



HAL
open science

Quantifying facial expression intensity and signal use in deaf signers

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

► **To cite this version:**

Chloé Stoll, Helen Rodger, Junpeng Lao, 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-02567187v1

HAL Id: hal-02567187

<https://hal.science/hal-02567187v1>

Submitted on 7 May 2020 (v1), last revised 25 May 2021 (v2)

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.

Quantifying facial expression intensity and signal use in deaf signers

*Chloé Stoll^{1,2}, *Helen Rodger³, *Junpeng Lao³, Anne-Raphaëlle Richoz³, Olivier Pascalis¹, Matthew Dye⁴, and Roberto Caldara³

¹*Laboratoire de Psychologie et de Neurocognition (CNRS-UMR5105), Université Grenoble-Alpes, France*

²*Laboratory for Investigative Neurophysiology, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Switzerland*

³*Eye and Brain Mapping Laboratory (iBMLab), Department of Psychology, University of Fribourg, Fribourg, Switzerland*

⁴*National Technical Institute for Deaf/Rochester Institute of Technology, Rochester, New York, USA*

** These first authors contributed equally to this work*

*This study was supported by the Swiss National Science Foundation (n° 100014_156490/1) awarded to Roberto Caldara***Abstract**

We live in a world surrounded by rich dynamic multisensory signals. Hearing individuals rapidly and effectively integrate multimodal signals to decode biologically relevant facial expressions of emotion. Yet, it remains unclear how facial expressions are decoded in deaf adults in the absence of an auditory sensory channel. We thus compared early and profoundly deaf signers ($n = 46$) with hearing non-signers ($n = 48$) on a psychophysical task designed to quantify their recognition performance for the six basic facial expressions of emotion. Using neutral-to-expression image morphs and noise-to-full signal images, we quantified the *intensity* and *signal* levels required by observers to achieve expression recognition. Using Bayesian modelling, we found that deaf observers require more signal and intensity to recognize disgust, while reaching comparable performance for the remaining expressions. Our results provide a robust benchmark for the intensity and signal use in deafness and

novel insights into the differential coding of facial expressions of emotion between hearing and deaf individuals.

Introduction

As a social species, one crucial ability for survival in our social environment is the effective communication of social information. We acquire the ability to understand others and express our inner feelings long before language develops. Indeed, nonverbal communication is a major part of our social, interpersonal interaction. It conveys a rich set of information, which is at times beyond the limits of human language. One such social information code is revealed by our facial expressions (Jack and Schyns, 2015), which are influenced by culture (Jack et al, 2009; 2012a; 2012b – for a review see Caldara, 2017) from an early stage of life (Geangu et al., 2016). The ability to decode emotional cues in our social environment is essential for normative social functioning. Effective representations of our own and others' internal affective states modelled through facial expressions, and accurately expressing these states through facial muscle movements, plays a central role in defining healthy social relations and well-being (e.g. Carton, Kessler, & Pape, 1999; Feldman, Philippot, & Custrini, 1991; Izard, Fine, Schultz, Mostow, Ackerman, et al., 2001; Nowicki & Duke, 1992).

The ability to decode and recognize facial expressions of emotion has been little studied in deaf individuals, considered by many as a distinct cultural group (Jones, 2002). In deaf communication, faces have a special status. For example, in sign language facial expression provides not only emotional but also grammatical and syntactic markers (Brentari & Crossley, 2002; Liddell, 2003; Reilly, McIntire, & Seago, 1992; Reilly & Bellugi, 1996). The face and its expressions can also function as intensity markers, and the same sign can have different meanings depending on the facial expression. Sign language communication therefore requires a specific ability to process facial expressions and to differentiate syntactic facial expressions from emotional facial expressions. Moreover, in deaf signers, syntactic and emotional facial expressions are processed by different cortical regions (Corina, Bellugi, & Reilly, 1999; McCullough, Emmorey, & Sereno, 2005). Surprisingly, as might otherwise have been predicted, Grossman and Kegl (2007) did not observe better categorization performance in deaf signers compared to hearing non-signers for both emotional and linguistic dynamic facial expressions. Instead, they found that hearing non-signers were better than deaf signers at categorizing facial expressions from interrogative *yes/no* and *Wh questions* (e.g.,

Where? What? Who? When?) in American Sign Language (ASL). Again, as might otherwise have been predicted, another study (Campbell, Woll, Benson, & Wallace, 1999) has shown that users of British Sign Language (deaf or hearing) were not better than hearing non-signers in categorizing facial expressions from BSL interrogative sentences (i.e., yes/no vs wh- question), or more common facial expressions from both BSL and emotional expressions (e.g. puzzled and surprised faces).

The absence of voice tone information could also change a deaf individual's representation of emotional expressions. An emotional response to a specific event often includes an association between a facial expression and a sound. For example, the expression of fear is frequently associated with a loud vocal expression. Fear already captures attention efficiently at a very early age (e.g., Bayet et al., 2017), and elicits stronger identity neural coding in adults compared to other expressions from the visual signal alone (e.g., Turano et al., 2017). In the classic shower scene from Alfred Hitchcock's *Psycho* (1960), a hearing person will immediately associate the terrified face of Janet Leigh with the loud scream heard seconds before she is stabbed in the shower. Different studies have indeed shown that multisensory information is integrated during the processing of affective facial expressions (Campanella & Belin, 2007; Collignon et al., 2008), and that this multisensory integration is likely to undergo perceptual narrowing (Lewkowicz, 2014; Lewkowicz & Ghazanfar, 2009).

Therefore, in the absence of auditory information, it is possible that there is a difference in the development and sensitivity to facial expressions between deaf and hearing individuals. However, in an assessment of the emotional valence of different static facial expressions, Watanabe, Matsuda, Nishioka, and Namatame (2011) did not observe any differences in the judgements made by deaf and hearing participants. In another study using potentially more ecological stimuli (Jones, Gutierrez and Ludlow, 2017), the authors explored emotion recognition in deaf and hearing children for both static and moving faces (study 1) and for different intensities of static emotion (study 2). Deaf children recognized facial expressions better in the dynamic compared to the static condition, whereas no difference was found between conditions for the control group of hearing children. Moreover, both groups similarly benefited from greater performance when the expressions were higher in emotion intensity. In both studies 1 and 2, deaf and hearing children showed similar performance across emotions, with the exception of disgust for which deaf children had fewer correct responses. The

differences in results between these developmental studies are most likely based on task differences as Watanabe, Matsuda, Nishioka, and Namatame (2011) investigated emotional valence, whereas Jones, Gutierrez and Ludlow (2017) investigated emotion recognition. However, similar emotion recognition tests with adult populations have not revealed any differences in recognition between deaf and hearing observers (Grossman and Kegl, 2007). The absence of differences in the adult population could reflect the possibility that the mental representations used to decode facial expressions of emotion are similar in both deaf and hearing adults; or the measures used in previous studies could lack sufficient sensitivity to uncover any differences, at least within the adult population.

To obtain a sensitive measure of facial expression recognition performance, we introduce a psychophysical approach. This approach provides a precise measure of recognition performance as the quantity of signal (a facial expression of emotion at its fullest intensity modified with random image noise) or intensity (a neutral expression to full intensity emotional expression) is parametrically manipulated. We predicted that in the absence of auditory information, it is possible that there is a difference in the sensitivity to facial expressions between deaf and hearing individuals. Specifically, we predicted that deaf signers would have greater sensitivity to the facial expression visual stimuli used in the signal and intensity tasks and therefore show better recognition performance than the hearing non-signers. Finally, we tested recognition performance for what are commonly referred to as the six basic emotions. The basic emotions (i.e., anger, fear, disgust, happiness, sadness, and surprise) have been described as basic because they are thought to be universally recognised, however this belief has now been strongly contested (e.g. Crivelli, Russell, Jarillo, & Fernández-Dols, 2016; Gendron, Roberson, van der Vyver, & Barrett, 2014; Jack, Garrod, Yu, Caldara & Schyns, 2012a; Jack, Sun, Delis, Garrod, & Schyns, 2016). Conventionally, these six expressions are most widely studied in facial expression recognition research and were therefore selected for this study. Due to the mixed results on recognition performance in deaf and hearing cohorts described in the literature earlier, we did not have a prediction that performance between these groups would be better for a specific emotion. We uniquely predicted, as described above, that deaf signers would have greater sensitivity to the facial expression visual stimuli used in the signal and intensity tasks and therefore show better recognition performance than the hearing non-signers.

Method and analysis

Participants:

We tested a total of 46 deaf signers (26 females), who all had severe to profound hearing loss (dB loss > 70) from birth or in the first years of their life, and who were all native or early ASL signers (before the age of 5 years old). Among the 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). The age range of the deaf participants was between 18 and 30 years (the mean age was 21.70 with a standard deviation of 2.35). In addition, 48 hearing participants (29 females) with no knowledge of any sign language were tested. The age range of the hearing participants was between 18 and 31 years (the mean age was 21.44 with a standard deviation of 3.30). All participants had normal or corrected-to-normal vision. The participants were students from the Rochester Institute of Technology and the National Technical Institute for the Deaf and received \$10 for their participation. The local Institutional Review Board at Rochester Institute of Technology approved this study, and all participants provided written informed consent.

Materials

The stimuli and paradigm were the same as those previously used to test recognition differences in typically developing children (Rodger, Lao, & Caldara, 2018). Facial stimuli expressing each of the six basic emotions (i.e., anger, fear, disgust, happiness, sadness, and surprise) and a neutral expression were selected from the Karolinska Directed Emotional Faces database (KDEF, Lundqvist, Flykt, & Öhman, 1998). Images were scaled to 256 x 256 pixels and mapped to grey scale. For the intensity condition, eight identities (4 females) were chosen. We used Abrosoft FantaMorph software to create morphs of 100 increments for each identity and emotional expression, ranging from a 1% morph of a neutral face to a 100% expressive face (original image). The total number of images used was 4800 (8 identities x 6 expressions x 100 increments). For the signal condition, the stimuli consisted of 252 images from the KDEF database comprising 36 distinct identities (18 females) each displaying six facial expressions and a neutral one. Example stimuli of different expression intensities

and signal strengths are shown in Figure 1. Participants only viewed images at the intensities or signal strengths calculated by the QUEST (Watson & Pelli, 1983) procedure. All images were cropped around the face to remove distinctive hairstyles using Adobe Photoshop, and were aligned along the eyes and mouth using Psychomorph software (Tiddeman, Burt & Perrett, 2001). Images were also normalized for contrast and luminance using the SHINE toolbox (Willenbockel, Sadr, Fiset, Horne, Gosselin, & Tanaka, 2010) in MATLAB 7.10.0 and displayed on an 800 x 600 grey background at a distance of 50 cm subtending $10^\circ \times 14^\circ$ to simulate a natural viewing distance during social interaction (Hall, 1966). The stimuli were presented using the Psychophysics toolbox (PTB-3) with MATLAB 7.10.0 and QUEST, a Bayesian adaptive psychometric method (described below) to estimate the level of stimuli strength (intensity or signal) for each trial.

Procedure

Participants were asked to label how the person in the picture was feeling by pressing a key on the keyboard corresponding to each of the emotions. We instructed the participants to respond as accurately as they could, as reaction time was not important for the current task. Participants could also press the space bar to indicate “I don’t know or uncertain”. Participants were given as much time as they needed to familiarise themselves with the response keys before beginning the experiment. Unknown to the participants, the first six trials were practice trials to familiarise the procedure and did not count towards the final threshold estimations. The experimental trials therefore followed seamlessly without distinction from the practice trials. Each trial began with a fixation cross for 500 ms, followed by a face stimulus presented for 500 ms. The displayed intensity or signal strength was provided by the QUEST psychometric procedure (described below), and followed by a mask of random noise until a response was given (see Figure 2 for an illustrated example of a trial). The expressions were randomly displayed. Once the threshold of a specific expression was estimated by the QUEST procedure, that particular expression was no longer displayed and only the remaining expressions were presented to the participants. As a consequence, the number of trials for each participant varied as a function of the QUEST procedure. The intensity and signal conditions were

randomised for each participant and the whole experiment lasted about 30 minutes. For deaf participants, instructions were both written and signed by the experimenter.

The QUEST Bayesian adaptive psychometric procedure

We implemented the same QUEST procedure as previously used in Rodger, Vizioli, Ouyang, and Caldara (2015) and Rodger et al. (2018). It is an adaptive staircase procedure based on a psychometric function to establish an observer's threshold sensitivity to some physical measure of a stimulus (e.g., stimulus strength, Watson & Pelli, 1983). The obtained threshold can be considered as a measure of how effectively an observer can discriminate a set of stimuli. Here, we tested threshold sensitivity for (1) intensity and (2) signal of emotional expressions between deaf signers and hearing non-signers. QUEST obtains thresholds by adaptively presenting a sequence of stimuli according to the observer's previous responses. For example, if the observer incorrectly labels the expression of happiness, the subsequent presentation of happiness will contain more signal or intensity. Conversely, if the expression is correctly labelled, the next presentation of the same expression will contain less signal or intensity. Thus, this adaptive staircase method is an efficient way to determine the individual perceptual threshold for a type of stimulus (here facial expression), as the represented stimuli are adaptively narrowed to the true underlying threshold.

The QUEST paradigm was implemented in MATLAB 7.10.0 with the Psychophysics Toolbox (PTB-3). We extended the original binary response in QUEST to parametrically determine an observer's perceptual threshold for discriminating each of the six emotional expressions. The final estimated threshold is determined as the intensity or signal strength when the participant maintains performance at 75%. In this way, equal performance is maintained across observers. For the intensity condition, we implemented one QUEST procedure with an initial expression intensity of 30%. This intensity was selected since by nature, 50% intensity denotes an image morph of 50% neutral and 50% expressive face stimuli, so the initial value should be below this level of morph. For the signal condition, we implemented three QUEST procedures with different initial stimulus strengths (60%, 40%, and 20%) to prevent possible bias in the final estimation towards the direction of the initial

value. The QUEST procedure terminates for an expression after three consecutive correct or incorrect trials in which the intensity or signal strength standard deviations are less than 0.025.

Threshold detection

To calculate the individual thresholds obtained by the QUEST procedure, we computed the mean and the standard deviation of the estimated threshold from the QUEST posterior probability density function (pdf) using the *QuestMean* and *QuestSd* function from PTB3 (Pelli, 1987; King-Smith et al. 1994). In the signal task where multiple QUEST procedures were employed, we computed the arithmetic mean to get the final threshold estimation, and the quadratic mean of the standard deviations to get the error of the estimated threshold (based on the Gaussian assumption of the estimation). In a previous paper (Rodger et al., 2015) we used the intensity of the last trial from the QUEST procedure as the threshold estimation. However, since some participants could not achieve 75% recognition performance even when intensity or signal was at the maximum (100%), the previous calculation returned a ceiling value of 100% (especially for some expressions, e.g., fear. See supplementary figure in Rodger et al., 2015). Using the new computation approach, the threshold instead returned values above 100%.

Statistical Modelling

Data analysis was performed in Python using Jupyter Notebook. Bayesian modelling was performed using PyMC3 version 3.2, and the results were displayed using Seaborn and Matplotlib. The full model parameterization is displayed below:

```

Intercept
  C ~ Normal(0.5, 1)
Coefficient (for Fixed effect)
  β ~ Normal(0, 1)
Random effect (for j subject)
  σ12 ~ HalfCauchy(5.)
  bsubjectj ~ Normal(0, σ12)
Latent threshold
  μ = Xβ + bsubjectj
  σ22 ~ HalfCauchy(2.5)
  ν ~ HalfNormal(10.)
  τij ~ Student-T(μ, σ22, ν)
Observed thresholds for each subject
  estimated_meanij ~ Normal(τij, estimated_stdij)
Imputed model of censored threshold
  right_censored ~ Student-T(μk1, σ22, ν) for k1 estimated_meanij > 1
  left_censored ~ Student-T(μk2, σ22, ν) for k2 estimated_meanij < 0

```

We modelled the estimated threshold along with the estimation uncertainty from the QUEST with a Bayesian Hierarchical Censored Model. The threshold of each expression and each task from one participant is modelled as a linear function of the full factorial of Group * Expression * Task, with the intercept of each subject as a random effect. It is equivalent to a linear mixed model with random intercept. Moreover, to account for the estimation error from the QUEST procedure, the estimated threshold is modelled as a realization of the latent (true) threshold with a standard deviation that is the same as the estimated standard deviation output from the QUEST procedure. Importantly, as explained above, since the presented intensity or signal information are limited to the range [0, 1], a threshold estimated above 1 is less reliable as we can never directly observe and test this estimate. Thus, to account for these limitations, we added a penalization to the model log-likelihood using a censored variable representing the threshold estimated above 1 or below 0. We implemented an imputed censored model, where estimated values outside of [0, 1] are modelled as a set of additional random values that would have been censored. Thus, each censored observation introduces a random variable that would be added to the model log-likelihood.

A set of simpler alternative models were also tested: general linear model (no random effect, no latent parameters, and no additional censoring term), linear mixed model (no latent parameters and

no censoring term), linear mixed model with latent parameters (but no censoring term), and the full model as described above. It is worth noting that the estimation was highly similar across all models, but the full model yields the highest leave-one-out cross-validation score and the highest widely available information criterion (Vehtari et al., 2015; Spiegelhalter et al., 2002). All the tested models and model comparisons are shown in the supplementary notebook.

The probabilistic model was built using PyMC3 and we sampled from the posterior distribution using NUTS. We ran four MCMC chains with 3000 samples each; the first 1000 samples were used for tuning the mass matrix and step size for NUTS, and were discarded following this. Model convergence was diagnosed by computing Gelman and Rubin's convergence diagnostic (R-hat, 1992), examining the effective sample size, inspecting the mixing of the traces, and checking whether there is any divergent sample that has been returned from the sampler. Inferences were performed by computing directly on the posterior samples.

Results

The descriptive results are shown in Figure 3. Both groups of participants showed similar recognition threshold means for each of the emotions. For example, happy was the easiest to recognise, as the estimated threshold was the lowest across all expressions. The most difficult expression was fear, with an estimated threshold consistently over 1. The mean recognition threshold for each group and each expression can be found in supplementary table 1.

Posterior distribution and estimation of the parameters of the Bayesian Hierarchical Censored Model are reported in supplementary figure 1 and supplementary table 2. The estimated threshold (i.e., overall intercept) is 0.642 [0.576, 0.706], in brackets is the 95% highest probability density interval (95% HPD). At the group level, the deaf signers require a similar level of intensity/signal overall compared to hearing non-signers, as the offset compared to the hearing non-signers is estimated at 0.002 [-0.090, 0.090]. There are non-zero estimates of the Group * Expression * Task interaction terms from the linear equation part of the model, which we further quantified by computing the posterior conditional mean and performing a group comparison within each expression.

The posterior distribution of the latent threshold for each group, task, and expression is shown in figure 4 and supplementary table 3. The two groups of observers do not show the same threshold estimation within each task and expression, as the posterior distributions are not completely overlapping with each other. To demonstrate this difference, we computed the posterior contrast between deaf and hearing observers for each expression and task (figure 5). We report the contrast of deaf minus hearing participants. Interestingly, we found that deaf observers require more intensity and signal than the hearing group to recognise the expression of *Disgust* accurately. The deaf observers need 0.130 [0.039, 0.219] more intensity and 0.115 [0.026, 0.198] more signal than the hearing participants. For the other expressions, the estimated thresholds for both tasks are similar for both groups of observers. Deaf observers also show a small increase of signal threshold for *Sadness* (0.051 [-0.038, 0.136]) and *Surprise* (0.056 [-0.026, 0.137]). However, as these latter estimated differences overlap with zero, more information is needed to draw conclusions about these conditions.

Discussion

The aim of this study was to assess facial expression recognition in deaf signers and hearing non-signers using a psychophysical method to measure the quantity of intensity and signal needed to recognise an expression. The sensitivity of this method combined with a large sample enabled us to obtain interesting and reliable observations about facial expression coding in deaf signers.

Overall, our results suggest that deaf signers' facial expression coding does not differ from hearing non-signers in the quantity of signal and intensity required to recognise the basic expressions, with the exception of disgust. For both deaf and hearing participants, the expression of happiness had the lowest signal and intensity thresholds and was therefore the easiest expression to recognise, whereas fear had the highest thresholds. The same pattern of results for happiness and fear has been reported in previous developmental studies using the same paradigm, which also tested hearing adults (Rodger et al., 2015; Rodger et al., 2018). Further studies of emotion recognition in deaf children are necessary to further understand the developmental trajectory of emotion processing in the deaf population, and more specifically potential reasons for the difference found here in disgust recognition between deaf and hearing groups. Many studies report poorer performances in deaf

children compared to age-matched hearing children in the labelling and categorisation of facial expressions of emotion (e.g., Dyck, Farrugia, Shochet, and Homes-Brown, 2004; Ludlow, Heaton, Rosset, Hills, 2010; Most & Michaelis, 2012). These differences have been observed in young children from 6 years of age up to 18 years of age. However, other studies report no significant or subtle differences between deaf and hearing children in similar tasks involving facial emotion processing (e.g. Hosie et al., 1998, Jones et al., 2017, Ziv, Most and Cohen, 2013). The heterogeneity of these results could be linked to different factors related to the study design, or to the participants characteristics such as age, use of hearing aids, or language abilities (e.g. sign language as a primary language or not). They also suggest that more research is necessary to better understand the developmental trajectory of facial emotion processing in deaf children. Moreover, studies including both children and adults in the same experimental paradigm would be beneficial.

Disgust was the unique expression showing differences in recognition thresholds between the deaf signing and hearing non-signing participants for both the signal and intensity conditions. Deaf signers had higher recognition thresholds than hearing non-signers for both conditions. Therefore, deaf signers needed more facial information than their hearing counterparts to recognise disgust in comparison to the other facial expressions of emotion at a level of 75% performance. Atypical performance for disgust recognition was similarly reported in deaf children in a recent study of deaf and hearing children's recognition of the six basic emotions (Jones, Gutierrez & Ludlow, 2017). In this study both deaf and hearing children showed the poorest performance for disgust and fear recognition overall, however deaf children showed poorer performance for disgust recognition compared to hearing children for both moving faces and faces that varied in emotional intensity. Jones et al. (2017) posited that deaf children's concepts of disgust in comparison to hearing children's may be less developed as they have been less exposed to conversations about emotions which may impact emotion recognition ability. However, it is not clear why the recognition of disgust uniquely is affected within this context. Another possibility for this atypical performance in disgust recognition is that deaf people need more signal or intensity information because disgust may overlap with other facial expressions like confusion, frustration, or uncertainty. It is possible that deaf people experience

such expressions more frequently, so their perceptual categories for disgust and other expressions which are similar to them may be different to those of hearing people.

Performance for disgust recognition in another developmental study of deaf and hearing children was again lowest across all six basic expressions for both deaf and hearing children (Hosie, Gray, Russell, Scott, and Hunter 1998). However, accuracy for disgust was higher for deaf children compared to hearing children. Nonetheless, this result is not a direct contradiction to our result with adult observers, as higher thresholds in our study were necessary to maintain a high level of performance (75% accuracy). Therefore, despite deaf children showing better scores than hearing children, accuracy overall was low in comparison to the 75% threshold in our study. High thresholds and accuracy are therefore not equivalent. In order to further address the question of atypical performance for disgust within deaf populations, a full parametric design is necessary to map out the complete psychometric function. While it is clear that the facial expression of disgust is special in deaf signers (both adults and children), the reason for this atypical coding remains unclear. Future studies are necessary to establish whether this difference relates to deafness or sign language experience. Further, now that this deficit for disgust recognition in the deaf population has been detected, clearly there is a need for training in disgust recognition. Emotion recognition training in typically developing children has been shown to improve recognition performance (Pollux, Hall, & Guo, 2014). One pilot study with deaf children showed training improvements in emotion understanding but not facial expression recognition, so further studies in this area are clearly necessary (Dyck & Denver, 2003).

The absence of difference in recognition thresholds between deaf and hearing participants does not necessarily mean that the special status of faces for the deaf population has no impact on the coding of facial expressions. Indeed, the current study only explored the coding of emotional facial expressions and not other types of facial expressions. Since linguistic and emotional expressions are processed differently in deaf signers (Corina et al., 1999; McCullough et al., 2005), it is possible that only facial expressions used in sign language (i.e., linguistic facial expressions) are processed differently in deaf signers. It would therefore be necessary to extend research on facial expressions

with a larger variety of expressions, in particular with both emotional and linguistic expressions in both hearing and deaf signers. In addition, it would also be interesting to clarify whether deaf adults would benefit from the presentation of dynamic emotional expressions over static images. We very recently showed that the performance for the recognition of *static* as compared to *dynamic* expressions was notably less effective in fragile face processing systems, such as brain-damaged patients (Ricoz, Jack, Garrod, Schyns, & Caldara, 2015), and elderly adults (Ricoz, Lao, Pascalis, & Caldara, 2018) by using a database of stimuli controlled for the amount of *low-level* information carried over time (Gold et al., 2014). As sign language is dynamic, we could expect similar results in the deaf population, a question that needs to be addressed in a future study.

In the current study, we aimed to provide a benchmark for the recognition of facial expressions of emotion in deaf signers, using young adult hearing participants as a baseline. Thus, instead of performing a conventional series of t-tests for all possible combinations of each groups' expression recognition measures corrected for multiple comparisons, we decided to improve the estimation of accuracy for each condition by properly accounting for measurement errors using weakly informative priors. With this novel threshold paradigm, we take advantage of the uncertainty of the individual threshold estimation output by the adaptive maximum likelihood procedure. By using a hierarchical mixed-effect model, we account for the individual differences and provide a more precise group estimation. Moreover, we account for the bias in the threshold estimations that are higher than 1 with a censoring likelihood as a penalty. We hope that future studies with similar paradigms can take advantage of the model estimation we provide here, to construct more informative models.

To conclude, this study offers new insights into the coding of facial expressions of emotion in deaf signers. Despite the central importance of facial expressions in deaf communication, overall we observe similar intensity and signal thresholds for both deaf-signers and hearing non-signers for facial expression recognition. Further studies are necessary to examine potential differences in the recognition of both emotional and linguistic expressions, in both hearing and deaf signers. Exceptionally, recognition performance for disgust was poorer for deaf-signers compared to hearing

non-signers. The atypical processing of disgust expressions has similarly been reported in deaf children. Future studies should try to establish how the experience of deafness and sign language may interfere in the construction and understanding of expressions of disgust. One consideration is that the Action Units for sign language may overlap with the Action Units for emotions, but not uniformly. That is, the Action Units for disgust may incorporate effectors that are more similar to those employed in sign language than is the case for other emotional expressions. Our results provide a reliable benchmark for the intensity and signal thresholds used for expression recognition in young deaf adults.

References

- Bayet, L., Quinn, P. C., Laboissière, R., Caldara, R., Lee, K., & Pascalis, O. (2017). Fearful but not Happy Expressions Boost Face Detection in Human Infants. *Proceedings of the Royal Society of London B*, *284* (1862), 1-9.
- Brentari, D., & Crossley, L. (2002). Prosody on the hands and face: Evidence from American Sign Language. *Sign Language & Linguistics*, *5*, 105–130. <https://doi.org/10.1075/sll.5.2.03bre>
- Caldara, R. (2017). Culture Reveals a Flexible System for Face Processing. *Current Directions in Psychological Science*, *26*(3), 249-255.
- Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Integrating Face and Voice in Person Perception*, *11*(12), 535–543. <https://doi.org/10.1007/978-1-4614-3585-3>
- Campbell, R., Woll, B., Benson, P. J., & Wallace, S. B. (1999). Categorical Perception of Face Actions: Their Role in Sign Language and in Communicative Facial Displays. *The quarterly Journal of Experimental Psychology*, *52*(A), 67–95
- Carton, J.S., Kessler E.A., & Pape C.L. (1999). Nonverbal decoding skills and relationship well-being in adults. *Journal of Nonverbal Behavior*, *23*(9), 1–100.
- Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research*, *1242*, 126–135. <https://doi.org/10.1016/j.brainres.2008.04.023>
- Corina, D. P., Bellugi, U., & Reilly, J. (1999). Neuropsychological studies of linguistic and affective facial expressions in deaf signers. *Language and Speech*, (2–3), 307–331. <https://doi.org/10.1177/00238309990420020801>
- Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J. M. (2016). The fear gasping face as a threat display in a Melanesian society. *Proceedings of the National Academy of Sciences, USA*, *113*(44), 12403-12407.
- Dyck, M., & Denver, E. (2003). Can the emotion recognition ability of deaf children be enhanced? A

- pilot study. *Journal of Deaf Studies and Deaf Education*, 8, 348–356.
- 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? *Journal of Child Psychology and Psychiatry*, 45(5), 789-800.
- Ekman, P., & Friesen, W. V. (1978). Facial action coding system. Palo Alto, CA
- Feldman, R. S., Philippot, P., & Custrini, R. J. (1991). Social competence and nonverbal behavior. In R. S. Feldman & B. Rime (Eds.), *Fundamentals of nonverbal behavior* (pp. 329- 350). New York: Cambridge University Press.
- Geangu, E., Ichikawa, H., Lao, J., Kanazawa, S., Yamaguchi, M. K., Caldara, R., & Turati, C. (2016). Culture shapes 7-month-olds perceptual strategies in discriminating facial expressions of emotion. *Current Biology*, 26(14), R663-664.
- Gelman, A and Rubin, DB (1992) Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, 457-511.
- Gendron, M., Roberson, D., van der Vyver, J. M., and Barrett, L. F. (2014). Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. *Emotion* 14, 251–262.
- 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. *Journal of Vision*, 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. *Journal of Nonverbal Behavior*, 31(1), 23–38. <https://doi.org/10.1007/s10919-006-0022-2>
- Hall, E. (1966). *The Hidden Dimension*. Garden City, NY: Doubleday.
- Hess, U., Blairy, S., & Kleck, R. E. (1997). The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior*, 21, 241–257.
- 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 emotional labels. *Motivation & Emotion*, 22, 293–313.
- Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman B., & Youngstrom, E. (2001). Emotion Knowledge as a Predictor of Social Behavior and Academic Competence in Children at Risk. *Psychological Science*, 12(1), 18-23.
- Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, 19(18), 1543-1548.
- Jack, R., Garrod, O., Yu, H., Caldara, R., & Schyns, P. G. (2012a). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 109 (19), 7241-4.
- Jack, R., Caldara, R., & Schyns, P. (2012b). Internal representations reveal cultural diversity in expectations of facial expressions of emotions. *Journal of Experimental Psychology: General*,

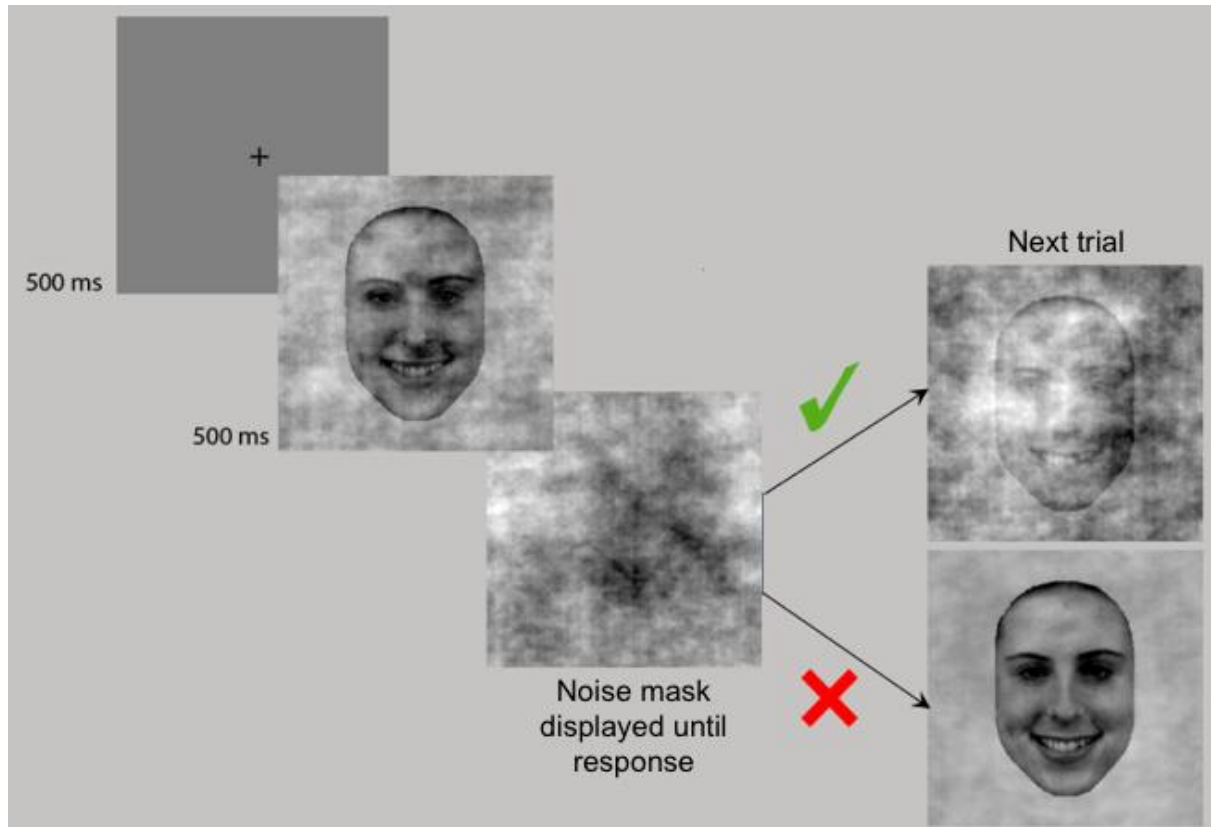
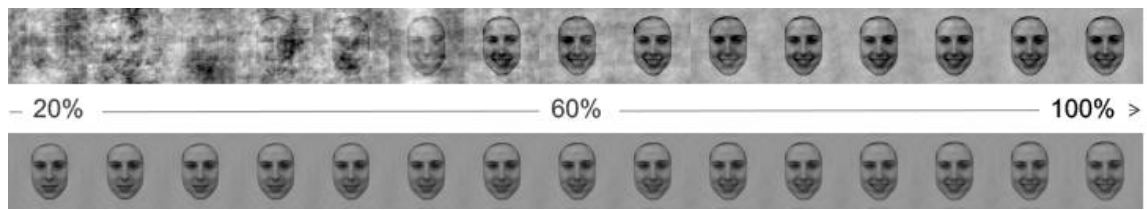
141 (1), 19-25

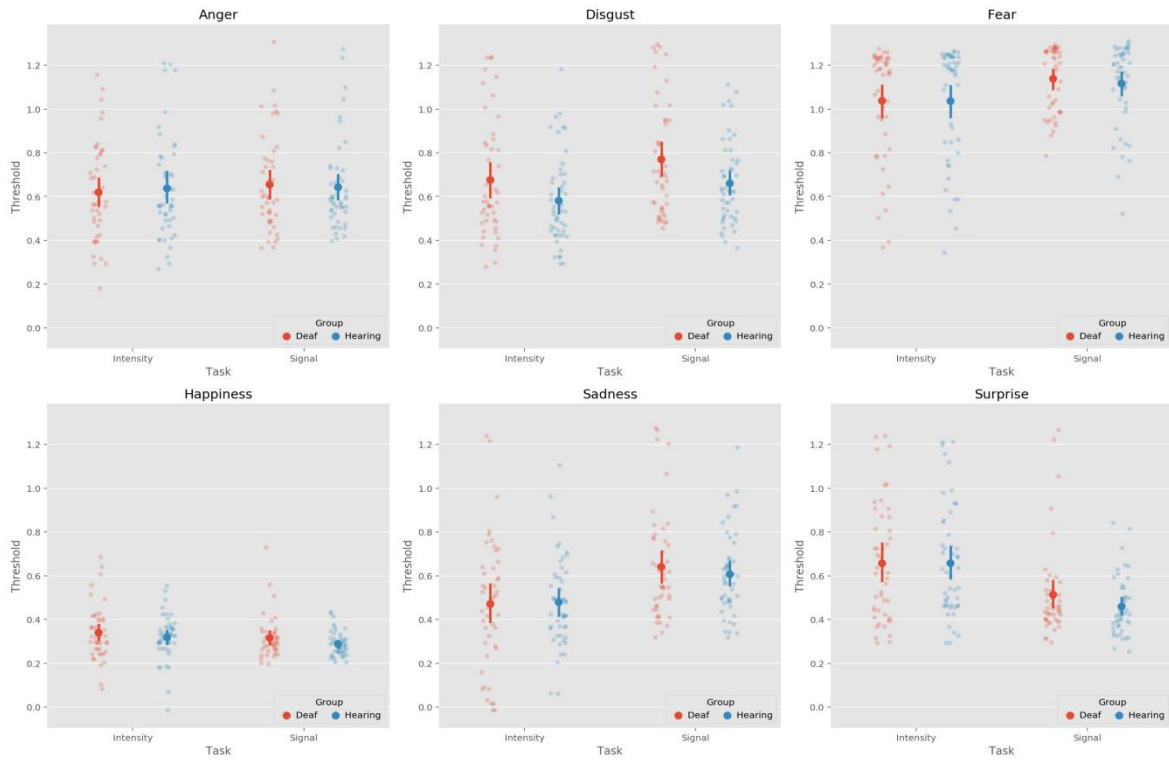
- Jack, R. E., & Schyns, P. G. (2015). The Human Face as a Dynamic Tool for Social Communication. *Current Biology*, 25(14), R621–R634. <https://doi.org/10.1016/j.cub.2015.05.052>
- Jack, R. E., Sun, W., Delis, I., Garrod, O. G. B. and Schyns, P. G. (2016) Four not six: revealing culturally common facial expressions of emotion. *Journal of Experimental Psychology: General*, 145(6), pp. 708-730.
- 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. *Cognition and Emotion*, 9931, 1–14. <https://doi.org/10.1080/02699931.2017.1289894>
- Jones, M.A. (2002). Deafness as culture: A psychosocial perspective. *Disability Studies Quarterly*, 22(2), 51-60.
- King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Research*, 34 (7), 885-912.
- Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental Psychobiology*, 56(2), 292–315. <https://doi.org/10.1002/dev.21197>
- Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, 13(11), 470–478. <https://doi.org/10.1016/j.tics.2009.08.004>
- Liddell, S. K. (2003). *Grammar, gesture, and meaning in American Sign Language* (Cambridge). New-York.
- Ludlow, A., Heaton, P., Rosset, D., Hills, P., & Deruelle, C. (2010). Emotion Recognition in children with profound and severe deafness : Do they have a deficit in perceptual processing? *Journal of Clinical and Experimental Neuropsychology*, 32, 923-28.
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces - KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9.
- McCullough, S., Emmorey, K., & Sereno, M. (2005). Neural organization for recognition of grammatical and emotional facial expressions in deaf ASL signers and hearing nonsigners. *Cognitive Brain Research*, 22(2), 193–203. <https://doi.org/10.1016/j.cogbrainres.2004.08.012>
- Most, T., & Michaelis, H. (2012). Auditory, visual, and auditory– visual perceptions of emotions by young children with hearing loss versus children with normal hearing. *Journal of Speech Language and Hearing Research*, 55(4), 1148–1162
- Nowicki, S., Jr., & Duke, M. P. (1992). The association of children's nonverbal decoding abilities with their popularity, locus of control, and academic achievement. *The Journal*

of *Genetic Psychology*, 153, 385-394

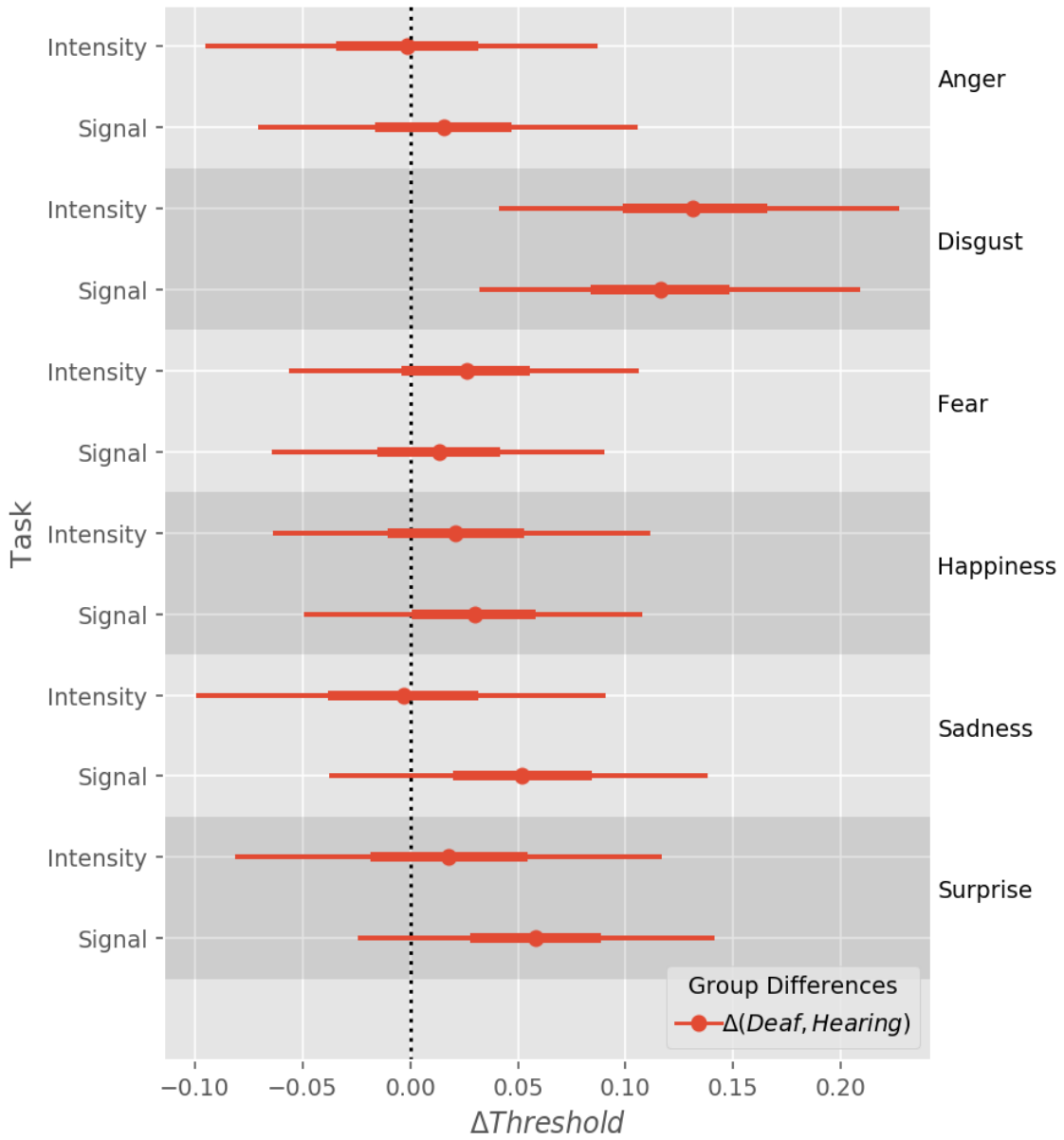
- Pelli, D. G. (1987). *The ideal psychometric procedure*. *Investigative Ophthalmology & Visual Science*, 28, 366.
- Reilly, & Bellugi, U. (1996). Competition on the face: affect and language in ASL motherese. *Journal of Child Language*, 23(1), 219–239. <https://doi.org/10.1017/S0305000900010163>
- Reilly, J., McIntire, M. L., & Seago, H. (1992). Affective prosody in American Sign Language. *Sign Language Studies*, 75, 113–128.
- 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. *Journal of Vision*, 18(9), 1-27.
- Richoz, A.-R., Jack, R. E., Garrod, O. G., Schyns, P. G., & Caldara, R. (2015). Reconstructing dynamic mental models of facial expressions in prosopagnosia reveals distinct representations for identity and expression. *Cortex*, 65, 50-64.
- Rodger, H., Lao, J., & Caldara, R. (2018). Quantifying facial expression signal and intensity use during development. *Journal of Experimental Child Psychology*, 174, 41-59.
- Rodger H., Vizioli L., Ouyang X., & Caldara R. (2015). Mapping the development of facial expression recognition. *Developmental Science*, 18(6), 926-93
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B*, 64(4), 583-639.
- Tiddeman, B., Burt, M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications*, 21, 42–50.
- Turano, M. T., Lao, J., Richoz, A.-R., de Lissa, P., Degosciu, S. B. A., Viggiano, M. P., & Caldara, R. (2017). Fear boosts the early neural coding of faces. *Social Cognitive and Affective Neuroscience*, 12(12), 1959-1971.
- Vehtari, A., Gelman, A., & Gabry, J. (2016a). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 1-20.
- Watanabe, K., Matsuda, T., Nishioka, T., & Namatame, M. (2011). Eye gaze during observation of static faces in deaf people. *PloS One*, 6(2), e16919. <https://doi.org/10.1371/journal.pone.0016919>
- Watson, A., & Pelli, D. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33(2), 113–120.F
- Willenbockel, V., Sadr, J., Fiset, D., Horne, G.O., Gosselin, F., & Tanaka, J.W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior Research Methods*, 42(3), 671-684.
- Ziv, M., Most, T., & Cohen, S. (2013). Understanding of emotions and false beliefs among children versus deaf children. *Journal of Deaf Studies and Deaf Education*, 18(2), 161– 174.

Figures





95% Credible Intervals



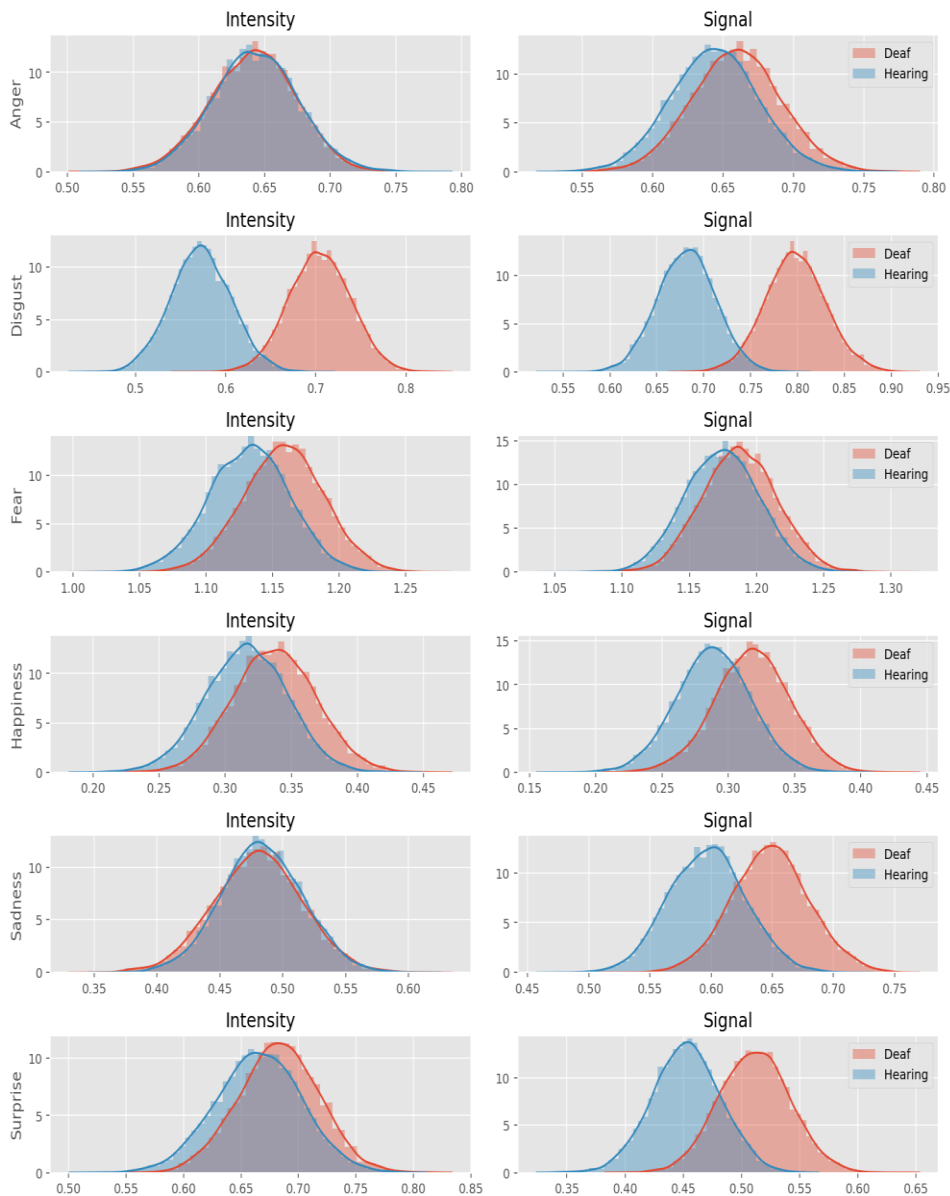


Figure Captions

Figure 1: Example stimuli from the signal and intensity conditions. In this image, the stimuli are shown in 5% increments, starting at 20% signal or intensity.

Figure 2: Example trial from the signal condition. Each trial begins with a fixation cross for 500 ms, followed by a face stimulus presentation for 500 ms. The displayed signal or intensity strength is provided by the QUEST psychometric procedure, followed by a mask of random noise until a response is made. Depending on accuracy, the next trial was followed by a face containing more (in the case of an erroneous response) or less (in the case of an accurate response) signal or intensity.

Figure 3: Descriptive statistics. Each subplot shows the threshold estimation for one facial expression of emotion. The scatter plots show the individual threshold estimations for one observer. Each group

is color coded: red for deaf signers and blue for hearing non-signers. The group mean is shown as a solid non-transparent dot on top of the scatter plot, the error bars show the 95% Bootstrap confidence interval of the mean.

Figure 4: Posterior distribution of the thresholds for 75% recognition performance for each expression and task. Groups are color coded red for deaf signers and blue for hearing non-signers.

Figure 5: Forest plot of the posterior contrast of the thresholds for each expression and task between deaf signers and hearing non-signers.

Supplementary Tables and Figures

Supplementary Table 1

Mean Expression Recognition Thresholds per group and task with 97.5% bootstrap-t Confidence Intervals

Mean Expression Recognition

Thresholds

Express ion	Task	Group	Mean	SD	97.5% btCI
Anger	Intensity	Deaf	0.619	0.224	(0.554, 0.685)
		Hearing	0.637	0.234	(0.574, 0.705)
	Signal	Deaf	0.655	0.215	(0.598, 0.716)
		Hearing	0.642	0.204	(0.587, 0.700)
Disgust	Intensity	Deaf	0.676	0.273	(0.603, 0.757)
		Hearing	0.581	0.206	(0.526, 0.641)
	Signal	Deaf	0.770	0.262	(0.691, 0.843)
		Hearing	0.660	0.191	(0.609, 0.717)
Fear	Intensity	Deaf	1.038	0.258	(0.962, 1.109)
		Hearing	1.036	0.260	(0.951, 1.108)
	Signal	Deaf	1.139	0.148	(1.096, 1.180)
		Hearing	1.116	0.187	(1.062, 1.161)
Happiness	Intensity	Deaf	0.340	0.124	(0.305, 0.377)
		Hearing	0.319	0.101	(0.291, 0.345)
	Signal	Deaf	0.316	0.093	(0.293, 0.347)
		Hearing	0.288	0.055	(0.274, 0.304)
Sadness	Intensity	Deaf	0.471	0.297	(0.384, 0.558)
		Hearing	0.479	0.208	(0.422, 0.540)
	Signal	Deaf	0.641	0.248	(0.571, 0.715)
		Hearing	0.607	0.192	(0.556, 0.659)
Surprise	Intensity	Deaf	0.657	0.270	(0.581, 0.736)
		Hearing	0.657	0.255	(0.588, 0.727)
	Signal	Deaf	0.513	0.212	(0.456, 0.577)
		Hearing	0.459	0.132	(0.425, 0.497)

Supplementary Table 2

Posterior distribution and estimation of the parameters of the Bayesian Hierarchical Censored Model.

The estimated threshold (i.e., overall intercept) is 0.642 [0.576, 0.706], in brackets is the 95% highest

probability density interval (95% HPD).

Bayesian Hierarchical Censored Model

		Mean	SD	Monte Carlo Error	HPD Interval (Lower, Upper)	Number of effective sample size	Rhat
Intercept	0.642	0.033	0.001	0.576	0.706	1139.0	1.002
Group__0	0.002	0.046	0.001	-0.090	0.090	1175.0	1.005
Task__0	0.017	0.041	0.001	-0.066	0.090	1109.0	1.002
Expression__0	0.062	0.043	0.001	-0.022	0.147	1536.0	1.001
Expression__1	0.516	0.040	0.001	0.435	0.591	1222.0	1.001
Expression__2	-0.305	0.039	0.001	-0.380	-0.226	1565.0	1.001
Expression__3	-0.162	0.042	0.001	-0.248	-0.082	1624.0	1.001
Expression__4	0.041	0.043	0.001	-0.048	0.122	1618.0	1.002
Group_Task__0	-0.016	0.056	0.002	-0.122	0.093	1126.0	1.004
Group_Expression__0	-0.132	0.059	0.001	-0.247	-0.018	1571.0	1.002
Group_Expression__1	-0.026	0.055	0.002	-0.133	0.082	1362.0	1.003
Group_Expression__2	-0.022	0.056	0.001	-0.133	0.083	1630.0	1.003
Group_Expression__3	0.003	0.057	0.001	-0.110	0.111	1658.0	1.002
Group_Expression__4	-0.019	0.059	0.001	-0.132	0.099	1628.0	1.003
Task_Expression__0	0.077	0.058	0.001	-0.038	0.189	1489.0	1.001
Task_Expression__1	0.012	0.054	0.001	-0.089	0.120	1361.0	1.002
Task_Expression__2	-0.036	0.054	0.001	-0.143	0.069	1465.0	1.001
Task_Expression__3	0.152	0.057	0.001	0.041	0.265	1618.0	1.001
Task_Expression__4	-0.189	0.057	0.001	-0.302	-0.080	1541.0	1.002
Interaction__0	0.031	0.081	0.002	-0.130	0.185	1530.0	1.001
Interaction__1	0.028	0.074	0.002	-0.109	0.178	1359.0	1.003
Interaction__2	0.008	0.074	0.002	-0.137	0.147	1557.0	1.003
Interaction__3	-0.040	0.078	0.002	-0.190	0.114	1715.0	1.002
Interaction__4	-0.024	0.079	0.002	-0.186	0.127	1570.0	1.003

Supplementary Table 3

Posterior distribution of the latent threshold for each group, task, and expression.

	Mean Recognition Accuracy			
	(%)			
	Mean	SD	HPD Interval	
			(Lower,	Upper)
Anger_Deaf_Intensity	0.642	0.033	0.576	0.706
Anger_Hearing_Intensity	0.644	0.033	0.582	0.710
Anger_Deaf_Signal	0.659	0.032	0.597	0.723
Anger_Hearing_Signal	0.645	0.031	0.583	0.704
Disgust_Deaf_Intensity	0.704	0.035	0.633	0.768
Disgust_Hearing_Intensity	0.574	0.032	0.511	0.636
Disgust_Deaf_Signal	0.797	0.032	0.731	0.859
Disgust_Hearing_Signal	0.683	0.031	0.620	0.740
Fear_Deaf_Intensity	1.158	0.031	1.095	1.216
Fear_Hearing_Intensity	1.134	0.030	1.077	1.196

Intensity				
Fear_Deaf_Signal	1.187	0.028	1.135	1.244
Fear_Hearing_Signal	1.175	0.028	1.120	1.228
Happiness_Deaf_Intensity	0.337	0.031	0.279	0.398
Happiness_Hearing_Intensity	0.317	0.030	0.259	0.379
Happiness_Deaf_Signal	0.317	0.028	0.263	0.372
Happiness_Hearing_Signal	0.290	0.028	0.233	0.342
Sadness_Deaf_Intensity	0.480	0.035	0.410	0.546
Sadness_Hearing_Intensity	0.485	0.032	0.424	0.549
Sadness_Deaf_Signal	0.649	0.032	0.586	0.709
Sadness_Hearing_Signal	0.598	0.031	0.538	0.660
Surprise_Deaf_Intensity	0.683	0.035	0.615	0.752
Surprise_Hearing_Intensity	0.666	0.036	0.594	0.737
Surprise_Deaf_Signal	0.510	0.031	0.451	0.571
Surprise_Hearing_Signal	0.454	0.029	0.401	0.515

Supplementary Figure 1

Posterior distribution and estimation of the parameters of the Bayesian Hierarchical Censored Model

