Methods in Psychology 14 (2026) 100241

Contents lists available at ScienceDirect

Methods in Psychology

journal homepage: www.sciencedirect.com/journal/methods-in-psychology

Quantifying individual recognition sensitivity to static and dynamic facial expressions

L. Stacchi [a,1], F. Poncet [a,1], V. Benedetti [a,b], R. Caldara [a], A.-R. Richoz [a,*]

[a] Eye and Brain Mapping Laboratory (iBMLab), Department of Psychology, University of Fribourg, Fribourg, Switzerland
[b] Cognitive Psychophysiology Lab (CogPsyLab), Department of Neuroscience, Psychology, Drug Research and Child's Health (NEUROFARBA), University of Florence, Florence, Italy

ARTICLE INFO

Keywords:
Facial expression recognition
Static-dynamic individual differences
QUEST
Perceptual thresholds

ABSTRACT

Dynamic facial expressions of emotion are typically recognized more accurately than static expressions, especially among individuals with immature, vulnerable or impaired facial expression recognition (FER) systems, such as children, older adults, and individuals with clinical conditions. These findings underscore the need of assessing both dynamic and static stimulus formats and suggest that the dynamic advantage could serve as a potential individual-level marker of FER impairment. However, previous research has primarily focused on group-level effects, often overlooking critical individual differences. Here, we tested whether the QUEST threshold-seeking algorithm can efficiently estimate the minimal signal required for accurate recognition of static and dynamic facial expressions at the individual level. We also quantified the minimum number of trials needed to obtain stable threshold estimates. To evaluate its sensitivity, we compared FER thresholds across neurotypical young adults, older adults, and a well-documented case of acquired prosopagnosia.

Our findings demonstrate that the QUEST algorithm is a robust and efficient tool for rapidly estimating meaningful FER thresholds at the individual level. We also provide guidance on the minimum number of trials required to obtain stable threshold estimates and identify the facial expressions that serve as the most sensitive markers for probing the dynamic advantage. This psychophysical approach is particularly well-suited for single-subject analyses, assessment of FER in populations with limited capacities, and inclusion in comprehensive testing batteries. Collectively, this psychophysical approach enables scalable, time-efficient screening and longitudinal monitoring of FER, facilitating cross-individual and population comparisons in both research and clinical settings.

1. Introduction

Accurate decoding facial expressions of emotion (FEEs) is fundamental to effective human communication, enabling individuals to infer the mental states of others and adjust their behavior accordingly. Given its central role in social interactions, facial expression recognition (FER) has been extensively studied. While early research primarily relied on static images to probe FER, recent work has demonstrated superior recognition performance for dynamic compared with static facial expressions (for reviews, see Krumhuber et al., 2023; Alves, 2013). This benefit for dynamic relative to static FER is referred to as the dynamic advantage. This dynamic advantage is particularly evident in populations with less robust FER systems. It has for example been reported in young children, older adults (Richoz et al., 2018; but see Widen and Russell, 2015), and individuals with clinical or neuropsychological conditions including prosopagnosia (Richoz et al., 2015) and autism spectrum disorder (Tardif et al., 2007). Notably, neurotypical young adults also exhibit a dynamic advantage, but only under visually challenging conditions (Cunningham and Wallraven, 2009; Dobs et al., 2018). For instance, Richoz et al. (2024) reported that recognition accuracy in young adults was comparable for static and dynamic stimuli under optimal viewing conditions, but that dynamic stimuli yielded a clear benefit when images were degraded by phase decoherence. This pattern suggests heterogeneity in the mechanisms supporting facial expression decoding: less robust systems appear to rely more on motion-based cues than on static information. Accordingly, the

* Corresponding author. University of Fribourg Department of Psychology Faucigny 2, 1700, Fribourg, Switzerland.
E-mail address: anne-raphaelle.richoz@unifr.ch (A.-R. Richoz).
[1] Co-shared first authors.

https://doi.org/10.1016/j.metip.2026.100241
Received 10 October 2025; Received in revised form 29 January 2026; Accepted 22 February 2026
Available online 24 February 2026

magnitude of the dynamic advantage (i.e., the extent to which dynamic stimuli offer benefits compared to static ones) may serve as a functional marker of fragility in FEE decoding.

To date, however, research on the dynamic advantage has focused predominately on group-level analyses (e.g., Richoz et al., 2024), with limited attention to individual differences. Traditionally, individual differences have been considered a source of noise, which could be attenuated through group-level averaging to reveal common patterns. While this is appropriate for many research questions, this approach can also obscure meaningful information. In fact, the integration of idiosyncratic signatures in response patterns offers deeper insights into how the visual system reaches similar outcomes via distinct strategies. For instance, Stacchi et al. (2019) demonstrated that neural facial identity discrimination responses are tuned to eye movements, providing the neural system with observer-specific diagnostic information. By investigating individual differences, this study highlighted the existence of idiosyncratic rather than universal face processing mechanisms.

Beyond its theoretical implications, individual variability is also particularly relevant in clinical and diagnostic settings, where the focus is not only on whether a group exhibits a specific effect, but on whether this can be quantified in single cases and whether deviations from normative patterns can provide diagnostic insight. For example, in the context of facial expression recognition, an unusually large dynamic advantage compared to a control group might indicate specific vulnerabilities of the patient's FER system.

Nevertheless, comparing FER across populations or conditions using traditional accuracy-based approaches can be difficult. These methods typically assess recognition performance at a fixed level of task difficulty. However, selecting an appropriate difficulty level is not a trivial task. If the task is too easy or too difficult, a ceiling or floor effect might obscure individual differences or within-individual changes across conditions. This point is illustrated by recent findings from Richoz et al. (2024), who assessed static and dynamic FER accuracy in young adults using a combined uniform and adaptive signal sampling approach that varied the amount of visual information available. Their results revealed that although a dynamic advantage is present in young adults, it emerges most clearly under perceptually challenging conditions. Within the same participants, this effect becomes significantly less robust or disappears entirely under optimal viewing conditions, depending on the FEE considered.

Taken together, studies in both young and older adults (Richoz et al., 2018, 2024) suggest that reliably observing a dynamic advantage requires modulating task difficulty in a way that ensures all populations are tested under comparable challenges. This can be achieved using threshold-based approaches, which estimate the amount of visual information required to reach a fixed level of performance (i.e., perceptual threshold). Within this context, differences in FER are indexed by differences in perceptual thresholds, with more efficient FER requiring less information (i.e., lower thresholds) to achieve the same level of accuracy.

The methodology used by Richoz et al. (2024) allowed the estimation of this perceptual threshold by fitting a Weibull function to recognition performance across signal levels. However, while effective at the group-level, this approach does not yield reliable estimates at the individual level. Moreover, it requires testing numerous signal levels per condition, resulting in lengthy assessment procedures. This limitation is problematic in both experimental and clinical settings, where a rapid evaluation is essential. Efficient testing enables multiple assessments within a single testing session, facilitating comparisons across conditions, and minimizing participant fatigue. This is especially important for populations with limited tolerance for lengthy procedures, such as clinical populations, young children, and older adults. The need for efficient assessment tools becomes even more pronounced when investigating the dynamic advantage, which requires testing multiple facial expressions across both static and dynamic formats.

To address these limitations, the current study evaluates whether the QUEST algorithm (Watson and Pelli, 1983), a Bayesian adaptive psychometric method, could provide a suitable solution. The Bayesian QUEST algorithm uses participants' responses to iteratively update a prior probability distribution, with its mean corresponding to the best estimate of the perceptual threshold. Over the course of testing, if QUEST converges on a perceptual threshold, its estimate begins to oscillate around a single value across subsequent trials and the range of this fluctuation becomes smaller, indicating stabilization. This approach would allow setting the accuracy level to a low-to-moderate performance, ensuring that the FER system of any population is sufficiently challenged, thereby increasing the likelihood of detecting a dynamic advantage. Crucially, this approach would enable direct comparisons of the signal required to achieve equivalent performance across modalities, expressions and individuals.

This procedure has been widely used due to its efficiency in rapidly converging on threshold estimates, effectively reducing the duration of testing. However, its use has largely been confined to low-level visual processing research (Leek, 2001; Treutwein, 1995), with more limited application in high-level visual domains (e.g., Blais et al., 2017; Rodger et al., 2015; Wyssen et al., 2019). In this framework, QUEST has often been used to adaptively adjust stimulus difficulty to maintain a fixed level of performance, rather than as a method for precisely estimating perceptual thresholds. To date, only a handful of studies have used the QUEST algorithm to estimate FER sensitivity – specifically, the level of visual phase coherence or expression intensity required to achieve a given accuracy. These studies have examined FER across different developmental stages and in clinical populations such as deaf individuals and women with eating disorders or mixed mental disorders (Rodger et al., 2015, 2018; Wyssen et al., 2019; Stoll et al., 2019). Although these studies have provided valuable insights into FER development and clinical conditions, they have been limited to static facial expressions with a group-level approach. As such, it remains unknown whether the QUEST algorithm could also provide information on dynamic FEE processing and the dynamic advantage at the individual level.

In addition, previous implementation of the QUEST algorithm in FER research have typically stopped testing once participants provided three consecutive correct or incorrect responses and the standard deviation of the signal presented over these last trials fell below .025. While such reduced variability in stimulus intensity may suggest algorithm stabilization, it reflects changes in the input rather than in the estimate itself. A more reliable index of stabilization is the stabilization of the estimated threshold itself – specifically when the algorithm's threshold estimate remains stable within a narrow range – no more than 2 % variation – over multiple consecutive trials. Finally, a further limitation in previous research is the lack of documentation regarding the minimum number of trials needed to achieve stabilization. This factor is crucial for an efficient application of the QUEST algorithm, as requiring a large number of trials to ensure a stable estimate would undermine its effectiveness.

Given these limitations, the present study aimed to assess the suitability of the QUEST algorithm for determining, at the individual level, the quantity of signal required to reach a predefined recognition accuracy (i.e., perceptual threshold) for both static and dynamic FEEs. We tested three distinct populations: young healthy adults, older healthy adults and a single case of pure acquired prosopagnosia. Establishing the applicability of the QUEST method across a wide range of individuals is fundamental to enable future comparative studies, in both theoretical and clinical research contexts. Our study addressed three core methodological questions.

(1) Can the QUEST algorithm produce stable perceptual thresholds for static and dynamic FER at the individual level?
(2) What is the minimum number of trials required to reach stabilization?
(3) Are these thresholds sensitive to well-established effects on FER such as age-related differences and the dynamic advantage?

First, we determined whether it was possible to obtain a stable perceptual threshold for each emotion in each modality at the individual level. A perceptual threshold was considered stable when its estimate - defined as the mean of the probability distribution - fluctuated within a small range for four consecutive trials. This first step was necessary, because, unlike in low-level visual tasks where the QUEST is commonly applied, FER can be influenced by many factors, increasing the risk of noisy measurements. For instance, participants must recognize the same expression across different identities. While this allows for results generalization, it requires the visual system to flexibly adapt and filter out identity-specific variations to access the underlying expression. If FEE extraction is not efficient and robust, the same participant might provide inconsistent responses for the same expression, even at high signal levels, potentially preventing the QUEST algorithm from converging. Therefore, before conducting any further data analysis, we first examined whether the procedure could reliably produce perceptual threshold estimates across conditions and participants. A failure to do so would compromise the relevance on any interpretation of the results.

Second, we assessed the minimum number of trials required by the algorithm to reach stabilization. This step aimed to provide critical guidelines for future studies optimizing the efficiency of the QUEST method in obtaining a stable perceptual threshold, while avoiding unnecessarily long testing sessions.

Third, we tested whether the derived thresholds were sensitive to known FER effects. Specifically, we assessed the presence of a dynamic advantage and investigated whether young adults would outperform older adults, with superior performance indexed by lower perceptual thresholds (i.e., less signal required). Additionally, we explored whether the thresholds estimated for the prosopagnosic patient PS (Rossion, 2022a, 2022b) differed from those of age-matched controls. Previous studies have shown that this patient exhibits poorer recognition performance of static FEE compared to her peers (Rossion et al., 2003; Richoz et al., 2015; Fiset et al., 2017). However, FER is known to decline with age, even in the absence of acquired cerebral lesions (Richoz et al., 2018; Calder et al., 2003). As the patient continues to age (i.e., she was born in 1950), the gap between her performance and that of her age-matched peers may be narrowing because performance also deteriorates with aging in healthy seniors. Therefore, our analysis aimed to determine whether degraded visual stimuli could not only elicit a dynamic advantage across populations but also reveal subtle differences between the patient and her peers, which might not be apparent under optimal viewing conditions.

To the best of our knowledge, this is the first study that assesses the precision of individual perceptual threshold estimates obtained using the QUEST algorithm in the context of both static and dynamic FER. Additionally, it establishes methodological guidelines for its most efficient use. By rigorously assessing the effectiveness of this approach, our study not only validates the use of the QUEST for FER research but also lays the groundwork for its broader implementation in single-subject studies and clinical applications, where precise and time-efficient assessment tools are critically needed.

2. Material and methods

2.1. Participants

2.1.1. Young and older adults

We conducted a priori power analysis using G*Power software (Faul et al., 2007; latest ver. 3.1.9.7; Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany), to determine the samples size necessary to compare a large group of young adults and a smaller group (one fifth) of older participants, with a Mann-Whitney-Wilcoxon test. With this method, aiming to be a relevant tool for revealing differences in sensitivity to FER in various populations – including clinical ones – we set the expected effect size, the smallest effect we were interested in, at $d = .8$ (see Anderson et al., 2017) ($\alpha = .05$; $1- \beta = .8$), leading to estimate that a minimum sample size of 63 young-adult and 13 older participants was needed. Because the study required multiple testing sessions and participant completion rates were unknown at enrolment, we recruited beyond these minimum estimates using a conservative oversampling strategy to ensure that the final analyzable sample met a priori power requirements. Accordingly, we recruited 104 young adults and 21 older adults. Our final sample data was 100 young adults (YA; $M = 21.4$ years; $SD = 2.3$; 82 females), which were tested between 2023 and 2024 in Fribourg: they were volunteers of the experimenters' entourage or students. Four additional young adults' participants were removed due to extreme values (participants 101 to 104, see in the detailed data file available on Open Science Framework https://osf.io/t2csb/?view_only=8cbf165a4df746daaacee2baf1a0cd72). Twenty-one healthy older individuals (EC; $M = 69.1$ years; $SD = 3.8$; years, 14 females), considered as age-matched controls for the prosopagnosic patient PS were recruited and tested in 2023 in the region of Fribourg and Malta. They all volunteered and were tested either in senior housing or at their own home. All healthy participants had normal or corrected-to-normal visual acuity, no problematic drug consumption, and did not suffer from neurological, developmental, or psychiatric disorders. All participants provided written consent, and the study was approved by the local ethical committee.

2.1.2. Patient PS's case report

Patient PS, born in 1950, is a well-documented case of pure acquired prosopagnosia. In 1992, she suffered a closed head injury, resulting in major lesions to the left middle fusiform gyrus (left Fusiform Face Area) and right inferior occipital gyrus (right Occipital Face Area) - both critical for face processing (for a review, see Rossion et al., 2003). Minor damage also occurred in the right middle temporal gyrus and left posterior cerebellum (see Sorger et al., 2007). Importantly, brain regions known to be involved in facial expression recognition, such as the amygdala, insula, and posterior superior temporal sulcus (pSTS), are structurally spared in patient PS.

Following her injury, PS initially experienced broad cognitive deficits but recovered well with medical and neuropsychological support. However, as a consequence of her lesions she developed profound prosopagnosia, being severely impaired to recognize familiar faces, even close relatives or herself (Rossion et al., 2003; Rossion, 2022a, 2022b). Despite this, she could still distinguish faces from other objects and performed normally on object recognition tests (BORB; Riddoch and Humphreys, 1993). Her reading, visual acuity, and visual fields were largely intact, except for a small left paracentral scotoma.

Patient PS's lesions also significantly impair her performance on facial expression recognition tasks with *static* faces (Richoz et al., 2015). Interestingly, her ability to recognize *dynamic* facial expressions remains preserved (Richoz et al., 2015).

Patient PS was tested at her home in 2023 and provided written consent for her participation.

2.2. Stimuli and task

A total of 48 stimuli were taken from Gold et al. (2013), with 4 female and 4 male identities expressing the six basic FEE (i.e., anger, disgust, fear, happiness, sadness, and surprise). Stimuli included dynamic and static versions of each expression. Dynamic stimuli portrayed each identity evolving from a neutral into a fully developed expression over 30 frames, lasting exactly 1 s. Static stimuli consisted of 30 repetitions of the final frame of the corresponding dynamic stimulus, also lasting 1 s, corresponding to the apex of each expression. All stimuli were used as originally created by Gold et al. (2013), without modification to sampling rate or duration. The faces were grayscale and cropped at the hairline. All stimuli were normalized for their low-level properties (i.e., spatial frequency, luminance, and contrast) using the SHINE toolbox (Willenbockel et al., 2010). Faces subtended a vertical visual angle of 12° on the screen at a viewing distance of 65 cm and were displayed on a color liquid-crystal display with a resolution of 1440 × 900 pixels and a 60 Hz refresh rate. The experiment was programmed in The MathWorks Inc, 2022 R2022b (MathWorks, Natick, MA) using the Psychophysics Toolbox (PTB-3; Brainard, 1997; Kleiner et al., 2007).

Participants took part in a FER task including a total of 1920 trials for the young adults and 960 trials for PS and the older adults. The young adults were tested across two recognition accuracy levels, 44% and 58%. Testing the 44% recognition level means evaluating the level of visual noise at which participants reach a fixed threshold of 44% correct FER performance. Similarly, testing the 58% recognition threshold assesses participants' ability to achieve 58% correct FER performance. These two accuracy levels were selected based on the work of Richoz et al. (2024). In their study, they examined neurotypical adults for static and dynamic FER at various signal levels and applied curve-fitting to their results. When estimating the performance values (95% confidence intervals) for 100% signal, the lower bound of the confidence interval was above 58% for every FEE in both conditions, except for fear, for which the lower bound was just under 44% in the dynamic presentation. Based on these results we reasoned that these two performance levels would be both challenging, enhancing the dynamic advantage, but still within the participants' capabilities, leading to meaningful results.

The older adults and PS were assessed in the current study only for the 44% threshold. This choice was driven by the need to reduce testing time and the expectation that attaining 58% of accuracy even with 100% of signal might not be possible for this population. In such case the QUEST algorithm would likely estimate that more than 100% of signal is required to reach the predefined performance level. In each trial, a fixation cross was presented at the center of the screen for 500ms, followed by a 1-s face stimulus whose frames were displayed at 30Hz. Participants then pressed the keyboard button corresponding to the perceived expression. If they were unable to make this decision, they were allowed to answer: "I don't know". In this case the program randomly chose one expression for them. This was done to prevent participants from systematically responding one specific expression. When testing PS and the older adults, participants verbally communicated their response to the experimenter who pressed the corresponding key.

The percentage of signal presented in each trial was determined using the QUEST toolbox, which, through a Bayesian adaptive procedure, estimates an observer's sensitivity threshold based on the amount of signal presented in previous trials and the corresponding performance (see Fig. 1). Based on the data reported in Richoz et al. (2024), the prior threshold guess was set to 20% of signal (i.e., phase coherence) for the 44% accuracy condition, and to 24% of signal for the 58% accuracy condition.

All participants were tested using the same order, beginning with the 44% accuracy condition, followed by the 58% condition. Within each accuracy level, the static modality was presented first, followed by the dynamic modality. We chose not to randomize the order for two main reasons. First, using the same sequence allowed direct comparison between patient PS and all older adult participants. Second, because the 44% accuracy condition requires lower target performance, the stimuli contain less visual information than in the 58% condition. Therefore, starting with the lower-signal condition reduces potential learning

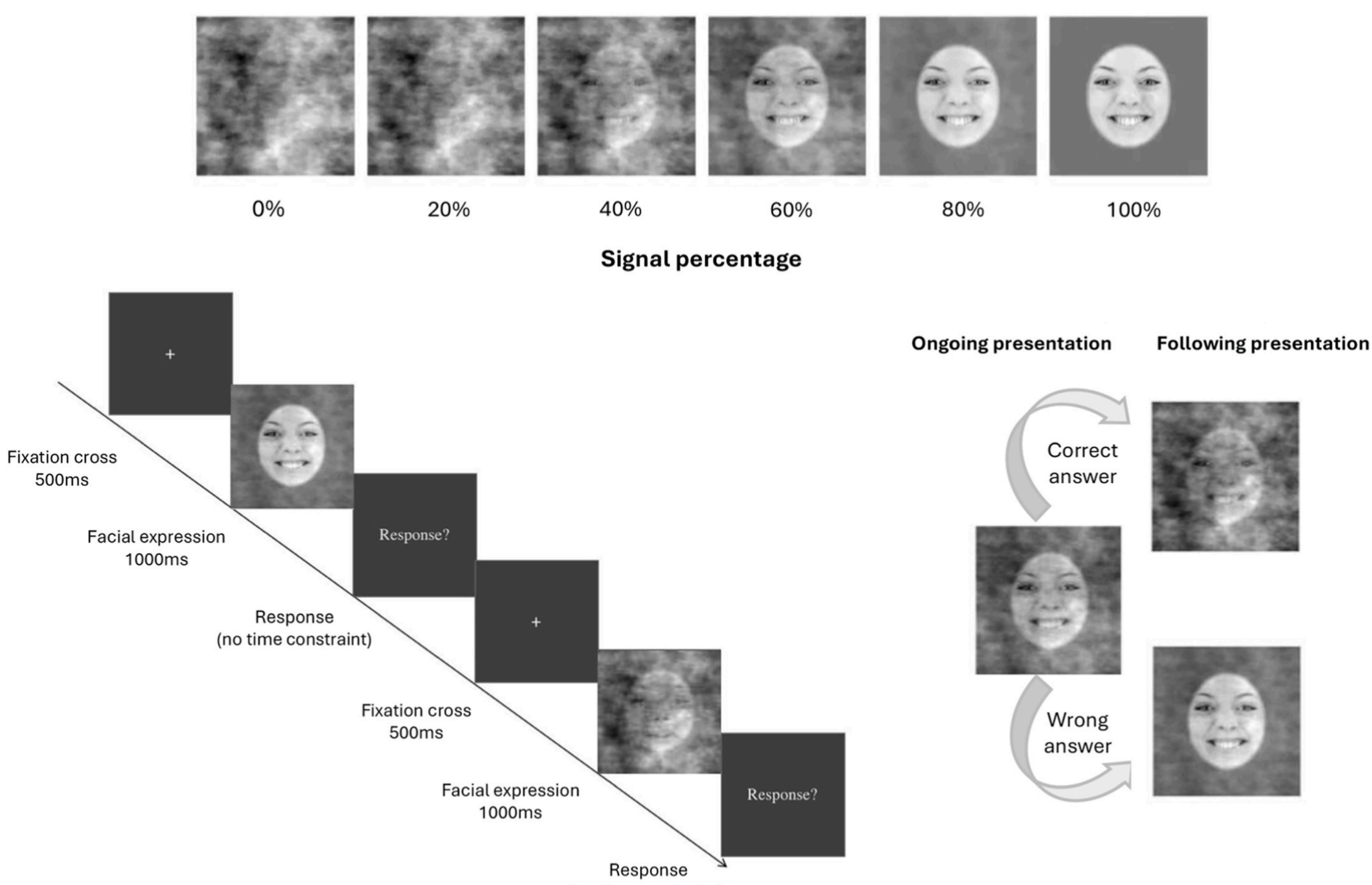

Fig. 1. Schematic representation of the procedure
Note. From top to bottom and left to right. On top: example of one identity expressing happiness at different levels of phase signal (0%, 20%, 40%, 60%, 80%, and 100%). On the bottom: schematic representations of the testing procedure (left), and of the QUEST threshold-seeking algorithm functioning (right). We adapted the stimuli with permission from Gold et al. (2013).

effects. Similarly, presenting the static modality before the dynamic modality prevented participants from associating a static frame with its dynamic version, which could otherwise allow them to rely on the memory of the more informative dynamic stimulus to correctly interpret the static frame. To further minimize learning effects and reduce fatigue due to long testing sessions, each accuracy-modality combination was run on a separate day, resulting in four testing sessions for young adults and two testing sessions for older adults and patient PS.

Each session included one QUEST procedure per expression with 80 trials each. These were presented in a semi-randomized fashion through the session, which however ensured that every 60 trials, expressions were presented equally.

2.3. Data analysis

2.3.1. Optimal accuracy level determination

Before conducting the formal analysis, we performed a preliminary assessment to determine which accuracy level - 44% or 58% - was more appropriate for studying FER sensitivity in young adults. Therefore, our preliminary analysis aimed to evaluate which target accuracy level would be more likely to be associated with realistic signal thresholds. We defined a signal threshold as *realistic* if it required less than 100% of signal to reach the target accuracy. If an observer required more than 100% of signal to achieve a specific accuracy level, that level was considered unattainable for that individual.

For each emotion and modality, we compared the 44% and 58% performance levels in terms of the number of participants whose estimated threshold exceeded 100% of signal. Perceptual threshold estimates were derived from the mean of the posterior probability distribution computed by the QUEST algorithm after 80 trials.

A McNemar exact test revealed that, in half of the conditions, significantly more participants had unrealistic threshold estimates at the 58% performance level compared to the 44% ($p < .05$).

The scope of this work is to explore stability and efficiency of the QUEST algorithm in measuring individuals' sensitivity in the context of FER. Therefore, we decided to retain and further analyze only the results for the 44% target performance, which was more likely attainable.

Within this condition, we also performed data filtering to ensure that participants with extreme estimates would not be included in the formal analysis. Specifically, we averaged the perceptual thresholds estimated across all emotions for the static and dynamic modalities separately. We also computed the difference between these two modalities (static minus dynamic) to index the *dynamic advantage*. Participants whose scores – on any of these three measures – fell more than 2.5 standard deviations above or below the group mean were excluded from the formal analysis.

2.3.2. Evaluation of stabilization behavior and efficiency of the QUEST procedure

Our first and second research questions examine whether the QUEST algorithm could reach a stabilization point during its estimation of the perceptual threshold for FER and how many trials it required to do so. We identified a stabilization point by examining the threshold estimates after each trial using two criteria: (1) the presence of a specific oscillatory pattern across four consecutive trials and (2) the magnitude of the fluctuation within that pattern.

Our operationalization of the oscillatory pattern was based on the typical behavior of QUEST as it homes in on a threshold. In this phase, the estimate tends to oscillate around a stable value. We defined an oscillation as a directional reversal in the estimate (up-down or down-up). A repetitive oscillation therefore requires at least two such reversals (i.e., up-down-up or down-up-down), which corresponds to four consecutive trials. We chose this pattern length, rather than a single oscillation, to increase the likelihood of observing a genuine homing-in rather than a random fluctuation. At the same time, we reasoned that an even longer pattern could be disrupted by momentary attentional lapses. Indeed, although each facial expression had its own QUEST procedure, trial order was randomized within the same session, making it unlikely that the same expression would be tested on consecutive trials.

The second criterion concerned the amplitude of the oscillation. Because we had no a priori knowledge of how narrowly the threshold estimates might vary, we implemented an iterative procedure to identify the smallest amplitude attainable by most of our participants. We first identified the proportion of participants exhibiting a repetitive oscillation with an amplitude below 1.5%, and then we gradually increased this limit by .1%. The iterative process ended when a stabilization point could be found in at least 90% of the YA and OA participants in all FEE-modality combinations.

2.3.3. Handling of Missing stabilized threshold estimates

In the current work, we adopted a conditional inclusion strategy to maximize useable data. Specifically, participants who did not reach a stabilized threshold for a given expression in a given modality were excluded only from analyses involving that specific FEE-modality combination, while being retained for all other analyses.

For analyses involving the global score, a stricter criterion was applied: the global score was computed only when a participant exhibited a stabilized threshold for all six FEEs within a given modality. Additionally, participants exhibiting unrealistic threshold estimates suggesting they required less or close to 0% of visual signal were removed according to the same procedure. To ensure interpretability of the results, the number of participants retained in each FEE-modality is reported in Tables 1 and 3 as well as in Supplementary Table S1 and S3.

2.3.4. Statistical comparison of the number of trials necessary to reach stabilization

For each participant in each modality and for each emotion, we determined if a stabilization point could be found and the number of trials that were necessary to reach it. Additionally, within each modality, we computed a *global* score represented as the average across emotions.

We used a Mann-Whiney-Wilcoxon test to compare the number of trials required to reach stabilized threshold estimates between modalities (static vs. dynamic) for all emotions and the global score within the YA and the OA groups separately. The same statistical test was also used to compare YA and OA across all emotions, the global score and both modalities.

Finally, patient PS was compared to the OA group across all emotions, the global score, and both modalities using the Crawford *t*-test (Crawford and Garthwaite, 2002; Crawford and Howell, 1998).

Importantly, if a participant did not reach stabilization in any given emotion, they were not included in the modality comparison for that emotion nor for the global score. As we did not have any hypothesis on the direction of the effect, we performed all tests as two-tailed. Bonferroni correction was applied to control for multiple comparisons across emotions.

2.3.5. Statistical comparison of perceptual thresholds across groups and modalities

Perceptual thresholds were estimated using the mean of the posterior probability distribution from the QUEST algorithm after the final trial of the stabilization pattern (i.e., stabilized threshold).

Because this approach produces thresholds based on a variable number of trials across participants and conditions, we also estimated perceptual thresholds using all 80 trials (i.e., 80-trials threshold). Analyses on these data will serve as a control condition to ensure that the observed pattern of results are not dependent on the number of trials performed.

In addition to analyzing the perceptual thresholds for the six basic facial expressions of emotion separately, we also computed a *global threshold* within each modality by averaging the estimates across emotions. Furthermore, we calculated the *dynamic advantage* by subtracting the threshold estimate in the dynamic modality from the threshold estimate in the static modality. Statistical comparisons were conducted between modalities within each group, and between YA and OA for each emotion (including the global threshold), each modality, and the dynamic advantage. These analyses were performed using the Wilcoxon signed-rank test for paired observation and the Wilcoxon rank-sum test for independent observations, as appropriate.

Importantly, when comparing perceptual thresholds extracted at the stabilization point, participants who did not reach this point for a given emotion were not included in the modality comparison for that emotion, the global score, the dynamic advantage for that emotion, and the global threshold.

All tests were one-tailed, based on our hypotheses predicting lower perceptual thresholds for YA compared to OA, and for the dynamic modality compared to the static modality.

Finally, we compared patient PS with the OA group across all emotions, the global threshold, the two modalities and the dynamic advantage using the Crawford t-tests. These tests were two-tailed as we did not have strong a priori hypotheses on the differences between PS and the OA group. Bonferroni correction was applied to control for multiple comparisons across emotions.

2.3.6. Statistical comparison of perceptual thresholds at the stabilization point vs. after 80 trials

For participants who reached a stabilization point, we examined whether their perceptual threshold at that point was related to, and significantly different from, the final estimate obtained after 80 trials. These supplementary analyses were conducted separately for each group (i.e., YA and OA), emotion and modality. The relationship between the two threshold estimates was assessed using Spearman rank correlations, and differences were tested using the two-tailed Wilcoxon signed-rank test. Multiple comparisons across emotions were corrected using the Bonferroni method.

2.4. Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data are available online on the Open Science Framework repository at: https://osf.io/t2csb/?view_only=8cbf165a4df746daaacee2baf1a0cd72, the analysis code and research materials are available upon request. This study's design and its analysis were not pre-registered.

3. Results

3.1. Stabilization of the perceptual threshold estimate

Stabilization of the perceptual threshold was defined as an oscillation within a predefined range across at least four consecutive trials. To define a suitable oscillation range, we quantified the proportion of participants meeting the criterion across increasingly larger amplitude limits (see Methods). Results showed that a 2% fluctuation amplitude yielded a stabilization for at least 90% of YA and OA participants across expressions and modalities (Supplementary Table S3; four additional YA participants were excluded from one condition due to aberrant threshold values). Patient PS also exhibited stabilized thresholds in all FEE-modality combinations.

Statistical analysis revealed that, within each group (YA and OA), the number of trials required to reach the stabilization point did not significantly differ between the static and dynamic modalities, whether considered for each emotion individually or for the global score.

In contrast, comparisons between groups showed that, based on the global score, older adults required significantly more trials to reach stabilization than young adults in both the static ($p < .01$) and dynamic modalities ($p < .05$; Table 2).

Analysis at the single-expression level indicated that, compared to the YA, the OA population required significantly more trials to reach a stable threshold estimate only for static Anger ($p < .05$; Table 1).

Analyses comparing patient PS to the OA population showed no significant difference in the number of trials necessary to reach a stable estimate for either the static or dynamic modality, regardless of whether based on the global score or individual expressions. Nevertheless, we observed a trend for the static expression of fear, where PS required more trials to reach stabilization than the OA population ($p = .051$; Table 2).

3.2. Perceptual threshold across modalities

Perceptual thresholds were defined as the mean of the posterior distribution generated by the QUEST algorithm, either at the stabilization point (i.e., stabilized threshold) or after the full 80-trials sequence (i.e., 80-trials threshold). Unless otherwise specified the reported results refer to both the stabilized and 80-trials thresholds.

Comparison between modalities revealed that, globally, young adults (YA) exhibited significantly lower thresholds in the dynamic compared to the static condition (stabilized threshold: $p < .001$; 80-trials threshold: $p < .001$). This dynamic advantage was also significant for most individual expressions (stabilized threshold: $p < .05$; 80-trials threshold: $p < .01$), except for fear ($p > .05$).

Similarly, at the global level, older adults (OA) also exhibited significantly lower thresholds in the dynamic condition compared to the static one (stabilized threshold: $p < .001$; 80-trials threshold: $p < .05$). When considering individual expressions, this dynamic advantage reached significance only for anger, happiness and surprise (stabilized thresholds: $p < .05$; 80-trials threshold: $p < .01$). Results also revealed a significant dynamic advantage for disgust but only when considering the 80-trials threshold ($p < .05$) Full results are provided in Table 4 for the stabilized thresholds and in Supplementary Table S2 for the 80-trials thresholds (see also Fig. 2).

3.3. Perceptual threshold across populations

As in the previous section, reported results refer to both the stabilized and 80-trial thresholds unless specified otherwise.

Statistical comparisons between the OA and YA groups revealed significantly lower perceptual thresholds in YA for most expressions and modalities (stabilized threshold: $p < .05$; 80-trials threshold: $p < .05$). No significant differences were found for sadness in either modality for the stabilized threshold ($p > .05$), nor for the static expression of sadness for the 80-trials threshold ($p > .05$). When considering the dynamic advantage, results showed that it was significantly larger in OA compared to YA but only at the global level (stabilized threshold: $p < .01$; 80-trials threshold: $p < .05$); no significant differences were found at the single-expression level (all $p > .05$).

Comparison between the OA group and patient PS revealed that globally there was no significant difference in perceptual thresholds, regardless of presentation modality ($p > .05$). At the single-expression level, PS did not differ significantly from the OA group for any expression in either the static or dynamic modality, with one exception: for the 80-trials threshold, PS required significantly more signal to recognize the static expression of happiness ($p < .05$).

When considering the dynamic advantage, results revealed that globally it was significantly larger for PS than for the OA group for the 80-trials thresholds ($p < .05$) but not for the stabilized ones ($p > .05$). At the single-expression level, PS showed a significantly larger dynamic advantage than the OA group for fear ($p < .05$) and surprise ($p < .05$) for the stabilized-threshold and only for fear for the 80-trials threshold ($p < .05$). Additionally, PS also showed a significantly larger static advantage for sadness (stabilized threshold: $p < .01$; 80-trials threshold: $p < .05$). Full results are provided in Table 4 for the stabilized thresholds and in Supplementary Table S2 for the 80-trials thresholds (see also Fig. 2).

Table 1
Descriptive statistics for the number of trials required to reach stabilization across populations, modalities, and facial expression.

Population	Modality	FEE	N	Mean (SD)	Percentile		
					25%	50% [95% CI]	75%
YA	Static	Anger	98	22.14 (10.49)	14	19 [17 – 21]	26.75
		Disgust	99	20.17 (6.79)	15	19 [18 – 21]	24.5
		Fear	98	26.1 (13.39)	17	22 [19 – 25]	33.75
		Happiness	100	20.56 (7.54)	15	19 [17 – 21]	24
		Sadness	100	21.02 (9.6)	14	19 [17 – 21]	25
		Surprise	100	20.99 (8.74)	14	20.5 [18 – 22]	26
		Global	95	21.74 (4.01)	19.42	21.33 [20.33 – 22]	23.67
	Dynamic	Anger	100	21.95 (9.8)	15	19.5 [18 – 22]	25
		Disgust	100	21.68 (9.94)	14	20 [18 – 21]	26
		Fear	96	26.41 (12.96)	17	23 [21 – 26]	31
		Happiness	98	20.95 (7.08)	16	20 [18 – 22]	25
		Sadness	100	19.77 (7.39)	14	18 [16 – 20]	24
		Surprise	99	19.88 (7.9)	13.5	19 [16 – 20]	23
		Global	93	21.83 (3.76)	19	21.5 [20.93 – 21.83]	23.33
	Static-Dynamic difference	Anger	98	.49 (12.84)	−8.75	2.5 [-2 – 4]	8
		Disgust	99	−1.49 (12.14)	−7	−1 [-3 – 2]	6
		Fear	94	−.5 (15.93)	−8.75	−2 [-5 – 2]	8
		Happiness	98	−.32 (9.92)	−6	−1.5 [-3 – 1]	5
		Sadness	100	1.25 (12.22)	−4	0 [-2 – 2]	6.25
		Surprise	99	1.08 (11.52)	−5.5	0 [-2 – 2]	6.5
		Global	88	.27 (5)	−2.71	.5 [-1.17 – 1.33]	3.58
OA	Static	Anger	21	30.81 (14.08)	18	30 [18 – 33]	38
		Disgust	21	27.48 (14.73)	18	23 [18 – 29]	29
		Fear	21	28.14 (13)	20	26 [20 – 32]	32
		Happiness	21	19.48 (6.01)	15	18 [15 – 22]	24
		Sadness	21	21.95 (11.03)	17	18 [17 – 21]	22
		Surprise	21	23.76 (9.48)	18	21 [18 – 28]	29
		Global	21	25.27 (4.7)	21.83	25.17 [21.83 – 27.67]	28
	Dynamic	Anger	21	22.81 (10.61)	15	20 [15 – 28]	31
		Disgust	21	22.48 (12.86)	15	18 [15 – 24]	24
		Fear	19	33.89 (15.38)	21.5	33 [21– 40]	42.5
		Happiness	21	20.19 (5.88)	18	20 [18 – 22]	24
		Sadness	21	22.76 (9.9)	15	22 [15 – 29]	29
		Surprise	21	23.33 (10.58)	17	21 [17 – 22]	24
		Global	19	24.39 (4.18)	21.5	23.5 [21.33 – 26.83]	27.42
	Static-Dynamic difference	Anger	21	8 (16.3)	−4	5 [-4 – 17]	18
		Disgust	21	5 (21.33)	−2	5 [-2 – 16]	18
		Fear	19	−6.11 (17.41)	−16.5	−6 [-19 – 5]	7
		Happiness	21	−.71 (7.19)	−4	−1 [-4 – 2]	2
		Sadness	21	−.81 (14.69)	−11	1 [-11 – 3]	5
		Surprise	21	.43 (15.54)	−4	1 [-4 – 7]	10
		Global	19	1.08 (4.68)	−2.08	1.33 [-2.33 – 5]	5.17
PS	Static	Anger		28			
		Disgust		20			
		Fear		67			
		Happiness		21			
		Sadness		17			
		Surprise		46			
		Global		33.17			
	Dynamic	Anger		20			
		Disgust		40			
		Fear		29			
		Happiness		22			
		Sadness		37			
		Surprise		14			
		Global		27			
	Static-Dynamic difference	Anger		8			
		Disgust		−20			
		Fear		38			
		Happiness		−1			
		Sadness		−20			
		Surprise		32			
		Global		6.17			

Note. YA = young adults; OA = older adults; PS = prosopagnosic patient; FEE = facial expression of emotions.

3.4. Statistical comparison of the stabilized vs. 80-trials thresholds

Correlation analyses revealed a strong relationship between the perceptual thresholds estimated at the stabilization point (stabilized threshold) and those obtained after the full 80-trials sequence (80-trials threshold) for almost all emotions and modalities. For correlations reaching significance ($p < .05$) after Bonferroni correction, Spearman's ρ ranged from .30 to .95 in YA and from .70 to .96 in OA. The only condition that failed to reach significance was the dynamic advantage of happiness in both YA and OA. Full correlation results are reported in Supplementary Table S4.

Finally, Wilcoxon signed-rank tests revealed no significant difference between the stabilized and 80-trials perceptual thresholds across most conditions. In those cases, the median differences between estimates

Table 2
Inferential statistics for the number of trials required to reach stabilization for modality and population comparisons.

		Static-Dynamic contrast			
		YA		OA	
	FEE	V	r_s [95%CI]	V	r_s [95%CI]
	Anger	2416	.04 [-.19 – .26]	170.5	.48 [.03 – .76]
	Disgust	2205.5	−.09 [-.31 – .14]	141.5	.35 [-.14 – .7]
	Fear	1994	−.05 [-.28 – 19]	61.5	−.35 [-.71 – .14]
	Happiness	2021.5	−.08 [-.3 to −.16]	100.5	−.13 [-.55 – .34]
	Sadness	2245.5	.05 [-.18 – .28]	103.5	−.1 [-.53 – .37]
	Surprise	2437	.09 [-.14–.31]	131	.13 [-.34 – .55]
	Global	2002.5	.07 [-.17 – .3]	122	.28 [-.22 – .67]
		Population contrast			
		YA-OA		OA-PS	
Modality	FEE	W	r_{rb} [95%CI]	t	zcc
Static	Anger	1438.5*	.4 [.15 – .6]	−.19	−.2
	Disgust	1366.5	.31 [.05 – .54]	−.50	−.51
	Fear	1172.5	.14 [-.13 – .39]	2.92	2.99
	Happiness	1000.5	−.05 [-.31 – .22]	.25	.25
	Sadness	1076	.02 [-.24 – .29]	−.44	−.45
	Surprise	1224.5	.17 [-.1 – .41]	2.29	2.35
	Global	1435**	.44 [.19 – .63]	1.64	1.68
Dynamic	Anger	1093.5	.04 [-.23 – .3]	.26	−.26
	Disgust	1039	−.01 [.28 – .26]	1.33	1.36
	Fear	1204	.32 [.05 – .55]	−.31	−.32
	Happiness	999	−.03 [-.29 – .24]	.30	.31
	Sadness	1213.5	.16 [-.12 – .41]	1.40	1.44
	Surprise	1247.5	.2 [-.07 – .44]	−.86	−.88
	Global	1204*	.36 [.09 – .58]	.61	.62

Note. YA = young adults; OA = older adults; FEE = facial expression of emotions; V = Wilcoxon signed-rank test; r_s = rank biserial correlation (effect size) for the Wilcoxon signed-rank test; W = Wilcoxon rank-sum test; r_{rb} = rank biserial correlation (effect size) for the Wilcoxon rank-sum test; t = modified Crawford-Howell *t*-test; zcc = effect size for the modified Crawford-Howell *t*-test.
***$p < .001$; **$p < .01$; *$p < .05$.

ranged from −.87% to 1.63% in YA and from −4.19% to 3.42% in OA. Significant differences were observed only in YA for sadness in either modality ($p < .01$), for static happiness ($p < .05$), and for the global static and dynamic conditions ($p < .05$). For these conditions, the median differences between stabilized and 80-trials thresholds ranged from −.82% to 2.20%.

4. Discussion and conclusions

This study evaluated the reliability and efficiency of the QUEST threshold-seeking algorithm as a tool for estimating individual sensitivity in facial expression recognition. Addressing the need for rapid and precise assessment tools that capture individual variability, we applied the QUEST algorithm across three populations: neurotypical young adults, older adults, and a patient with acquired prosopagnosia, PS. We estimated the minimum amount of signal each individual required to reach a predefined level of recognition accuracy for both static and dynamic expressions by using a psychophysical approach. Specifically, we examined: (1) whether the QUEST algorithm can generate stable perceptual threshold estimates at the single-subject level; (2) how many trials are required for the algorithm to reach stabilization across individuals, expressions, and modalities; and (3) whether the resulting thresholds capture well-established effects in FER, such as age-related decline and the dynamic advantage – and whether this advantage could serve as a marker for frailties in the FER system. In the following sections, we discuss each of these points in turn and consider their theoretical and methodological implications. Finally, we provide expression-specific guidance for future studies that aim to examine FER sensitivity with the QUEST. Before addressing these core questions, we first provide a brief discussion of the potential interpretation of the 44% accuracy level in terms of perceptual challenge.

In this work, we examined the amount of noise required to achieve two separate accuracy levels: 44% and 58%. The latter revealed to pose too much of a challenge, especially to older adults, who often required more than 100% of visual signal to perform at this level. In contrast, 44% accuracy was more attainable by both young and older adults as well as by patient PS. Although interpreting a single accuracy level should be done with caution, performance at 44% in a 6AFCS task indicates that participants are performing well above chance (16.67%) while still facing substantial perceptual difficulty. One way to contextualize this threshold is by considering performance at full signal. Richoz et al. (2018) showed that accuracy around 40% is typically observed in young children and older adults when no visual noise is present. Under this framing, young adults in the current study might be facing a level of perceptual challenge comparable, at least quantitatively, to that experienced by older adults under natural viewing conditions. This interpretation is in line with our observation of a dynamic advantage at 44% of accuracy, which supports the notion that dynamic information provides significant support during facial expression recognition when the FER system is stressed, whether by physiological factors such as aging, or by externally imposed visual noise. Nevertheless, whether the underlying mechanisms are qualitatively comparable remains an open question.

4.1. *Stability of individual-level threshold estimates*

First and foremost, to validate the effectiveness of the QUEST algorithm in assessing FER sensitivity at the individual level, we examined its capacity to produce stable and precise threshold estimates across different populations and for both the recognition of static and dynamic FEEs.

Our study builds directly on the work of Richoz et al. (2024) who used a parametric manipulation of the signal strength (phase coherence ranging from 0% to 100%) to examine FER in young adults. Their approach combined uniform and adaptive sampling methods to systematically vary the signal strength of the FEEs presented. They conducted group-level analyses using Weibull fits of binned accuracy data to estimate recognition curves. While their findings offered important

Table 3
Descriptive statistics for the stabilized perceptual threshold estimates across populations, modalities, and facial expression.

Population	Modality	FEE	N	Mean (SD)	Percentile		
					25%	50% [95% CI]	75%
YA	Static	Anger	98	33.01 (21.29)	16.76	26.02 [21.68 – 33.25]	45.52
		Disgust	99	27.06 (11.12)	19.11	24.38 [22.37 – 25.71]	32.61
		Fear	98	44.10 (33.32)	16.06	30.29 [22.04 – 46]	63.05
		Happiness	100	14.03 (6.81)	11.41	13.23 [12.73 – 13.82]	14.73
		Sadness	100	25.90 (18.99)	14.47	18.36 [16.29 – 21.24]	29.97
		Surprise	100	30.96 (21.31)	18.46	23.87 [21.58 – 27.49]	35.46
		Global	95	28.98 (8.63)	23.01	27.65 [25.44 – 29.63]	33.17
	Dynamic	Anger	100	27.85 (27.53)	9.64	14.6 [12.52 – 20.14]	39.29
		Disgust	100	23.21 (16.29)	13.6	18.35 [15.19 – 20.73]	24.55
		Fear	96	45.92 (38.56)	13.33	29.88 [20.89 – 46.22]	69.16
		Happiness	98	7.48 (2.97)	5.62	7.15 [6.58 – 7.68]	8.91
		Sadness	100	23.24 (19.83)	11.48	15.4 [13.38 – 17.01]	27.75
		Surprise	99	21.38 (22.04)	7.75	11.58 [9.57 – 15.28]	27.01
		Global	93	25.28 (11.57)	16.42	23.29 [21.05 – 25.96]	31.86
	Static-Dynamic difference	Anger	98	6.51 (28.44)	−1.56	5.95 [4.28 – 11.96]	23.17
		Disgust	99	3.94 (15.96)	−1.3	4.73 [2.43 – 7.73]	12.13
		Fear	94	−2.89 (40.94)	−25.13	2.61 [-3.45 – 8.26]	18.03
		Happiness	98	6.69 (7.70)	3.59	6.48 [4.75 – 7.09]	8.17
		Sadness	100	2.66 (21.83)	−2.85	3.17 [.70 – 5.48]	11.76
		Surprise	99	9.79 (25.36)	2.12	10.44 [6.78 – 12.79]	17.93
		Global	88	3.88 (9.82)	−2.28	5.01 [1.64 – 7.31]	10.69
OA	Static	Anger	21	69.70 (33.65)	43.51	66.94 [43.51 – 93.33]	99.67
		Disgust	21	54.13 (32.31)	29.7	48.46 [29.70 – 64.40]	71.85
		Fear	21	82.19 (42.63)	39.04	90.45 [39.04 – 121.66]	121.77
		Happiness	21	16.66 (7.15)	12.44	16.5 [12.44 – 18.39]	18.98
		Sadness	21	40.35 (33.52)	18.93	23.93 [18.93 – 45.63]	49.51
		Surprise	21	47.33 (26.65)	24.33	37.48 [24.33 – 67.02]	71.3
		Global	21	51.72 (16.96)	37.66	52.08 [37.66 – 61.07]	63.22
	Dynamic	Anger	21	46.50 (30.83)	23.42	40.25 [23.42 – 49.75]	57.68
		Disgust	21	38.60 (23.71)	17.92	34.25 [17.92 – 44.42]	45.18
		Fear	19	81.50 (42.72)	41.61	99.62 [35.53 – 114.85]	118
		Happiness	21	10.91 (2.61)	9.24	10.83 [9.24 – 11.95]	13.03
		Sadness	21	35.73 (30.67)	14.74	21.12 [14.73 – 43.10]	46.15
		Surprise	21	32.13 (22.21)	13.83	23.44 [13.83 – 47.20]	48.83
		Global	19	41.98 (16.46)	28.89	42.49 [27.17 – 53.87]	54.1
	Static-Dynamic difference	Anger	21	23.19 (31.35)	1.02	17.09 [1.02 – 41.73]	50.62
		Disgust	21	15.54 (34.35)	−2.96	13 [-2.96 – 36.78]	37.09
		Fear	19	−3.51 (21.70)	−19.52	.51 [-19.73 – 8.08]	9.46
		Happiness	21	5.74 (5.99)	2.19	5.05 [2.19 – 8.72]	9.5
		Sadness	21	4.61 (21.01)	−4.38	1.96 [-4.38 – 3.51]	5.24
		Surprise	21	15.20 (24.53)	4.55	14.66 [4.55 – 22.45]	22.47
		Global	19	9.99 (10.12)	2.15	13.38 [1.81 – 16.44]	16.51
PS	Static	Anger		91.39			
		Disgust		43.3			
		Fear		103.23			
		Happiness		21.05			
		Sadness		20.6			
		Surprise		105.54			
		Global		64.18			
	Dynamic	Anger		19.37			
		Disgust		24.9			
		Fear		33.45			
		Happiness		18.42			
		Sadness		110.07			
		Surprise		12.1			
		Global		36.39			
	Static-Dynamic difference	Anger		72.01			
		Disgust		18.4			
		Fear		69.78			
		Happiness		2.62			
		Sadness		−89.47			
		Surprise		93.44			
		Global		27.8			

Note. YA = young adults; OA = older adults; PS = prosopagnosic patient; FEE = facial expression of emotions.

insights into how dynamic and static FEEs are recognized under varying visual conditions, their method did not implement a threshold-seeking algorithm and was not designed to yield reliable perceptual thresholds at the individual level.

In contrast, in the current study, we used the QUEST algorithm to derive threshold estimates at the single-subject level. Initial QUEST parameters were derived from the slope of FER performance - as a function of visual noise - reported by Richoz et al. (2024), providing plausible starting values for threshold estimation. Two accuracy targets - 44% and 58% - were evaluated. A prior test in a large sample of young adults revealed that the 44% accuracy level provided the most reliable and sensitive measure across facial expressions and modalities. Consequently, we adopted this threshold for subsequent analyses in young adults, older adults, and a patient with acquired prosopagnosia.

Table 4
Inferential statistics for the stabilized perceptual threshold estimates for modality and population comparisons.

		Static-Dynamic contrast			
		YA		OA	
	FEE	V	r_s [95%CI]	V	r_s [95%CI]
	Anger	3275**	.35 [.14 – .53]	193*	.67 [.31 – .86]
	Disgust	3496**	.41 [.21 – .58]	176	.53 [.09 – .79]
	Fear	2157	−.03 [-.26 – .20]	84	−.12 [-.56 – .38]
	Happiness	4618**	.90 [.85 – .94]	215***	.86 [.67 – .95]
	Sadness	3232*	.28 [.06 – .47]	138	.19 [-.28 – .60]
	Surprise	3882***	.57 [.40 – .70]	197**	.71 [.37 – .88]
	Global	2789***	.42 [.21 – .60]	174***	.83 [.59 – .94]
		Population contrast			
		YA-OA		OA-PS	
Modality	FEE	W	r_{rb}[95%CI]	t	zcc
Static	Anger	1675***	.63 [.43 – .77]	.63	.64
	Disgust	1666***	.6 [.4 – .75]	−.33	−.34
	Fear	1592***	.55 [.33 – .71]	.48	.49
	Happiness	1421*	.35 [.10 – .57]	.6	.61
	Sadness	1350	.29 [.02 – .51]	−.57	−.59
	Surprise	1471*	.40 [.15 – .60]	2.13	2.18
	Global	1788***	.79 [.67 – .88]	.72	.73
Dynamic	Anger	1517**	.44 [.20 – .64]	−.86	−.88
	Disgust	1558**	.48 [.25 – .66]	−.56	−.58
	Fear	1338**	.47 [.22 – .66]	−1.09	−1.12
	Happiness	1690***	.64 [.45 – .78]	2.8	2.88
	Sadness	1377	.31 [.05 – .53]	2.37	2.42
	Surprise	1426*	.37 [.12 – .58]	−.88	−.9
	Global	1405***	.59 [.37 – .75]	−.33	−.34
Static-Dynamic difference	Anger	1264	.23 [-.04 – .47]	1.52	1.56
	Disgust	1289	.24 [-.03 – .48]	−.08	.08
	Fear	849	−.05 [-.32 – .23]	3.29*	3.38
	Happiness	961	−.07 [-.33 – .20]	−.51	−.52
	Sadness	971	−.08 [-.33 – .19]	−4.37**	−4.48
	Surprise	1216	.17 [-.10 – .42]	3.11*	3.19
	Global	1144**	.37 [.10 – .59]	1.71	1.76

Note. YA = young adults; OA = older adults; FEE = facial expression of emotions; V = Wilcoxon signed-rank test; r_s = rank biserial correlation (effect size) for the Wilcoxon signed-rank test; W = Wilcoxon rank-sum test; r_{rb} = rank biserial correlation (effect size) for the Wilcoxon rank-sum test; t = modified Crawford-Howell *t*-test; zcc = effect size for the modified Crawford-Howell *t*-test.
***$p < .001$; **$p < .01$; *$p < .05$.

Given that FER sensitivity varies with age (Richoz et al., 2018), expression type (e.g., Calder et al., 2003), and visual complexity (e.g., Ambadar et al., 2005; Bould et al., 2008) it was critical to assess whether the QUEST could yield stable thresholds across diverse participants and conditions. A posteriori, data-driven analysis revealed that QUEST converged in nearly all individuals, facial expressions and conditions. Even among older adults, data exclusion rates remained below 10%, consistent with prior evidence that QUEST provides sensitive and unbiased threshold estimates in both developmental and clinical populations (Rodger et al., 2015, 2018; Watson and Pelli, 1983).

A key methodological novelty of the present study lies in how stabilization was defined and evaluated. Previous studies applying the QUEST algorithm to high-level visual tasks such as FER (e.g., Rodger et al., 2015) typically stopped testing once participants provided a sequence of three correct or incorrect responses and the standard deviation of the presented signal levels dropped below a given threshold. While such criteria may indicate that the stimulus presentation (i.e., the input) has stabilized, they do not guarantee that the algorithm's estimate of the perceptual threshold itself has genuinely converged. Low variation in signal levels alone cannot confirm stabilization, as participants may still be guessing or responding inconsistently, and the underlying threshold estimate may continue to fluctuate despite an apparently stable input. In contrast, our approach introduced a more rigorous and reliable stabilization criterion. Rather than relying on the stability of the stimuli, we assessed the stabilization of QUEST's internal threshold estimate across successive trials. Specifically, stabilization was defined as three consecutive directional changes of the estimated threshold within a 2% range – a criterion that ensures that the estimate itself, not just the stimulus, is stable.

Applying this refined criterion, we showed that QUEST yields stable individual-level perceptual thresholds for the recognition of both static and dynamic facial expressions. This methodological advance strengthens threshold estimation by ensuring that the derived thresholds accurately reflect perceptual sensitivity and are less susceptible to response noise or random guessing. As a result, our approach provides a strong foundation for future research on individual differences in facial expression recognition. This is particularly important given recent evidence that age-related FER decline is not uniform across individuals but is shaped by cognitive factors. For instance, Pua and Yu (2024) demonstrated that executive functions—such as inhibition, cognitive flexibility, and working memory—mediate or moderate the impact of aging on FER, with distinct patterns for men and women. These findings underscore the importance of considering cognitive profiles when interpreting individual differences in perceptual thresholds. In this context, ongoing work currently examines whether traits such as emotional intelligence correlate with perceptual thresholds (Gillioz et al., in preparation). Preliminary findings suggest that individuals who score higher on the emotion recognition and emotion understanding facets of emotional intelligence require less signal to accurately recognize dynamic facial expressions. These findings provide deeper insights into the cognitive and affective mechanisms underlying human facial emotion decoding and further validate the utility of perceptual threshold measures in capturing individual differences in emotion processing.

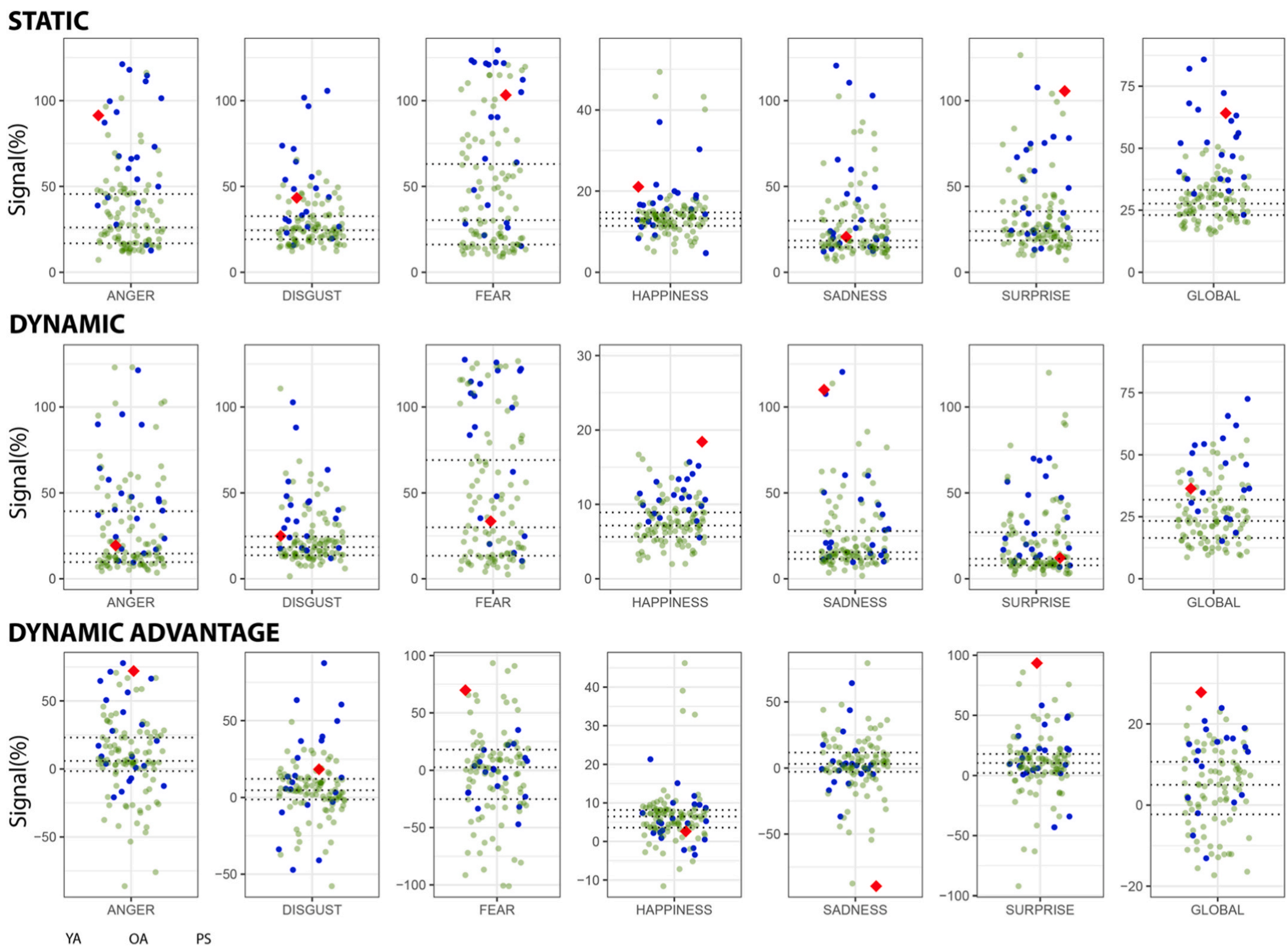

Fig. 2. Percentage of Signal Required for Static and Dynamic FER and Dynamic Advantage in Young-Adults, Older Adults and PS.
Note. From the top to the bottom: signal required to achieve 44% accurate static and dynamic FER and the dynamic advantage (signal static – signal dynamic) in young-adults (YA – green dots), older adults (OA – blue dots) and PS (red diamond) for the six basic facial expressions of emotion and for the recognition performance of all expressions averaged (Global). The scales have been adapted according to the distribution obtained in each condition. Please note that the QUEST algorithm provides a statistical estimation of the quantity of signal the participants require to perform the task. When participants fail to reach the 44% accuracy level even with 100% signal, the algorithm returns a value above the maximum signal.

4.2. Efficiency: minimum trials to reach stabilization

After establishing the reliability and stability of QUEST-derived thresholds, we examined the algorithm's efficiency by comparing the number of trials required to reach stabilization across individuals, expressions, and populations. Overall, our findings indicate that QUEST can achieve stabilization within a limited number of trials for all expressions and modalities in more than 90% of the participants in every population, supporting its suitability for use in both research and clinical contexts.

Notably, the number of trials required to reach stabilization varied across participant groups. Older adults required on average significantly more trials than young adults, to achieve stable threshold estimates in both static and dynamic conditions. This increased trial count in older adults likely reflects greater response variability, consistent with well-documented age-related declines in FER abilities (e.g., Richoz et al., 2018; Hayes et al., 2020; Pua and Yu, 2024). At the level of individual facial expressions, older adults needed significantly more trials than young adults to reach stabilization for the static expression of anger.

In contrast, the prosopagnosic patient PS did not require significantly more trials than older adults across either modality or individual expressions. Despite her known impairments in face processing (Rossion et al., 2003) and facial expression recognition (Richoz et al., 2015), she reached stabilization within a comparable number of trials to those of older adults. Altogether, these findings indicate that the QUEST algorithm can efficiently yield stable threshold estimates even in populations with specific perceptual difficulties.

To our knowledge, this is the first study to systematically determine the number of trials required for QUEST stabilization in the context of facial expression recognition. Although QUEST has been previously implemented in high-level visual tasks, such as FER (e.g., Rodger et al., 2015; Wyssen et al., 2019), none have reported concrete trial counts, which is a critical information when working with populations for whom long testing sessions are not possible. This is important information, as knowing how many trials are actually necessary to derive stable threshold estimates allows researchers to design shorter, more efficient testing protocols without compromising the validity of the data. Populations with limited attentional or cognitive resources, such as older adults (Pua and Yu, 2024), young children (e.g., Conte and Richards, 2021; Reynolds et al., 2013) or clinical groups with neurodegenerative conditions such as Alzheimer (Perry and Hodges, 1999; Mavritsaki et al., 2019) as well as patients with brain lesions (Spaccavento et al., 2019) would greatly benefit from this methodological advancement, for its potential use in diagnostic or screening tools, where rapid and effective assessment of individual abilities is crucial.

4.3. Sensitivity to known FER effects

Our final objective was to assess whether the QUEST algorithm could reveal well-known effects in FER, thus providing meaningful insights into the mechanisms underlying FER. Specifically, we investigated three key phenomena: (1) age-related differences in FER; (2) the ability to distinguish individual profiles by comparing patient PS to age-matched older adults and (3) the dynamic advantage.

To assess the algorithm's sensitivity to these effects, we conducted two analyses for each facial expression and modality. The first focused on the perceptual threshold estimated at the stabilization point, while the second was based on the estimates obtained after the full 80-trials sequence. Importantly, both analyses yielded the same pattern of effects across populations and modalities with only minor variations in significance. Thresholds obtained at stabilization were strongly correlated with those derived from the full 80-trials set and statistical comparison between these two estimates showed no significant differences for most expressions. The only significant differences were found in young adults for sadness in both conditions and happiness in the static condition. We hypothesize that these exceptions may be due to increased noise when administering more trials than necessary, especially for emotions that require fewer trials to achieve stable threshold estimates. Additionally, the lack of correlation for happiness may be explained by the limited inter-individual variability observed for this expression, as reflected in the smallest standard deviation. When variability is minimal, correlations are less likely to emerge, even if underlying associations exist.

Overall, our findings demonstrate that using fewer trials preserves well-established effects and does not compromise inter-observer statistical comparisons. Based on these observations, the following sections will focus exclusively on the thresholds estimated at stabilization.

4.3.1. Age-related differences in signal requirements for FER

Our results revealed that young adults required overall significantly less signal than older adults to recognize facial expressions across both static and dynamic conditions. This difference present across all expressions – excepted sadness – supports the well-established age-related decline in FER abilities and aligns with prior work showing reduced recognition accuracy for dynamic facial expressions (e.g., Cortes et al., 2021; Richoz et al., 2018) and vocal stimuli in aging (Cortes et al., 2021). Furthermore, Pua and Yu (2024) emphasize that executive functioning modulates these declines, offering a framework for understanding why some older adults exhibit greater FER performance than others and providing potential explanations for inter-individual variability. Our findings indicate that older adults require more visual information than young adults to reach the same level of FER accuracy. They are in line with a study by Smith et al. (2018) that used a reverse correlation emotion categorization task with the Bubbles paradigm to map out the facial features that are the most important to recognize facial expressions. Their results revealed that while older and younger adults used the same perceptual strategies to recognize facial expressions, older adults required on average 25 more bubbles (approximatively 15% more visual information) than their younger counterparts to correctly identify emotions. Critically, while both the *Bubbles* technique and QUEST provide quantitative measures of FER sensitivity, QUEST achieves this with a substantially reduced number of trials (<80 trials per expression vs. 2700 trials with *Bubbles*) and in both static and dynamic modalities. This advantage positions QUEST as a unique tool for the rapid and reliable assessment of FER sensitivity, particularly in populations where time and attentional limitations are critical constraints.

4.3.2. Patient PS vs. age-matched controls

Contrary to expectations, PS did not significantly differ from older adults in terms of quantity of signal required to reach recognition accuracy, on both the overall and the individual expressions. Several factors may explain this lack of difference between PS and her age-matched controls. Prior studies have shown that perceptually challenging conditions and task difficulty have a negative impact on facial expression recognition in older adults, particularly when facial features are hidden (for instance while wearing masks; Lenoir et al., 2024) or when emotional expressions are less distinct (e.g., Hayes et al., 2020; Shen and Zuo, 2023; Orgeta, 2010). Given the high difficulty of the current task, it is possible that the performance gap typically observed with full signal faces between older adults and individuals with specific impairments, such as PS, was reduced. Notably, despite her known deficits, PS's performance did not decline further relative to that of older adults. This observation suggests that the increased task demand may have encouraged PS to rely on alternative processing strategies that disproportionally benefited her. Indeed, previous research has evidenced that PS tends to rely more on local facial features than global configurations (Rossion, 2022a, 2022b; Caldara et al., 2005; Fiset et al., 2017; Schaller et al., 2023) - a strategy that may be particularly effective in situations where configural cues are degraded or inaccessible. In these conditions, where older adults may also be forced to rely on alternative and less familiar strategies, PS's long-standing use of non-configural cues may give her a relative advantage. Thus, the absence of a significant difference between PS and her age-matched controls should be interpreted with caution, and not as evidence of preserved FER in PS. Instead, we interpret her performance as an indication that age-related decline and compensatory mechanisms interact to shape FER performance under increased perceptual challenge.

4.3.3. The dynamic advantage

All three populations benefited from dynamic over static presentations, requiring less signal on average to reach accuracy levels in the dynamic condition. This global dynamic advantage was modulated by both the type of expression and the population. More specifically, a dynamic advantage was observed across nearly all FEEs in young adults (except for fear), for anger, happiness and surprise in older adults, and descriptively for all expressions in PS except sadness, which instead showed a static advantage. These findings are in line with previous research showing a dynamic advantage for the recognition of facial expressions in patient PS (Richoz et al., 2015) in older adults (Richoz et al., 2018; Cortes et al., 2021) and in healthy young adults under visually challenging conditions (Richoz et al., 2024).

When examining the strength of the dynamic advantage, our findings revealed that it was more pronounced in PS and in older adults, suggesting that frailer visual systems may rely more on dynamic information to support recognition. PS did not exhibit a stronger overall dynamic advantage than older adults, which might be explained by her expression-specific response profile. In fact, when considering the expressions individually, she showed a larger advantage for fear and surprise, but also a stronger static advantage for sadness, likely reducing the overall dynamic advantage. The static advantage for sadness is consistent with previous research showing that this expression is often better recognized from static images (e.g., Bomfim et al., 2019; Recio et al., 2013; Bould et al., 2008) or when evolving slowly over time (Kamachi et al., 2001). This may be particularly the case for individuals who have difficulties in extracting information from the eye region – such as PS (Caldara et al., 2005) – a facial region known to be critical for the recognition of sadness (Eisenbarth and Alpers, 2011). In such cases, the temporal dynamics of moving faces may disrupt, rather than facilitate, the processing of these critical diagnostic cues.

4.3.4. Potential clinical applications

Taken together these findings highlight QUEST's sensitivity to well-known FER effects. They also demonstrate that while dynamic cues generally enhance FER, this benefit is not uniform across expressions or individuals. Importantly, adaptive tools like the QUEST can help uncover specific response profiles – as illustrated by PS's atypical static advantage for sadness – offering new insights into the functioning of the FER systems under challenging visual conditions, particularly in specific populations. Notably, this adaptive tool could be of strong potential interest for clinical applications, particularly for the early detection of FER deficits in specific populations. These may include, for example, individuals with neurodegenerative conditions such as frontotemporal or Alzheimer's dementia (Jiskoot et al., 2021), those with mild cognitive impairment in the context of Parkinson's disease, for whom impairments in the recognition of disgust have been reported (Chiang et al., 2024), and neurodevelopmental populations such as individuals with autism spectrum disorder (ASD), who often show impaired recognition of anger and fear (Ashwin et al., 2007; Van der Donck et al., 2020).

4.4. *Expression-specific guidance for future studies using QUEST*

In addition to validating the QUEST algorithm as an efficient tool to assess FER sensitivity, our findings also offer practical guidelines for optimizing its use in future studies. Specifically, our results suggest that while a full set of expressions should be tested to preserve the integrity of the decisional space, it may be possible to focus the interpretation on a subset of the most informative expressions. Such an approach would improve testing efficiency and reduce overall testing time as only the selected expressions would need to reach stabilization – even if all the expressions are initially tested to keep the same chance level. The resulting stabilized estimates could then be used as measures of interest or potential clinical markers of FER sensitivity. This is critical when working with developmental or clinical populations or older individuals for whom long testing sessions might not be possible.

The use of single expressions as clinical markers is supported by previous research. For instance, impaired recognition of anger and fear have been identified as markers of autism spectrum disorder (e.g., Ashwin et al., .2007; Van der Donck et al., 2020), while deficits in recognizing disgust are characteristic of Parkinson's disease (Chiang et al., 2024). These findings highlight that certain expressions may be especially informative for characterizing FER vulnerabilities, and therefore particularly useful for targeted analyses in both research and clinical settings.

Our results further suggest that in some contexts, significant differences in FER sensitivity between dynamic and static expressions, or between populations, can be captured by a limited number of informative expressions rather than the full set. For example, among young adults, a dynamic advantage was observed for all expressions except fear, suggesting that any basic expression other than fear – or happiness, which generally shows limited interindividual variability – could be selected to probe the dynamic advantage. In older adults, anger or surprise emerged as particularly informative expressions as they showed a dynamic advantage. When comparing young and older adults, all expressions except sadness were effective in both static and dynamic conditions.

However, the choice of focusing on specific expressions should be considered with care as this approach may not be appropriate in all contexts. In fact, when comparing older adults to the prosopagnosic patient PS, our findings do not strongly support the use of a specific expression to explore FER threshold differences. A significant difference between PS and AM emerges only when examining the magnitude of the dynamic advantage, with PS exhibiting a significantly larger dynamic advantage for fear and surprise, and a stronger static advantage for sadness.

4.5. *Limitations and future directions*

Despite the strengths of the current study, several limitations should be acknowledged. First, while our approach enabled precise estimation of FER thresholds, it was restricted to basic expressions and a fixed set of stimuli, which may constrain generalizability to more nuanced or socially complex emotions (see Rodger and Caldara, 2025). Second, although our findings demonstrate that stabilization can be reached efficiently, we did not assess the temporal stability of these threshold estimates through replication or test–retest reliability. Future studies should therefore examine whether similar thresholds are obtained across repeated sessions and independent samples to confirm robustness, ideally using new participants to avoid potential familiarity or practice biases. Third, our clinical insights are based on a single case of acquired prosopagnosia, which may not reflect the variability typically observed across clinical populations. Additionally, this study considered only one stimulus dimension—the percentage of signal or "phase coherence"—and one psychometric parameter, namely the predetermined threshold for correct recognition performance. To explore the percentage of signal required for FER, we employed the original QUEST method, which was deemed sufficient for such straightforward cases. However, alternative adaptive psychophysical methods exist, including the more recently developed QUEST + method (Watson, 2017) or the Bayesian Psi method (Kontsevich and Tyler, 1999). Future studies should explore whether these or other methods may provide even more sensitive and efficient measures in this context. This is particularly important for paradigms assessing FER and the dynamic advantage as they involve assessing a large number of conditions.

Finally, in the current study we did not randomize the order of accuracy level-modality combinations. This choice was motivated by the need to compare a single case to a control group and to minimize potential learning effects arising from testing easier conditions first. Accuracy level-modality combinations were tested on separate days to reduce fatigue. Nevertheless, the lack of randomization should be considered when interpreting the findings.

4.6. *Constraints on generality*

All participants tested in this experiment were Western Europeans. Therefore, our findings should only be generalized to individuals sharing this cultural background. This limited cultural diversity represents an important limit on the generalization of our findings to Western observers, as cultural factors can influence face processing (for reviews, see Caldara, 2017; Blais and Caldara, 2021; Rodger and Caldara, 2025) as well as the processing and representation of facial expressions (Jack et al., 2012, 2013; Jack, Caldara & Schyns, 2012). Future studies should extend these findings to individuals from different cultures to ensure broader applicability and understanding of FER sensitivity across diverse populations.

5. Conclusions

This study demonstrates that the QUEST threshold-seeking algorithm is a reliable, efficient, and sensitive method for estimating individual FER sensitivity across diverse populations, including young adults, older adults, and a patient with acquired prosopagnosia. By introducing a rigorous stabilization criterion based on the stability of the threshold estimate itself, the algorithm produced stable perceptual thresholds within a limited number of trials for both static and dynamic expressions. Our findings replicated well-established FER effects: older adults and patient PS required more signal to achieve recognition accuracy, confirming age-related decline and impairment. Notably, dynamic stimuli yielded lower thresholds than static ones, particularly in older adults and patient PS, supporting the notion that the dynamic advantage may serve as a sensitive marker of FER vulnerability in fragile systems. Overall, our study provides practical guidance for future research using QUEST with a similar research design. We also discussed how the testing can be optimized by focusing on selected expressions to reduce testing time without sacrificing sensitivity.

By establishing a robust framework for assessing FER thresholds at the individual level, this study lays the groundwork for future single-subject research in affective science and clinical neuropsychology.

Such approaches are critical for capturing individual variability and for advancing our understanding of the mechanisms underlying human emotion decoding.

CRediT authorship contribution statement

L. Stacchi: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **F. Poncet:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation. **V. Benedetti:** Writing – review & editing, Methodology, Conceptualization. **R. Caldara:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **A.-R. Richoz:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization.

Acknowledgments, fundings

This study was supported by the grant no 10001C_201145 from the Swiss National Science Foundation awarded to RC and ARR (Funding acquisition). RC, ARR, LS and VB contributed to Conceptualization and Methodology. LS and FP contributed to Data curation and Formal Analysis. FP and LS served as lead for writing–original draft. RC, ARR, LS, VB and FP contributed equally to writing, review and editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Roberto Caldara and Anne-Raphaelle Richoz reports financial support was provided by Swiss National Science Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.metip.2026.100241.

Data availability

They are available on OSF, the link to access is in the manuscript

References

Alves, N.T., 2013. Recognition of static and dynamic facial expressions: a study review. Estud. Psicolog. 18, 125–130. https://doi.org/10.1590/S1413-294X2013000100020.

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, 403–410. https://doi.org/10.1111/j.0956-7976.2005.01548.x.

Anderson, S.F., Kelley, K., Maxwell, S.E., 2017. Sample-size planning for more accurate statistical power: a method adjusting sample effect sizes for publication bias and uncertainty. Psychol. Sci. 28 (11), 1547–1562. https://doi.org/10.1177/0956797617723724.

Ashwin, C., Baron-Cohen, S., Wheelwright, S., O'Riordan, M., Bullmore, E.T., 2007. Differential activation of the amygdala and the 'social brain' during fearful face-processing in asperger syndrome. Neuropsychologia 45 (1), 2–14.

Blais, C., Caldara, R., 2021. 16 culture shapes face processing. The Oxford handbook of cultural neuroscience and global mental health 309.

Blais, C., Fiset, D., Roy, C., Saumure Régimbald, C., Gosselin, F., 2017. Eye fixation patterns for categorizing static and dynamic facial expressions. Emotion (Washington, D.C.) 17 (7), 1107–1119. https://doi.org/10.1037/emo0000283.

Bomfim, A.J. de L., Ribeiro, R.A., dos, S., Chagas, M.H.N., 2019. Recognition of dynamic and static facial expressions of emotion among older adults with major depression. Trends in Psychiatry and Psychotherapy 41 (2), 159–166. https://doi.org/10.1590/2237-6089-2018-0054.

Bould, E., Morris, N., Wink, B., 2008. Recognising subtle emotional expressions: the role of facial movements. Cognit. Emot. 22 (8), 1569–1587. https://doi.org/10.1080/02699930801921156.

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10 (4), 433–436. https://doi.org/10.1163/156856897X00357.

Caldara, R., 2017. Culture reveals a flexible system for face processing. Curr. Dir. Psychol. Sci. 26 (3), 249–255.

Caldara, R., Schyns, P., Mayer, E., Smith, M.L., Gosselin, F., Rossion, B., 2005. Does prosopagnosia take the eyes out of face representations? Evidence for a defect in representing diagnostic facial information following brain damage. J. Cognit. Neurosci. 17 (10), 1652–1666. https://doi.org/10.1162/089892905774597254.

Calder, A.J., Keane, J., Manly, T., Sprengelmeyer, R., Scott, S., Nimmo-Smith, I., Young, A.W., 2003. Facial expression recognition across the adult life span. Neuropsychologia 41 (2), 195–202. https://doi.org/10.1016/s0028-3932(02)00149-5.

Chiang, K.-W., Tan, C.H., Yu, R.L., 2024. Disgust-specific impairment of facial emotion recognition in parkinson's disease patients with mild cognitive impairment. Soc. Cognit. Affect Neurosci. https://doi.org/10.1093/scan/nsae073.

Conte, S., Richards, J., 2021. Attention in early development. In: Oxford Research Encyclopedia of Psychology.

Cortes, D.S., Tornberg, C., Bänziger, T., Elfenbein, H.A., Fischer, H., Laukka, P., 2021. Effects of aging on emotion recognition from dynamic multimodal expressions and vocalizations. Sci. Rep. 11 (1), 2647. https://doi.org/10.1038/S41598-021-82135-1.

Crawford, J.R., Garthwaite, P.H., 2002. Investigation of the single case in neuropsychology: confidence limits on the abnormality of test scores and test score differences. Neuropsychologia 40 (8), 1196–1208. https://doi.org/10.1016/s0028-3932(01)00224-x.

Crawford, J.R., Howell, D.C., 1998. Comparing an individual's test score against norms derived from small samples. Clin. Neuropsychol. 12 (4), 482–486. https://doi.org/10.1076/clin.12.4.482.7241.

Cunningham, D.W., Wallraven, C., 2009. Dynamic information for the recognition of conversational expressions. J. Vis. 9 (13), 1–7. https://doi.org/10.1167/9.13.7.

Dobs, K., Bulthoff, I., Schultz, J., 2018. Use and usefulness of dynamic face stimuli for face perception studies-a review of behavioral findings and methodology. Front. Psychol. 9, 1355. https://doi.org/10.3389/fpsyg.2018.01355.

Eisenbarth, H., Alpers, G.W., 2011. Happy mouth and sad eyes: scanning emotional facial expressions. Emotion 11 (4), 860–865. https://doi.org/10.1037/a0022758.

Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39, 175–191.

Fiset, D., Blais, C., Royer, J., Richoz, A.R., Dugas, G., Caldara, R., 2017. Mapping the impairment in decoding static facial expressions of emotion in prosopagnosia. Soc. Cognit. Affect Neurosci. 12 (8), 1334–1341. https://doi.org/10.1093/scan/nsx068.

Gillioz, C., Stacchi, L., Richoz, A.R., Caldara, R., & Fiori, M. (in preparation). Emotional Intelligence Lowers the Perceptual Threshold for Recognizing Facial Expressions of Emotion.

Gold, J.M., Barker, J.D., Barr, S., Bittner, J.L., Bromfield, W.D., Chu, N., Srinath, A., 2013. The efficiency of dynamic and static facial expression recognition. J. Vis. 13 (5), 1–12.

Hayes, G.S., McLennan, S.N., Henry, J.D., Phillips, L.H., Terrett, G., Rendell, P.G., Pelly, R.M., Labuschagne, I., 2020. Task characteristics influence facial emotion recognition age-effects: a meta-analytic review. Psychol. Aging 35 (2), 295–315. https://doi.org/10.1037/PAG0000441.

Jack, R.E., Caldara, R., Schyns, P.G., 2012a. Internal representations reveal cultural diversity in expectations of facial expressions of emotion. J. Exp. Psychol. Gen. 141 (1), 19.

Jack, R.E., Garrod, O.G., Yu, H., Caldara, R., Schyns, P.G., 2012b. Facial expressions of emotion are not culturally universal. Proc. Natl. Acad. Sci. 109 (19), 7241–7244.

Jack, R.E., Garrod, O.G., Yu, H., Caldara, R., Schyns, P.G., 2013. Reply to sauter and eisner: differences outweigh commonalities in the communication of emotions across human cultures. Proc. Natl. Acad. Sci. 110 (3), E181–E182.

Jiskoot, L.C., Poos, J.M., Vollebergh, M.E., Franzen, S., van Hemmen, J., Papma, J.M., et al., 2021. Emotion recognition of morphed facial expressions in presymptomatic and symptomatic frontotemporal dementia, and Alzheimer's dementia. J. Neurol. 268 (1), 102–113.

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., Akamatsu, S., 2001. Dynamic 2 properties influence the perception of facial expressions. Perception 30 (7), 875–887. https://doi.org/10.1068/p3131.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., 2007. What's new in 382 psychtoolbox-3. Perception 36, 1–16.

Kontsevich, L.L., Tyler, C.W., 1999. Bayesian adaptive estimation of psychometric slope and threshold. Vis. Res. 39, 2729–2737.

Krumhuber, E.G., Skora, L.I., Hill, H.C.H., Lander, K., 2023. The role of facial movements in emotion recognition. Nat. Rev. Psychol. 2, 283–296. https://doi.org/10.1038/s44159-023-00172-1.

Leek, M.R., 2001. Adaptive procedures in psychophysical research. Percept. Psychophys. 63 (8), 1279–1292. https://doi.org/10.3758/bf03194543.

Lenoir, H., Coqué, R., David, C., Demonceaux, E., Belkaid, D., Arnold, G., Siéroff, É., 2024. Le port du masque affecte l'identification des expressions faciales émotionnelles, surtout chez les personnes âgées. Gériatrie et Psychologie Neuropsychiatrie Du Vieillissement 22 (2), 200–208. https://doi.org/10.1684/pnv.2024.1175.

Mavritsaki, E., Bowman, H., Su, L., 2019. Attentional deficits in Alzheimer's disease: investigating the role of acetylcholine with computational modelling. In: Multiscale Models of Brain Disorders. Springer, pp. 113–126. https://doi.org/10.1007/978-3-030-18830-6_11.

Orgeta, V., 2010. Effects of age and task difficulty on recognition of facial affect. J. Gerontol. B Psychol. Sci. Soc. Sci. 65B (3), 323–327. https://doi.org/10.1093/geronb/gbq007.

Pua, S.Y., Yu, R.L., 2024. Effects of executive function on age-related emotion recognitiondecline varied by sex. Soc. Sci. Med. 361, 117392.

Perry, R.J., Hodges, J.R., 1999. Attention and executive deficits in Alzheimer's disease: a critical review. Brain 122 (3), 383–404. https://doi.org/10.1093/brain/122.3.383.

Recio, G., Schacht, A., Sommer, W., 2013. Classification of dynamic facial expressions of emotion presented briefly. Cognit. Emot. 27 (8), 1486–1494. https://doi.org/10.1080/02699931.2013.794128.

Reynolds, G.D., Courage, M.L., Richards, J.E., 2013. The development of attention. In: Oxford Handbook of Cognitive Psychology.

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. https://doi.org/10.1016/j.cortex.2014.11.015.

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. https://doi.org/10.1167/18.9.5.

Richoz, A.-R., Stacchi, L., Schaller, P., Lao, J., Papinutto, M., Ticcinelli, V., Caldara, R., 2024. Recognizing facial expressions of emotion amid noise: a dynamic advantage. J. Vis. 24 (1), 1–22. https://doi.org/10.1167/jov.24.1.7.

Riddoch, J.M., Humphreys, G.W., 1993. BORB: Birmingham Object Recognition Battery, first ed. Psychology Press. https://doi.org/10.4324/9781003069645.

Rodger, H., Caldara, R., 2025. The perception of facial expressions of emotion. In: Armony, J., Vuilleumier, P. (Eds.), The Cambridge Handbook of Human Affective Neuroscience, second ed. Cambridge University Press, Cambridge, UK.

Rodger, H., Lao, J., Caldara, R., 2018. Quantifying facial expression signal and intensity use during development. J. Exp. Child Psychol. 174, 41–59. https://doi.org/10.1016/j.jecp.2018.05.005. PMID: 29906651.

Rodger, H., Vizioli, L., Ouyang, X., Caldara, R., 2015. Mapping the development of facial expression recognition. Dev. Sci. 18 (6), 926–939.

Rossion, B., Caldara, R., Seghier, M., Schuller, A.M., Lazeyras, F., Mayer, E., 2003. A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. Brain 126 (11), 2381–2395. https://doi.org/10.1093/brain/awg241.

Rossion, B., 2022a. Twenty years of investigation with the case of prosopagnosia PS to understand human face identity recognition. Part I: function. Neuropsychologia 173, 1–53. https://doi.org/10.1016/j.neuropsychologia.2022.108278, 108278.

Rossion, B., 2022b. Twenty years of investigation with the case of prosopagnosia PS to understand human face identity recognition. Part II: neural basis. Neuropsychologia 173, 1–38. https://doi.org/10.1016/j.neuropsychologia.2022.108279, 108279.

Schaller, P., Caldara, R., Richoz, A.-R., 2023. Prosopagnosia does not abolish other-race effects. Neuropsychologia 180, 108479. https://doi.org/10.1016/j.neuropsychologia.2023.108479.

Shen, X., Zuo, L., 2023. Facial occlusion affects emotional face recognition differently in older adults and children. In: 2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP), 1778-1781. https://doi.org/10.1109/ICSP58490.2023.10248831.

Smith, M.L., Grühn, D., Bevitt, A., Ellis, M., Ciripan, O., Scrimgeour, S., Papasavva, M., Ewing, L., Ewing, L., Ewing, L., 2018. Transmitting and decoding facial expressions of emotion during healthy aging: more similarities than differences. J. Vis. 18 (9), 10. https://doi.org/10.1167/18.9.10.

Sorger, B., Goebel, R., Schiltz, C., Rossion, B., 2007. Understanding the functional neuroanatomy of acquired prosopagnosia. Neuroimage 35 (2), 836–852. https://doi.org/10.1016/j.neuroimage.2006.09.051.

Spaccavento, S., Marinelli, C.V., Nardulli, R., Macchitella, L., Bivona, U., Piccardi, L., Zoccolotti, P., Angelelli, P., 2019. Attention deficits in stroke patients: the role of lesion characteristics, time from stroke, and concomitant neuropsychological deficits. Behav. Neurol. 2019. https://doi.org/10.1155/2019/7835710. Article 7835710.

Stacchi, L., Ramon, M., Lao, J., Caldara, R., 2019. Neural representations of faces are tuned to eye movements. J. Neurosci. 39 (21), 4113–4123. https://doi.org/10.1523/JNEUROSCI.2968-18.2019.

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*.

Tardif, C., Lainé, F., Rodriguez, M., Gepner, B., 2007. Slowing down presentation of facial movements and vocal sounds enhances facial expression recognition and induces facial-vocal imitation in children with autism. J. Autism Dev. Disord. 37, 1469–1484.

The MathWorks Inc, 2022. MATLAB Version: 9.13.0 (R2022B). The MathWorks Inc, Natick, Massachusetts. https://www.mathworks.com.

Treutwein, B., 1995. Adaptive psychophysical procedures. Vis. Res. 35 (17), 2503–2522.

Van der Donck, S., Dzhelyova, M., Vettori, S., Mahdi, S.S., Claes, P., Steyaert, J., Boets, B., 2020. Rapid neural categorization of angry and fearful faces is specifically impaired in boys with autism spectrum disorder. JCPP (J. Child Psychol. Psychiatry) 61 (9), 1019–1029. https://doi.org/10.1111/JCPP.13201.

Watson, A.B., Pelli, D.G., 1983. QUEST: a Bayesian adaptive psychometric method. Percept. Psychophys. 33 (2), 113–120. https://doi.org/10.3758/bf03202828.

Watson, A.B., 2017. QUEST+: a general multidimensional Bayesian adaptive psychometric method. J. Vis. 17 (3), 1–27. https://doi.org/10.1167/17.3.10.

Widen, S.C., Russell, J.A., 2015. Do dynamic facial expressions convey emotions to children better than do static ones? J. Cognit. Dev. 16 (5), 802–811. https://doi.org/10.1080/15248372.2014.916295.

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. https://doi.org/10.3758/BRM.42.3.671.

Wyssen, A., Lao, J., Rodger, H., Humbel, N., Lennertz, J., Schuck, K., Isenschmid, B., Milos, G., Trier, S., Whinyates, K., Assion, H.-J., Ueberberg, B., Müller, J., Klauke, B., Teismann, T., Margraf, J., Juckel, G., Kossmann, C., Schneider, S., Caldara, R., Munsch, S., 2019. Facial emotion recognition abilities in women suffering from eating disorders. Psychosom. Med.: J. Behav. Med. 81, 155–164.