Journal of the International Neuropsychological Society (2015), **21**, 568–572.

doi:10.1017/S1355617715000557

BRIEF COMMUNICATION

"Now I see it, now I don't": Determining Threshold Levels of Facial Emotion Recognition for Use in Patient Populations

Isabelle Chiu,[1,2] Regina I. Gfrörer,[3] Olivier Piguet,[4,5] Manfred Berres,[6] Andreas U. Monsch,[2] AND Marc Sollberger[1,2]

[1]Department of Neurology, University Hospital Basel, Basel, Switzerland
[2]Memory Clinic, University Center for Medicine of Aging Basel, Felix Platter-Hospital, Basel, Switzerland
[3]Department of Psychology, University of Basel, Basel, Switzerland
[4]Neuroscience Research Australia, and the University of New South Wales, Sydney, Australia
[5]ARC Centre of Excellence in Cognition and Its Disorders, Sydney, Australia
[6]Department of Mathematics and Technology, University of Applied Sciences Koblenz, Koblenz, Germany

(Received January 29, 2015; Final Revision June 12, 2015; Accepted June 15, 2015)

Abstract

The importance of including measures of emotion processing, such as tests of facial emotion recognition (FER), as part of a comprehensive neuropsychological assessment is being increasingly recognized. In clinical settings, FER tests need to be sensitive, short, and easy to administer, given the limited time available and patient limitations. Current tests, however, commonly use stimuli that either display prototypical emotions, bearing the risk of ceiling effects and unequal task difficulty, or are cognitively too demanding and time-consuming. To overcome these limitations in FER testing in patient populations, we aimed to define FER threshold levels for the six basic emotions in healthy individuals. Forty-nine healthy individuals between 52 and 79 years of age were asked to identify the six basic emotions at different intensity levels (25%, 50%, 75%, 100%, and 125% of the prototypical emotion). Analyses uncovered differing threshold levels across emotions and sex of facial stimuli, ranging from 50% up to 100% intensities. Using these findings as "healthy population benchmarks", we propose to apply these threshold levels to clinical populations either as facial emotion recognition or intensity rating tasks. As part of any comprehensive social cognition test battery, this approach should allow for a rapid and sensitive assessment of potential FER deficits. (*JINS*, 2015, *21*, 568–572)

Keywords: Threshold, Emotion recognition, Facial expression, Basic emotions, Face morphing, Neuropsychological disorder

INTRODUCTION

Disturbances in emotion processing are often found in patients with psychiatric diseases such as schizophrenia (e.g., Kring & Elis, 2013) and neurodegenerative diseases such as behavioral variant frontotemporal dementia (e.g., Piguet, Hornberger, Mioshi, & Hodges, 2011). Thus, measures of emotion processing, such as facial emotion recognition (FER) tests, should be an important component in any comprehensive neuropsychological assessment of individuals with neuropsychological disorders. This is particularly so as emotion recognition is part of the new domain "social cognition" introduced in the classification of neurocognitive disorders in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Accordingly, measures of emotion recognition need to be part of any comprehensive social cognition test battery. Yet, it remains a challenge to adequately measure facial emotion recognition in individuals with neuropsychological disorders.

In clinical settings, FER tests need to be sensitive, short, and easy to administer, given the limited time available and patient limitations. The most common FER tests use stimuli that display prototypical full-blown facial emotions (e.g., Ekman, 1992; Funkiewiez, Bertoux, de Souza, Levy, & Dubois, 2012; Naranjo et al., 2011). These tests are easy and quick to administer but are also associated with ceiling effects in individuals without emotion processing deficits and bear the risk of unequal task difficulty (e.g., Suzuki, Hoshino, Shigemasu, & Kawamura, 2007). Accordingly, mild deficits in emotion recognition might be missed in patient populations and different types of emotions might be recognized differently, not exclusively due to the brain disorder, but because of methodological biases.

To overcome ceiling effects, FER tests with increasing facial emotion intensities (e.g., Montagne, Kessels, De Haan,

Correspondence and reprint requests to: Marc Sollberger, Department of Neurology, University Hospital Basel, Schanzenstrasse 55, 4031 Basel, Switzerland. E-mail: marc.sollberger@usb.ch

& Perrett, 2007; Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002) or mixed facial expressions (Suzuki, Hoshino, & Shigemasu, 2006; Young et al., 2002) have been developed. Moreover, tests comprising movie clips have been developed to measure FER in ecologically valid situations (e.g., McDonald, Glanagan, Rollins, & Kinch, 2003). However, these tests are time-consuming and cognitively quite demanding. Accordingly, they might be difficult to apply in patients with cognitive impairment, especially when combined with psychomotor slowing and/or little motivation in testing. In addition, such tasks may result in FER deficits in some patient populations because of associated cognitive deficits (e.g., attention) (Miller et al., 2012), rather than due to a primary emotion processing deficit.

The aim of this study was to identify the facial emotional stimuli that represent the “tipping points” of FER for each basic emotion in healthy individuals (i.e., at least 50% correct recognition) to develop a short, sensitive, and easily applicable FER test with a task difficulty as equal as possible across different emotions. We, therefore, defined FER threshold levels for each basic emotion (i.e., anger, disgust, fear, happiness, sadness, and surprise) by using stimuli from the Facial Expressions of Emotion – Stimuli and Tests (FEEST) at different intensity levels (i.e., 25%, 50%, 75%, 100%, and 125% of the prototypical emotion) (Young et al., 2002). We hypothesized that FER threshold levels would vary depending on the type of emotion.

METHOD

Participants

Forty-nine healthy individuals [male = 26(53%), female = 23(47%); age: $M = 65.3$ years, $SD = 7.9$, range = 52–79 years; education: $M = 14.8$ years, $SD = 3.2$] from the participant pool of the Memory Clinic Basel, Switzerland, took part in the study (for sample size calculation, see Supplementary Materials, which are available online). Notably, participants were recruited in similar numbers in terms of age [age 50–59 years ($n = 15$), age 60–69 years ($n = 18$), and age 70–79 years ($n = 16$)] and evenly distributed in sex by age [$\chi^2(2, N = 49) = 0.42$; $p = .811$]. The study was approved by the local ethics committee and all participants provided informed consent.

Before enrolment into the study, participants completed a standardized medical questionnaire of the Memory Clinic Basel, the Geriatric Depression Scale (GDS) (Sheikh & Yesavage, 1986) (GDS: $M = 0.67$; $SD = 0.99$), the Mini Mental State Examination (MMSE) (Folstein, Folstein, & McHugh, 1975) (MMSE: $M = 28.69$; $SD = 1.04$) and the Clock Drawing Test (CDT) (Thalmann et al., 2002) (CDT: $M = 6.80$; $SD = 0.46$). Inclusion criteria were a minimum education of seven years. Exclusion criteria were history of current drug or alcohol abuse [according to the Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV)], head trauma (with loss of consciousness greater than 30 min), systemic disorders or brain diseases that could result in neuropsychological deficits, chronic pain thought to significantly interfere with neuropsychological testing, psychiatric disorders (according to DSM-IV), general anesthesia within the last 3 months, and ≤6 points in a combined analysis of the MMSE and CDT (Thalmann et al., 2002).

Facial Emotion Recognition Test

Facial stimuli were taken from the morphed and caricatured continua of the FEEST (Young et al., 2002). These stimuli are in greyscale and the hairline is masked. A male (model J.J.) and a female (model M.O.) model reported to be of consistent quality for each emotion across the continua and showing a reasonably standardized pose and lighting were used (Young et al., 2002). In each model, the six basic emotions (i.e., anger, disgust, fear, happiness, sadness, and surprise) (Ekman, 1992; Ekman & Friesen, 1978) were displayed at different intensities from a neutral expression (0%) to a caricatured expression (125%), with intermediate intensities (25%, 50%, 75%, and 100%). By combining the three dimensions (i.e., type of emotion, intensity level, and sex of the facial stimulus) plus a neutral stimulus for each sex, a total of 62 stimuli ($6 \times 5 \times 2 + 2 = 62$) was obtained.

To control for sequence effects, stimuli were presented in a pseudo-randomized order so that no emotion was shown more than twice in a row. Three versions of the task were designed and randomly allocated to study participants. The series of pictures was presented on a 15-inch laptop computer using E-Prime 1.2 software (http://pstnet.com/index.cfm). Instructions were displayed on the screen. Stimuli remained on the screen until participants provided answer. Participants were asked to identify the type of facial emotion shown (see Supplementary Material Figure 1). They could respond either by pointing at the label on the screen or verbally. The answer was typed in on a separate keyboard by the investigator. After the answer was given, the picture remained on the screen and participants were asked to rate the confidence of their response on a 5-point Likert scale (i.e., 1 = very unsure, 2 = unsure, 3 = undecided, 4 = sure, and 5 = very sure) (see Supplementary Material Figure 2). Before the experiment started, this procedure was carried out in five practice trials to ensure that participants had understood the task. Confidence ratings were used to assess the reliability of participants’ responses and to minimize the possibility of correct responses by guessing (correct response with low confidence).

Data Analyses

For each stimulus, the average accuracy rate and the average confidence rating were calculated. In addition, to ensure that confidence ratings were legitimate (correct response and confidence rating ≥ undecided), we calculated confidence ratings for correct responses only. As we aimed at setting threshold levels of facial emotion recognition slightly above the zone of ambiguity, we determined threshold levels by the lowest intensity level for each emotion and model (male or female) with a minimum accuracy rate of 50% and a minimum confidence rating of 50% (≥ undecided) for correct responses only.

To better understand participants' response pattern, we created six categories of confidence: (a) legitimately sure (correct response and confidence rating > undecided), (b) illegitimately sure (incorrect response and confidence rating > undecided), (c) legitimately undecided (incorrect response and confidence rating = undecided), (d) illegitimately undecided (correct response and confidence rating = undecided), (e) legitimately unsure (incorrect response and confidence rating < undecided), and (f) illegitimately unsure (correct response and confidence rating < undecided).

To examine accuracy rates depending on type of emotion and intensity, a generalized linear mixed-effects model (GLMM) with a binomial error distribution and pairwise contrasts was performed. Neutral stimuli were excluded from these analyses. Within-subject factors were type of emotion and intensity level. Random effects accounted for correlations within participants.

We also performed GLMM to compare the emotion recognition accuracy rates of the different emotions at threshold level and at 100% intensity (i.e., prototypical emotion). In addition, to investigate if accuracy rates reached similar levels for all emotions, we performed GLMM across all emotion recognition accuracy rates at the intensity level each emotion was recognized best.

Lastly, to explore patterns of wrong answers, we created a table with percentages of emotion identified for each type of emotion presented.

Age, education, sex of participants, sex of the facial stimuli, and randomization of stimuli presentation had no significant effects on the pattern of results. Therefore, we excluded these covariates from the final statistical models.

RESULTS

Recognition threshold levels (i.e., level of intensity at which an emotion was correctly recognized by at least 50% of participants with a confidence of at least 50% for correct responses only) differed between emotions and sex of stimuli and ranged between 50% and 100% of emotional intensity (Table 1). Accuracy rates at threshold levels varied and ranged between 51% (fear, male facial stimulus) to 92% (happiness, male facial stimulus) (Supplementary Table 1). Mean confidence ratings for correct responses only were higher than undecided (≥ 3) at every level of intensity, including infra-threshold levels.

A significant main effect of emotion on accuracy was observed, $F(5,2910) = 22.26$, $p < .001$. Pairwise contrasts across all levels of intensity showed that the recognition of happiness was easier than the recognition of all other emotions (Supplementary Table 2). Notably, recognition of happiness differed most from the other emotions at 25% intensity (Figure 1, Supplementary Tables 1 and 3). Overall, both surprise and disgust were more easily recognized than anger, fear, and sadness. Sadness, whose accuracy rate was close to the one of fear, was the most difficult emotion to recognize (Supplementary Table 2). FER accuracy rates of the different emotions did not reach similarly high levels, $F(5,582) = 7.34$, $p < .001$ (Figure 1), suggesting ceiling effects in the recognition of each type of emotion at different accuracy levels.

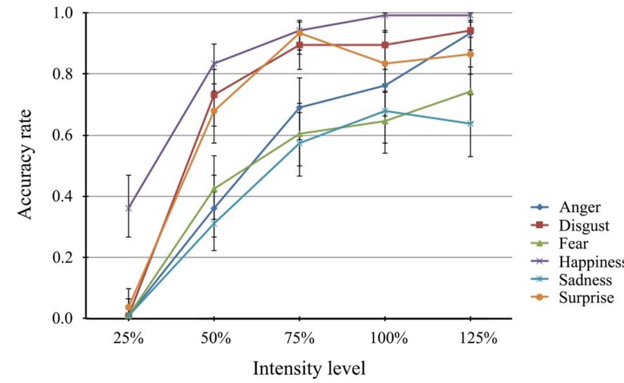

Fig. 1. Mean accuracy rate for each facial emotion over intensity levels. Error bars depict 95% confidence intervals.

A significant main effect of intensity on accuracy was also uncovered, $F(4,2910) = 77.15$, $p < .001$. Contrasts revealed that accuracy increased significantly from 25% to 50% intensity, $t(2910) = -18.88$, $p < .001$, $r = .55$, and from 50% to 75% intensity, $t(2910) = -8.24$, $p < .001$, $r = .25$. In contrast, accuracy did not increase significantly from 75% to 100% intensity, $t(2910) = -1.37$, $p = .171$, $r = .04$, and from 100% to 125% intensity, $t(2910) = -1.54$, $p = .124$, $r = .05$. Examining each emotion separately indicated that accuracy in recognizing happiness, disgust, fear, and sadness did not significantly increase beyond 75% intensity ($p > .05$) (Figure 1), indicating ceiling effects in the recognition of these four emotions above 75% intensity. Accuracy in recognizing surprise did not significantly increase above 100% intensity, whereas the accuracy in recognizing anger increased significantly up to 125% (Figure 1).

In addition, a significant emotion by intensity interaction was present, $F(20,2910) = 2.44$, $p < .001$, indicating that the accuracy in recognizing emotions varied as a function of the intensity level and type of emotion.

FER accuracy rates between the six emotions differed both at threshold level, $F(5,582) = 4.21$, $p = .001$, and at the level of the prototypical (i.e., 100% intensity) facial emotions, $F(5,582) = 6.75$, $p < .001$. The different magnitudes of the two F-statistics suggest that differences in emotion recognition between emotions were lower at threshold level than at the level of prototypical full-blown emotions. Notably, *post hoc* analyses revealed that differences in FER accuracy rates at threshold level were driven by different accuracy rates of happiness and fear ($p < .001$).

Table 1. Identified threshold levels (%) for each basic emotion and model

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Male model (J.J.)	75	50	75	50	75	50
Female model (M.O.)	75	50	50	50	100	50

Over all responses, the six confidence categories were distributed as follows: legitimately sure = 56.3%, illegitimately sure = 28.5%, legitimately undecided = 6.6%, illegitimately undecided = 3.2%, legitimately unsure = 3.0%, and illegitimately unsure = 2.4% (Supplementary Table 1). Notably, the percentage of illegitimately unsure responses was below 9% at any level of intensity. Apart from surprise (male model), the percentage of legitimately sure responses increased and the percentage of illegitimately sure responses decreased with increasing intensity level (Supplementary Table 1). Percentage of illegitimately sure responses was highest at 25% intensity (Supplementary Table 1), reflecting the high misrecognition of stimuli as neutral (Supplementary Table 3).

Regarding misrecognition, neutral was the most frequent wrong answer, particularly at low emotional intensities (Supplementary Table 3). Two types of misrecognition by another basic emotion reached a level of more than 10%: fear was misrecognized as surprise in 16% and anger was misrecognized as fear in 11%.

DISCUSSION

This study investigated FER threshold detection of the basic emotions in healthy adults. As hypothesized, we found differing threshold levels depending on the type of emotion and sex of the facial stimulus, ranging from 50% (e.g., happiness, male stimulus) to 100% (sadness, female stimulus) intensities. At threshold level, the different emotions were not recognized with the same ease, but were more comparable than at the level of the prototypical emotional stimuli.

Across all levels of intensity, happiness was the easiest and sadness the most difficult emotion to recognize. These findings are in line with previous studies examining FER in healthy subjects using prototypical facial emotional stimuli (e.g., Isaacowitz et al., 2007; Mill, Allik, Realo, & Valk, 2009).

Accuracy rates in FER increased with higher levels of intensity, reaching—apart from anger—ceiling effects at and below the level of prototypical facial emotional stimuli. Of interest, ceiling effects of the different emotions were not at similar levels of emotion recognition accuracy. In general, maximum accuracy rates of more easily recognized emotions such as happiness were higher than more difficult to recognize emotions such as sadness. These findings suggest that the task difficulty of the administered facial stimuli is inherently different across the different emotions. Differences in task difficulty, though, seem not to be based on the facial stimuli we applied. A recently published study on the multidimensional assessment of perception and recognition of facial emotions in healthy adults found that happiness was best perceived and recognized across all emotions and tasks (Wilhelm, Hildebrandt, Manske, Schacht, & Sommer, 2014). Similarly, in our study, happiness was the only emotion, which was already recognized at 25% intensity by some participants, showing accuracy rates of 39% (male stimulus) and 35% (female stimulus). The other emotions were not yet recognized as an emotion but as neutral instead.

Mean accuracy rate at threshold levels was 68%, ranging between 51% (fear, male stimulus) and 92% (happiness, male stimulus). The high accuracy rate of happiness at threshold level resulted from large changes in the accuracy rates of happiness between consecutive low-intensity levels, that is, between 25% (39% accuracy) and 50% (92% accuracy) intensities. Consequently, task difficulty was not equivalent across emotions at threshold levels. Task difficulty at threshold levels was, however, more comparable than at the level of prototypical facial emotions.

In line with previous studies examining FER in healthy individuals (e.g., Mill et al., 2009; Rapcsak et al., 2000), fear was most often misrecognized by another basic emotion. In 16%, fear was misrecognized as surprise, and surprise was misrecognized as fear in 9%. This relatively high rate of misrecognition may result from the action units (systematic categorization of physical expressions of emotion) shared by these two emotions (Ekman & Friesen, 1978). Whereas combinations of other basic emotions share between zero and two action units, fear and surprise share four action units (i.e., inner brow raiser, outer brow raiser, upper lid raiser, and jaw drop) (Young et al., 2002).

It is important to note that the rather large intervals in facial emotion intensities may partly account for the still unequal task difficulty of facial stimuli at threshold levels reported here. This potential limitation could be addressed in a study where smaller intervals in facial emotion intensity around the established threshold levels are used. In addition, future studies will be needed to ensure that our findings are generalizable beyond the two models used here.

In conclusion, we defined threshold levels of FER for the six basic emotions in mid- and later-life healthy individuals using a male (model J.J.) and female (model M.O.) stimulus from the FEEST. This manipulation allows us to determine a task difficulty that is quite similar across emotions at these experimentally identified levels. It, therefore, allows for reliable comparison of FER between different emotions in healthy individuals and also in patients with neuropsychological disorders. Using these findings as "healthy population benchmarks," we propose to apply these threshold levels to clinical populations either as facial emotion recognition or intensity rating tasks to investigate their power in discriminating patients with different brain diseases. As part of any comprehensive social cognition test battery, this approach would also allow for a rapid and sensitive assessment of potential FER deficits. In addition, this approach should also help refine patient groups' clinical phenotypes based on their performances in FER. Moreover, it might also serve as a clinical endpoint in investigations of patient populations known to experience emotion processing and social cognition disturbances such as behavioral-variant frontotemporal dementia.

ACKNOWLEDGMENTS

I.C. and M.S. are supported by the Research Fund of the University of Basel (grant number DMS2223), the Velux Foundation (grant number 831), and the Swiss Alzheimer's Association. O.P. is supported by a National Health and Medical Research Council of Australia Career Development Fellowship (grant number APP1022684). All authors report no conflict of interest. Drs. Chiu and Gfrörer contributed equally to this work.

Supplementary Materials

For supplementary material referred to in this article, Please visit http://dx.doi.org/10.1017/S1355617715000557

REFERENCES

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200.

Ekman, P., & Friesen, W.V. (1978). *Facial Action Coding System: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12* (3), 189–198.

Funkiewiez, A., Bertoux, M., de Souza, L.C., Levy, R., & Dubois, B. (2012). The SEA (Social cognition and Emotional Assessment): A clinical neuropsychological tool for early diagnosis of frontal variant of frontotemporal lobar degeneration. *Neuropsychology*, *26*(1), 81–90.

Isaacowitz, D.M., Lockenhoff, C.E., Lane, R.D., Wright, R., Sechrest, L., Riedel, R., & Costa, P.T. (2007). Age differences in recognition of emotion in lexical stimuli and facial expressions. *Psychology and Aging*, *22*(1), 147–159.

Kring, A.M., & Elis, O. (2013). Emotion deficits in people with schizophrenia. *Annual Review of Clinical Psychology*, *9*, 409–433.

McDonald, S., Flanagan, S., Rollins, J., & Kinch, J. (2003). TASIT: A new clinical tool for assessing social perception after traumatic brain injury. *The Journal of Head Trauma Rehabilitation*, *18*(3), 219–238.

Mill, A., Allik, J., Realo, A., & Valk, R. (2009). Age-related differences in emotion recognition ability: A cross-sectional study. *Emotion*, *9*(5), 619–630.

Miller, L.A., Hsieh, S., Lah, S., Savage, S., Hodges, J.R., & Piguet, O. (2012). One size does not fit all: Face emotion processing impairments in semantic dementia, behavioural-variant frontotemporal dementia and Alzheimer's disease are mediated by distinct cognitive deficits. *Behavioural Neurology*, *25*(1), 53–60.

Montagne, B., Kessels, R.P., De Haan, E.H., & Perrett, D.I. (2007). The Emotion Recognition Task: A paradigm to measure the perception of facial emotional expressions at different intensities. *Perceptual Motor Skills*, *104*(2), 589–598.

Naranjo, C., Kornreich, C., Campanella, S., Noel, X., Vandriette, Y., Gillain, B., … Constant, E. (2011). Major depression is associated with impaired processing of emotion in music as well as in facial and vocal stimuli. *Journal of Affective Disorders*, *128*(3), 243–251.

Piguet, O., Hornberger, M., Mioshi, E., & Hodges, J.R. (2011). Behavioural-variant frontotemporal dementia: Diagnosis, clinical staging, and management. *The Lancet Neurology*, *10*(2), 162–172.

Rapcsak, S.Z., Galper, S.R., Comer, J.F., Reminger, S.L., Nielsen, L., Kaszniak, A.W., … Cohen, R.A. (2000). Fear recognition deficits after focal brain damage: A cautionary note. *Neurology*, *54*(3), 575–581.

Sheikh, J.I., & Yesavage, J.A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist*, *5*(1-2), 165–173.

Suzuki, A., Hoshino, T., & Shigemasu, K. (2006). Measuring individual differences in sensitivities to basic emotions in faces. *Cognition*, *99*(3), 327–353.

Suzuki, A., Hoshino, T., Shigemasu, K., & Kawamura, M. (2007). Decline or improvement? Age-related differences in facial expression recognition. *Biological Psychology*, *74*(1), 75–84.

Thalmann, B., Spiegel, R., Stähelin, H.B., Brubacher, D., Ermini-Fünfschilling, D., Bläsi, S., & Monsch, A.U. (2002). Dementia screening in general practice: Optimised scoring for the Clock Drawing Test. *Brain Aging*, *2*(2), 36–43.

Wilhelm, O., Hildebrandt, A., Manske, K., Schacht, A., & Sommer, W. (2014). Test battery for measuring the perception and recognition of facial expressions of emotion. *Frontiers in Psychology*, *5*, 404.

Young, A., Perrett, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). *Facial Expressions of Emotion – Stimuli and Tests (FEEST)*. Bury St Edmunds: Thames Valley Test Company.