions of emotion are encoded and decoded as faces express language-specific grammatical signals in sign language communication (Aarons, 1996; Bahan, 1996; MacLaughlin, 1997; Neidle et al., 1996; Baker-Shenk, 1983; 1986; Neidle et al., 1997; Petronio and Lillo-Martin, 1997; Hoza et al., 1997). Stoll et al. (2017) showed that signers (both deaf and hearing) performed slower but more accurately in a face identification task than non-signers. While both deaf and hearing observers perform similarly in the current study, it does not necessarily mean that there is no effect of sign language on the recognition of facial expressions. Indeed, linguistic and emotional facial expressions are shown to be processed differently by deaf signers compared to hearing non-signers (Corina et al., 1999; McCullough et al., 2005). Therefore, it is possible that only facial expressions used in sign language (i.e., linguistic facial expressions) show clear processing differences in deaf signers. Or, that a difference in facial expression processing, similarly to facial identification processing, could exist between deaf signers and non-signers. To have a deeper understanding of the coding of facial expressions in deaf populations, it is necessary to test both hearing and deaf signers and non-signers with a larger variety of expressions, including linguistic expressions. Stoll et al. (2017) also found that signers cumulated visual information faster and had a larger threshold using a drift-model, which potentially indicates an over-sampling strategy. In the present study, the participants were required to respond as accurately as possible without any time constraints. Therefore, reaction time was not measured. Future studies which also consider potential reaction time differences are necessary to further investigate the interaction between deafness and sign language in facial expression processing.

There are many practical advantages provided by the Bayesian model estimation used here to compare deaf signers and hearing non-signers. Using a hierarchical model, we can fully account for the effect of group and condition (fixed effect), and the effect of individual differences (random effect). With a Bayesian modelling framework, it is possible to implement different priors to regularize the model estimates appropriately; this ensures that all parameters are identifiable and give appropriate shrinkage to the estimation in the hierarchical model. We applied an informative prior on the latent response rate for each expression, using a Softmax transformation on the Normal distribution to create a simplex. The Softmax normal prior put more prior weight on the extreme values towards 0 or 1, which generate a "denoise" effect on the observed response (Gelman et al., 1996). In addition, posterior credible intervals over model parameters are intuitively easier to interoperate, as they represent our belief that the intervals where the "true" parameter lies within a given probability, conditioned on all the available information.

Previous statistical inference on confusion matrices often involves null hypothesis testing independently performed on each cell of the matrix. However, such testing is inaccurate when the cell contains a low number of observations (e.g., if only a few participants make that particular response). Proper multiple comparison corrections are also potentially challenging. Here, we used a multinomial regression model that jointly inferences all responses to one presented expression, thus avoiding both issues described above. Moreover, in the current experiment, we included an additional "don't know" response. This is a more powerful way to capture the potential subtle differences in the representation and response of different (expression) categories, as uncertain categorisation would be marginalised into this additional response without causing a random response. Equally, in this way, all miscategorisations remain true miscategorisations, as uncertain responses are separately categorised. Therefore, as described earlier in the introduction, the dynamic advantage for surprise in the hearing group can be clearly explained by the lower incidence of miscategorisations of surprise for happiness in the dynamic compared to the static and shuffled conditions (see Figure 3). It is also possible to see that standard confusions between certain emotions are present. For example, in the literature the most well-known confusion between emotion categories is fear and surprise due to their visual similarity (Gagnon et al., 2010; Jack et al., 2009; Roy-Charland et al., 2014). This confusion is consistently found in adult and child populations (Jack et al., 2009; Matsumoto and Sung Hwang, 2011; Rodger et al., 2015: Rodger et al., 2018). This common confusion is also present for both the deaf and hearing participants studied here, as well as for all three stimulus types (dynamic, static and shuffled).

In conclusion, the categorisation of static and dynamic facial expressions of emotion is similar in both deaf and hearing young adults. By quantifying the response confusion matrix using a Bayesian model, we found that both groups show nearly identical accuracy and confusions for the six basic emotions. The dynamic or static presentation of stimuli did not have a strong effect on categorisation performance for either deaf or hearing individuals, with the exception of surprise, for which hearing non-signers showed a dynamic advantage. However, deaf signers followed this trend for surprise recognition, with six percent improvement in accuracy (which was not sufficient to reach significance) compared to the hearing group's ten percent performance improvement. For the first time, these findings chart the recognition of static and dynamic expressions of all six of the basic emotions in deaf signers and age-matched hearing non-signers. The results, therefore, with a relatively large sample and sophisticated modelling effort, provide a quantitative baseline for investigating the recognition of facial expressions in young deaf adults. This observation shows that the decoding of dynamic facial expression emotional signals is not superior even in the deaf *expert* visual system, suggesting the existence of optimal signals in static facial expressions of emotion at the apex. In a world in which deafness is often thought of in the context of an impairment which may negatively affect common social experiences, the common visual recognition of facial expressions of emotion between deaf signing and hearing non-signing adults marks an important understanding of the similarities rather than the differences in our lived social experiences. Deaf individuals match hearing individuals in the recognition of facial expressions of emotion.

Declarations

Author contribution statement

Helen Rodger, Junpeng Lao: Analyzed and interpreted the data; Wrote the paper.

Chloé Stoll, Anne-Raphaëlle Richoz: Performed the experiments; Analyzed and interpreted the data.

Olivier Pascalis, Matthew W Dye: Contributed reagents, materials, analysis tools or data.

Roberto Caldara: Conceived and designed the experiments.

Funding statement

This work was supported by the Swiss National Science Foundation (100014_156490/1).

Data availability statement

Data will be made available on request.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2021.e07018.

Acknowledgements

R.C. and J.L. were supported by the Swiss National Science Foundation (n° 100014_156490/1) awarded to R.C. C.S. was supported by the French-American Fulbright Comission, PhD program award. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Aarons, D., 1996. Topics and topicalization in American sign language. Stellenbosch Papers in Linguistics 30, 65–106.
Adolphs, R., Tranel, D., Damasio, A.R., 2003. Dissociable neural systems for recognizing emotions. Brain Cognit. 52 (1), 61–69.
Ambadar, Z., Schooler, J.W., Cohn, J.F., 2005. Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions. Psychol. Sci. 16 (5), 403–410.
Aviezer, H., Dudarev, V., Bentin, S., Hassin, R.R., 2011. The automaticity of emotional face-context integration. Emotion 11, 1406–1414.
Bahan, B., 1996. Nonmanual Realizations of Agreement in American Sign Language. Boston University, Boston MA. Unpublished doctoral dissertation.
Baker-Shenk, C., 1983. A Micro-analysis of the Nonmanual Component of Questions in American Sign Language. Unpublished doctoral dissertation, University of California, Berkeley.
Baker-Shenk, C., 1986. Factors affecting the form of question signals in American Sign Language. In: Sankoff, D. (Ed.), Diversity and Diachrony. John Benjamins, Amsterdam, Philadelphia, pp. 407–414.
Brainard, D.H., 1997. The psychophysics toolbox. Spatial Vis. 10, 433–436.
Caldara, R., 2017. Culture reveals a flexible system for face processing. Curr. Dir. Psychol. Sci. 26 (3), 249–255.
Calder, A.J., Keane, J., Manly, T., Sprengelmeyer, R., Scott, S., et al., 2003. Facial expression recognition across the adult life span. Neuropsychologia 41 (2), 195–202.
Campanella, S., Belin, P., 2007. Integrating face and voice in person perception. Itrends in Cognitive Sciences 11 (12), 535–543.
Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., Lepore, F., 2008. Audio-visual integration of emotion expression. Brain Res. 1242, 126–135.
Corina, D.P., Bellugi, U., Reilly, J., 1999. Neuropsychological studies of linguistic and affective facial expressions in deaf signers. Lang. Speech 42 (2–3), 307–331.
Cunningham, D.W., Wallraven, C., 2009. Dynamic information for the recognition of conversational expressions. J. Vis. 9 (13), 1–17, 7.
Dyck, M.J., Farrugia, C., Shochet, I.M., Holmes-Brown, M., 2004. Emotion recognition/understanding ability in hearing or vision-impaired children: do sounds sights, or words make the difference? JCPP (J. Child Psychol. Psychiatry) 45 (5), 789–800.
Ekman, P., Friesen, W.V., 1975. Unmasking the Face: a Guide to Recognizing Emotions from Facial Clues. Prentice-Hall, Englewood Cliffs.
Ekman, P., Friesen, W.V., 1978. Facial Action Coding System. Palo Alto, CA.
Fiorentini, C., Viviani, P., 2011. Is there a dynamic advantage for facial expressions? J. Vis. 11 (3), 1–15, 17.
Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., Gelman, A., 2017. Visualization in Bayesian Workflow. arXiv preprint arXiv:1709.01449.
Gagnon, M., Gosselin, P., Hudon-ven der Buhs, I., Larocque, K., Milliard, K., 2010. Children's recognition and discrimination of fear and disgust facial expressions. J. Nonverbal Behav. 34, 27–42.
Gelman, A., Bois, F., Jiang, J., 1996. Physiological pharmacokinetic analysis using population modeling and informative prior distributions. J. Am. Stat. Assoc. 91, 1400–1412.
Giard, M.H., Peronnet, F., 1999. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J. Cognit. Neurosci. 11 (5), 473–490.
Gold, J.M., Barker, J.D., Barr, S., Bittner, J.L., Bromfield, W.D., Chu, N., Goode, R.A., Lee, D., Simmons, M., Srinath, A., 2013. The efficiency of dynamic and static facial expression recognition. J. Vis. 13 (5), 1–12.
Grossman, R.B., Kegl, J., 2007. Moving faces: categorization of dynamic facial expressions in American sign language by deaf and hearing participants. J. Nonverbal Behav. 31 (1), 23–38.
Hoffman, M.D., Gelman, A., 2014. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J. Mach. Learn. Res. 15 (1), 1593–1623.
Hosie, J.A., Gray, C.D., Russell, P.A., Scott, C., Hunter, N., 1998. The matching of facial expressions by deaf and hearing children and their production and comprehension of emotion labels. Motiv. Emot. 22 (4), 293–313.
Hoza, J., Neidle, C., MacLaughlin, D., Kegl, J.A., Bahan, B., 1997. A Unified Syntactic Account of Rhetorical Questions in American Sign Language Syntactic Structure and Discourse Function: an Examination of Two Constructions in American Sign Language (American Sign Language Linguistic Research Project Report No. 4). Boston University, Boston, MA.
Jack, R.E., Schyns, P.G., 2015. The human face as a dynamic tool for social communication. Curr. Biol. 25 (14), R621–R634.
Jack, R.E., Blais, C., Scheepers, C., Schyns, P.G., Caldara, R., 2009. Cultural confusions show that facial expressions are not universal. Curr. Biol. 19 (18), 1543–1548.
Jack, R., Garrod, O., Yu, H., Caldara, R., Schyns, P.G., 2012. Facial expressions of emotion are not culturally universal. Proc. Natl. Acad. Sci. Unit. States Am. 109 (19), 7241–7244.
Johnston, P., Mayes, A., Hughes, M., Young, A.W., 2013. Brain networks subserving the evaluation of static and dynamic facial expressions. Cortex 49, 2462–2472.
Jones, A.C., Gutierrez, R., Ludlow, A.K., 2017. The role of motion and intensity in deaf children's recognition of real human facial expressions of emotion. Cognit. Emot. 9931, 1–14.
Knappmeyer, B., Thornton, I.M., Bulthoff, H.H., 2003. The use of facial motion and facial form during the processing of identity. Vis. Res. 43 (18), 1921–1936.
Krejtz, I., Krejtz, K., Wisiecka, K., Abramczyk, M., Olszanowski, M., Duchowski, A.T., 2020. Attention dynamics during emotion recognition by deaf and hearing individuals. J. Deaf Stud. Deaf Educ. 25 (1), 10–21.
MacLaughlin, D., 1997. The Structure of Determiner Phrases: Evidence from American Sign Lan- Gauge. Unpublished doctoral dissertation, Boston University, Boston, MA.
Martins, A.T., Faísca, T., Vieira, H., Gonçalves, G., 2019. Emotional recognition and empathy both in deaf and blind adults. J. Deaf Stud. Deaf Educ. 24 (2), 119–127.
Matsumoto, D., Sung Hwang, H., 2011. Judgments of facial expressions of emotion in profile. Emotion 11, 1223–1229.
McCullough, S., Emmorey, K., Sereno, M., 2005. Neural organization for recognition of grammatical and emotional facial expressions in deaf ASL signers and hearing nonsigners. Cognit. Brain Res. 22 (2), 193–203.
Most, T., Michaelis, H., 2012. Auditory, visual, and auditory– visual perceptions of emotions by young children with hearing loss versus children with normal hearing. J. Speech Lang. Hear. Res. 55 (4), 1148–1162.
Neidle, C., MacLaughlin, D., Bahan, B., Kegl, J., 1996. Non-manual Correlates of Syntactic Agreement in American Sign Language (American Sign Language Linguistic Research Project Report No. 2). Boston University, Boston, MA.
Neidle, C., MacLaughlin, D., Bahan, B., Lee, R., Kegl, J., 1997. The SignStreamTM Project (American Sign Language Linguistic Research Project Report No. 5). Boston University, Boston, MA.
Paulmann, S., Jessen, S., Kotz, S.A., 2009. Investigating the multimodal nature of human communication: an ERP study. J. Psychophysiol. 23 (2), 63–76.
Petronio, K., Lillo-Martin, D., 1997. WH-movement and the position of spec-CP: evidence from American sign language. Language 71 (1), 13–29.
Ramon, M., Sokhn, N., Caldara, R., 2019. Decisional space modulates visual categorization - evidence from saccadic reaction times. Cognition 186, 42–49.
Rapcsak, S.Z., Galper, S.R., Comer, J.F., Reminger, S.L., Nielsen, L., et al., 2000. Fear recognition deficits after focal brain damage. Neurology 54, 575–581.
Richoz, A.-R., Lao, J., Pascalis, O., Caldara, R., 2018. Tracking the recognition of static and dynamic facial expressions of emotion across the life span. J. Vis. 18 (9), 1–27, 5.
Rodger, H., Vizioli, L., Ouyang, X., Caldara, R., 2015. Mapping the development of facial expression recognition. Dev. Sci. 18 (6), 926–939.
Rodger, H., Lao, J., Caldara, R., 2018. Quantifying facial expression signal and intensity use during development. J. Exp. Child Psychol. 174, 41–59.
Roy-Charland, A., Perron, M., Beaudry, O., Eady, K., 2014. Confusion of fear and surprise: a test of the perceptual-attentional limitation hypothesis with eye movement monitoring. Cognit. Emot. 28 (7), 1214–1222.
Salvatier, J., Wiecki, T.V., Fonnesbeck, C., 2016. Probabilistic programming in Python using PyMC3. Peer J. Comput. Sci. 2, e55.
Sidera, F., Amadó, A., Martínez, L., 2017. Influences on facial emotion recognition in deaf children. J. Deaf Stud. Deaf Educ.
Stoll, C., Palluel-Germain, R., Caldara, R., Lao, J., Dye, M.W.G., Aptel, F., Pascalis, O., 2017. Face recognition is shaped by the use of sign language. J. Deaf Stud. Deaf Educ. (September), 1–9.

Stoll, C., Rodger*, H., Lao*, J., Richoz, A.-R.J., Pascalis, O., Dye, M.W.G., Caldara, R., 2019. Quantifying facial expression intensity and signal use in deaf signers. J. Deaf Stud. Deaf Educ. 24 (4), 346–355. *Joint first authors.

Trautmann, S.A., Fehr, T., Herrmann, M., 2009. Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. Brain Res. 1284, 100–115.

Wang, Y., Su, Y., Fang, P., Zhou, Q., 2011. Facial expression recognition: can preschoolers with cochlear implants and hearing aids catch it? Res. Dev. Disabil. 32 (6), 2583–2588.

Wehrle, T., Kaiser, S., Schmidt, S., Scherer, K.R., 2000. Studying the dynamics of emotional expression using synthesized facial muscle movements. J. Pers. Soc. Psychol. 78 (1), 105–119.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G.O., Gosselin, F., Tanaka, J.W., 2010. Controlling low-level image properties: the SHINE toolbox. Behav. Res. Methods 42 (3), 671–684.

Wyssen, A., Lao, J., Rodger, H., Humbel, N., Lennertz, J., Schuck, K., Munsch, S., 2019. Facial emotion recognition abilities in women suffering from eating disorders. Psychosom. Med.: J. Behav. Med. 81, 155–164.

Zhao, M.-F., Zimmer, H.D., Shen, X., Chen, W., Fu, X., 2016. Exploring the cognitive processes causing the age-related categorization deficit in the recognition of facial expressions. Exp. Aging Res. 42 (4), 348–364.

Ziv, M., Most, T., Cohen, S., 2013. Understanding of Emotions and False Beliefs Among Children versus deaf chil- dren. J. Deaf Stud. Deaf Educ. 18 (2), 161–174.