Title
Cross-Cultural Analysis of Infant Facial Expression: Automated Emotion Detection

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Zhang, Su
Smith, Marchella
Lee, Kean Mun
et al.

Publication Date
2026

Cross-Cultural Analysis of Infant Facial Expression: Automated Emotion Detection

Su Zhang[1], Marchella Smith[1], Kean Mun Lee[1], Daniel Fatori[2], Guilherme V Polanczyk[2], Evelyn C Law[3], Victoria Leong (victorialeong@ntu.edu.sg)[1] & Cuntai Guan (CTGuan@ntu.edu.sg)[1]
[1]Nanyang Technological University, Singapore
[2]University of São Paulo, Brazil
[3]National University of Singapore, Singapore

Abstract

The current study aims to investigate cross-cultural and gender-related differences in infants' facial expression using a novel, infant emotion detection model – the Hierarchical Multimodal Attention Neural (HMAN) Network – for automated Action Unit (AU) prediction. Analyses investigate how facial AUs (1) are *distributed* across positive and negative emotion classes, (2) *co-occur* within positive and negative emotion classes, and how (3) their *temporal dynamics* unfold over time. Compared to Singaporean infants (n=48), Brazilian infants (n=59) were more likely to lower brows (AU04) during negative emotion, and took longer to fully develop a facial expression. Male infants across countries were more likely than female infants to pull lip corners (AU12), and less likely to lower brows (AU04) during positive emotion. These findings advance theoretical understanding of early emotional development and highlight the need for culturally and contextually sensitive tools for interpreting infant emotion, particularly for the design of diagnostic and intervention tools.

Keywords: Infant Emotion; Facial Action Units; Cross-Cultural Analyses

Introduction

Emotions are an important tool for communicating information between individuals (Keltner & Haidt, 2010). The ability to understand infant emotions has significant adaptive value because infants cannot verbally report their emotions. The Differential Emotions Theory (DET; Izard, 1971, 1977, 1991) proposes that humans have a biologically pre-wired set of discreet basic emotions (happiness, sadness, fear, anger, surprise, disgust) that have unique adaptive functions. In adults, each basic emotion is proposed to be associated with a distinct neural and physiological response, and a specific configuration of facial movements – i.e. facial expression – that is *universally* perceived as expressing that emotion (e.g., happiness is associated with smiling). Evidence for the *universality thesis* (Nelson & Russell, 2013) came from early reports that members of small-scale societies (primarily The Fore people of Papua New Guinea) who were pre-literate and had limited exposure to (and hence opportunity to learn about) cultural practices and norms (e.g., media), were able to match posed configurations of facial movements (e.g., fear) with correct researcher provided emotional scenario descriptions (e.g., "He is frightened because he is about to be attacked") (Ekman et al., 1969; Ekman & Friesen, 1971a,b).

Importantly, DET proposes that while emotions are innate, their use, regulation, and interpretation (i.e. integration with cognition) are shaped by environmental influence. As such, human emotions develop from being relatively undifferentiated (e.g., positive or negative in valence) during infancy, to acquiring discrete categories (happiness, sadness, fear, anger, surprise, disgust) by later childhood. This developmental progression requires experience of the social world and is accompanied by maturation of executive control abilities (Cuevas & Bell, 2014; Garon et al., 2008 ; Bennett et al., 2002; Camras, 1991; Camras & Fatani, 2008; Leppänen & Nelson, 2009; Lewis, 2000; Lewis et al., 1990; Widen & Russell, 2003, 2008). Specifically, at 0-2 months, infants display only basic reflexive emotions (e.g., distress, contentment) that are critical for survival (e.g., signalling discomfort, hunger, etc.) (Izard, 1994). From 2-6 months, they develop social emotions (e.g., smiling/joy in response to social interactions) and increasing differentiation of negative emotions (e.g., disgust, anger, sadness, fear) (Mitsven et al., 2020). By 12-24 months, they begin to display self-conscious emotions (e.g., embarrassment, pride, shame), coinciding with the development of self-recognition (Bennett et al., 2005; Camras, 2019; Camras & Fatani, 2008; Izard, 1991).

Social Learning in Emotion Development

Because infant emotion is undifferentiated and shaped by the environment, caregivers are critical in guiding how infants express and regulate emotions. *Social Learning Theory* (Bandura, 1971) posits that infants learn emotional expressions by observing caregivers and associating specific expressions with social contexts (Campos & Stenberg, 1981). Through imitation and reinforcement learning – where behaviours are encouraged or discouraged based on outcomes (e.g., reward or lack thereof) (Bandura, 1971) – infants begin to refine their discrete emotional repertoires (Meltzoff & Moore, 1977; Anisfeld, 1991; Jones, 2009; Ray

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

& Heyes, 2011; Geangu et al., 2010; 2016; Isomura & Nakano, 2016; de Klerk et al., 2018; Wang & Hamilton, 2012).

Cultural and Gender Influences in Families

The expression of emotions is also influenced by culture-specific emotion display rules (Ekman, 2003). Individualistic cultures, commonly associated with Western nations, value individual autonomy, personal goals, and the expression of personal opinion. Consequently, assertiveness and high emotional expressivity are standard social behaviours (Matsumoto et al., 2008a). Collectivist cultures, commonly associated with East Asian nations, emphasise group goals, duties and obedience. Hence, emotion suppression is valued, which helps to maintain social harmony (Leong, 2022; Matsumoto et al., 2008b). Empirical evidence from cross-cultural investigations of emotional expressivity and a recent systematic review (Weiss et al., 2022), suggest that individuals of Asian ethnicity are more likely to rate the display of emotions in various social situations as less appropriate (Matsumoto, 1993) and to suppress both positive and negative emotions compared to individuals of non-Asian ethnicity (e.g. White or Hispanic) (Gross & John, 2003). Latin American cultures, such as Brazil, are often categorised as collectivist, placing value in family and community belonging. Despite these collectivist values, they often display heightened emotion expressivity compared to East Asian, and Western cultures (Matsumoto et al., 2008a), as a way of maintaining interpersonal connections (Halgunseth et al., 2006).

These cultural values and emotional display rules inevitably penetrate family life and parent-child relationships. For example, Asian parents place more emphasis on achievement (Okagaki & Frensch, 1998) and respect for elders (Chao, 1994) in early childhood and do not value high emotional expressiveness in the family as much as European-immigrant mothers (Chen et al., 2014). Additionally, Doan and Wang (2010) showed that during a story-telling task European American mothers were significantly more likely to engage their three-year-old children in conversations about the character's mental states, including thoughts and emotions, compared with Chinese immigrant mothers, who placed more emphasis on character's behaviour. Relatedly, the parenting practices and styles commonly assumed to be optimal for children's development might have divergent effects depending on cultural context. For example, a study of East Asian immigrant mothers and their pre-school-aged children conducted in Canada found a negative relationship between maternal sensitivity and a child's behavioural problems (Chan et al., 2010). Maternal warmth, too, was found to be expressed differently by US Chinese immigrant mothers and by European American mothers (Cheah et al., 2015).

Just as cultural norms shape emotional display rules, gender-specific socialisation further influences emotional expression. For example, in an adult sample, McDuff et al. (2017) found that in people of Western cultures, smiling, characterised by contraction of the zygomatic major muscles that pull lip corners toward the ears (AU12), was displayed more frequently in females. Conversely, lip corner depression, defined as contraction of the depressor anguli oris muscles that pull lip corners down (AU15), was displayed for a shorter duration in females. Lastly, brow furrowing, defined as contraction of the corrugator muscles that pull eyebrows down and together (AU4), which is a prototypic expression of anger (or increased cognition and confusion, McDuff et al., 2017), was displayed more frequently in males. Gender differences may reflect socialisation pressures for females to display more positive emotion and less externalising negative emotion (Brody & Hall, 2008; Chaplin, 2015).

If facial movement configurations are categorised based on learned culture- and gender-specific emotion concepts (Weiss et al., 2022), this would predict that the gender- and culture-specific emotion expression accents displayed by adults should emerge at a particular time point in development and strengthen across ontogeny. Evidence of the development of emotion expression accents in infancy and childhood are limited, however, research on emotion recognition and face recognition more generally suggests that cultural differences begin to emerge in infancy/childhood (Ip et al., 2021; Kelly et al., 2011; Liu et al., 2011; see also Yuki et al. 2007).

Manual Coding of Infant Facial Expression

The Facial Action Coding System (FACS; Ekman & Friesen, 1971) and the infant version – Baby FACS (Oster, 2007) – are the most comprehensive and widely established methods of annotating facial actions. FACS specifies actions of individual facial muscles (action units; AUs) that either individually or when grouped, describe specific emotional expressions (e.g., happy, sad). Coding more AUs increases specificity. For example, in adults, a voluntary "Pan-Am" smile involves only the zygomatic major (mouth; AU12) muscles, whereas a sincere and involuntary "Duchenne" smile additionally involves the inferior part of the orbicularis oculi (eye; AU6) muscles (Del Giudice & Colle, 2007; Ekman et al., 1990).

However, since there are 33 core AUs (related to facial muscle movements), manual video coding of video recordings is highly cost- and time- intensive in terms of recruiting video coders to train and complete coding. As such, while FACS can be highly informative about participants' emotions, it is not as widely adopted as its potential would suggest, particularly for large datasets (Bartlett et al., 2006; Cohn & Ekman, 2008; Sariyanidi et al., 2015). Without FACS, one way that researchers have reduced the labour intensity is by providing a subjective rating of participants' holistic facial expressions (e.g., positive/negative/neutral) at an agreed rate (e.g., per second). However, this method is conflated by coder subjectivity. FACS allows for objectivity in that it distinguishes between face actions, and inferences about what they mean are extrinsic to the coding system, thus increasing reliability (Cohn et al., 2007; Oster et al., 1992)

Therefore, automated detection of AUs using computer vision and machine learning methods can be highly cost- and time-efficient (providing almost automatic ratings), allowing it to be scaled to larger numbers of participants in a way that is not possible with manual coding. Recently, automated detection of emotion has been used to provide fine-grained understanding of how infants respond to their environment (e.g., Hammal et al., 2017; Zaharieva et al., 2024) and social world (e.g., caregiver; Messinger et al., 2009). If reliable, this has the potential to answer questions about typical and atypical emotion dynamics. For example, social communication metrics such as emotion variability, interpersonal synchrony, and turn-taking dynamics, may indicate biological (e.g., infant and/or parent with neurodevelopmental condition such as Autism Spectrum Disorder, Attention Deficit (hyperactivity) Disorder, etc.) and/or environmental (e.g., parent mental health difficulties, neglect, socioeconomic status) factors associated with developmental risk (Leong & Schilbach, 2019). As such, the scalable assessment of child development using automated detection of emotion (and other sensory) indices can inform social and/or clinical intervention (Leong, 2022).

The current study will investigate cross-cultural and gender-related differences in infant facial expressions using a novel, automated infant emotion detection model – the Hierarchical Multimodal Attention Neural (HMAN) Network – for automated Action Unit (AU) prediction. Analyses will investigate how AUs (1) are *distributed* across positive and negative emotion classes, (2) *co-occur* within positive and negative emotion classes, and how (3) their *temporal dynamics* unfold over time.

Method

Participants

59 Brazilian and 48 Singaporean mother-infant dyads were recruited for the Wellcome LEAP 1kD project. Data collection involving Brazilian dyads took place at University of São Paulo, Brazil, and caregivers spoke Brazilian Portuguese. Data collection involving Singaporean dyads took place at Nanyang Technological University, Singapore, and caregivers spoke English. Brazilian infants (29 female; 30 male) age ranged 7.2-16.2 months (M = 12.1, SD = 2.3); Singaporean infants (31 female; 17 male) age ranged 11.2 to 25.7 months (M = 16.7, SD = 5.0). Mothers had no history of severe mental health disorders, no infant delivery complications requiring medical intervention, and any auditory/visual impairment was corrected. Infants had no known genetic disorder or auditory/visual impairment.

Task Protocol

Mothers and infants completed the A-not-B task (Piaget, 1957), seated across from each other at a table. On each trial, while the infant watched, the mother hid a toy in one of two identical tubs, covered both simultaneously, pushed them toward the infant, and asked "where is the toy?". If the infant identified the correct toy location (reaching toward, or removing the lid, of their intended tub), the mother celebrated with them ("yay, you found it!"), and allowed them to play with the toy for a moment. Incorrect responses were corrected by the mother on the first trial ("no, the toy is not there, it is here, look", followed by *showing* the infant the correct location of the toy), and only verbally acknowledged on subsequent trials ("no, the toy is not there"). Trials were There were up to 12 trials, with the task ending early if infants became upset or unresponsive. A 4K resolution camcorder (Sony FDR-AX700 Camcorder) was used to videorecord the infants frontal view at 25 frames per second.

Data Analysis Method

Manual Coding Infants' emotions were manually annotated by trained video coders as either positive (smiling, cooing, etc), negative (anger, sadness, crying, fussing noises, etc.), or neutral valence. There was no fixed number or duration of events per infant. Events were split: Subset U comprises short events (0.5-10 sec), used for IER model training due to their consistency; Subset V comprises long events (>10 sec) used for cross-cultural analysis due to their rich temporal information (see Table 1).

Table 1. Number of total, positive, neutral, and negative emotion frames, and Mean event duration (MD. (sec)).

	Total	Positive	Neutral	Negative	MD. (sec)
Subset U	485	144	148	194	3.1±2.5
Subset V	1049	6y	614	368	31.3±44.7

Machine Learning Pipeline First, YOLOv7 (Wang et al., 2023), which was pretrained using multiple databases (Lucey et al., 2010, 2011; Mavadati et al., 2013; McKeown et al., 2012; Savran et al., 2008; Zhang et al., 2014), was used to track and obtain facial still images from videos. To further fine-tune YOLOv7 model, 50 video frames for each infant from the current sample were manually labelled with bounding boxes. Second, OpenFace (Baltrusaitis et al., 2016) was applied to align the detected faces within a consistent coordinate system, and Non-facial background regions were masked out. As a result, 74.3% of the video frames from the entire database were successfully extracted, totalling ~1.5 million cropped and aligned infant facial images. Additionally, intensity values for 17 Action Units (AUs) were extracted (Ekman & Friesen, 2019). Log-mel spectrograms from the audio of each labelled segment were also used as input to a pretrained VGG16 network. Finally, the cleaned faces are sampled (sliding window size 12), together with the audio log-mel, to form a sample for HMAN.

The Hierarchical Multimodal Attention Neural (HMAN) Network Model First, a multimodal feature extractor uses pretrained ResNet50 and VGG networks to extract visual features from the samples of facial images and audio features from log-mel spectrograms, which are concatenated into a joint representation. Second, a cross-modal attention module employs linear layers and softmax activation to compute attention scores that weight the importance of each modality, fusing them into a unified feature vector. Third, a temporal attention module identifies the most informative temporal segments across consecutive frames, using similar attention mechanisms to weight temporal importance before making final emotion predictions. Sequences of 12 consecutive data points with a sliding window approach to capture temporal dynamics and resolve ambiguities in static observations. To make full use of the data, the HMAN was trained with leave-one-out protocol. It achieved 70.7% on accuracy and 72.6% f1-score on Subset U.

Data Preparation and Analysis The trained HMAN model was used to re-label human-identified positive/negative emotions (Subset V) with finer temporal resolution (for every 12 frames, or 0.48s). The subsequent cross-cultural analysis (below points 1, 2, and 3) focused on seven representative AUs, which have been shown to map universally to basic emotions (Du et al., 2014). They include (see Fig. 1), AU04 (Brow Lowerer), AU06 (Cheek Raiser), AU07 (Lid Tightener), AU12 (Lip Corner Puller), AU15 (Lip Corner Depressor), AU17 (Chin Raiser), AU26 (Jaw Drop).

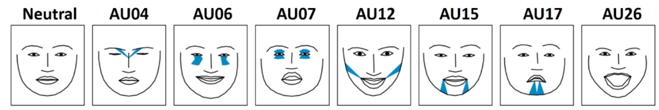

Figure 1. Muscle movements of the 7 representative AUs (visualised using PyFeat, Muhammod et al., 2019).

1. Cross-cultural AU Distribution by Emotion Class HMAN predictions were grouped according to country (Singaporean or Brazilian), age (< 1 year or 1-3 years), and sex (male or female). To analyse how much each AU contributed towards the positive and negative emotion classes, AU intensity values were converted into a binary occurrence format (i.e. > 1 = present, < 1 = absent). Time points with no active AUs were excluded. Within each group, AU occurrences were summed across frames and normalised by frame number.

2. Cross-cultural AU Concurrency by Emotion Class To understand which AUs co-occur within emotions, a 7 × 7 AU concurrency matrix was computed. For each frame, all pairs of simultaneously present AUs were counted, such that co-occurring AUs incremented the corresponding matrix cells. After aggregating across frames, the resulting matrix was normalized row-wise, so that the values represent the proportion of times a given AU co-occurred with other AUs.

3. Cross-cultural AU Temporal Dynamics Facial expressions were characterised by three stages: *onset* (the point at which intensity begins to increase), *apex* (the point of maximum intensity), *offset* (the point at which intensity begins to decrease). Using frame-level AU intensities from OpenFace, AU events were identified as a continuous sequence of frames in which the AU was active (i.e. non-zero). Each event starts at the first non-zero frame and ends when the intensity returns to 0. The maximum intensity value within this event is defined as the apex. Events with peak intensity <1 were excluded. For each AU, the mean onset to apex, and mean apex to offset durations were calculated. For each participant, AU duration, peak latency, and peak intensity, were averaged across episodes for each AU.

Results

1. Cross-cultural AU Distribution by Emotion Class Results are illustrated in Fig. 2. **AU04** (Brow Lowerer): higher in females than males during positive emotion; higher in Brazilian than Singaporean infants during negative emotion. **AU06** (Cheek Raiser): higher in negative than positive emotions; highest in Singaporean male infants than all other groups. **AU07** (Lid Tightener) and **AU17** (Chin

Raiser): were the most common and active AUs in all circumstances. **AU12** (Lip Corner Puller): higher in males than females in both countries groups during positive emotion; highest in Singaporean male infants. **AU15** (Lip Corner Depressor): equally active across corresponding groups for both emotions. **AU26** (Jaw Drop): highest during positive emotions, especially for Singaporean infants.

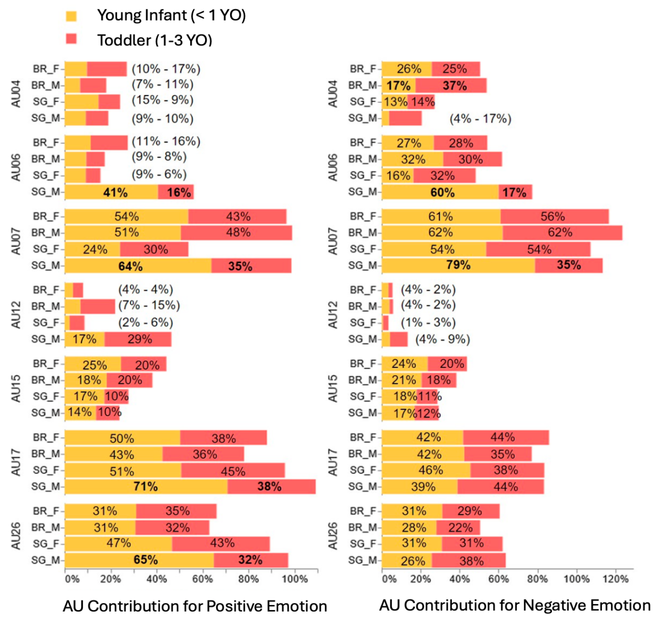

Figure 2. AUs contributions to positive and negative emotions according to age, sex (F, M), and country (SG, BR).

2. Cross-cultural AU Concurrency by Emotion Class Concurrency matrices are visualised in Fig. 3. Analysis shows universality across each emotion category: During AU04 display, both Singaporean and Brazilian infants concurrently display AU07, AU17, and AU26. During AU12 display, Singaporean and Brazilian infants concurrently display AU06 and AU07. During negative emotions, AU12 is the least occurred AU. During positive emotions, AU06, most frequently co-occurred with AU12 and AU07 for Singaporean and Brazilian infants, respectively.

3. Cross-cultural AU Temporal Dynamics To analyse the effect of Country and AU on emotion (i) duration, (ii) peak latency, and (iii) peak intensity three separate 2 (Country: Singapore/Brazil) x 7 (AU: AU04/AU06/AU07/AU12/AU15/AU17/AU26) mixed ANOVAs were conducted with repeated measures on the second factor. See Table 2. Mauchly's test of sphericity was violated for all analyses; Greenhouse-Geisser corrections were applied: **Duration:** Main effects of Country, $F(1, 97) = 54.10$, $p < .001$, and AU, $F(2.79, 270.50) = 36.07$, $p < .001$, were subsumed by a Country x AU interaction, $F(2.79, 270.50) = 9.98$, $p < .001$. Post-hoc tests showed longer AU duration in Brazilian than Singaporean infants for AU04 (M diff = 1.44), AU06 (M diff = 1.60), AU07 (M diff = 1.36), and AU15 (M diff = 0.16) (all $p < .05$). **Peak Latency:** Main effects of Country, $F(1, 97) = 54.86$, $p < .001$, and AU, $F(2.68, 259.55) = 30.52$, $p < .001$, were subsumed by a Country x AU interaction, $F(2.68, 259.55) = 8.71$, $p < .001$. Post-hoc tests showed that Brazilian infants showed longer peak latencies than Singaporean infants for AU04 (M diff = 0.67), AU06 (M diff = 0.88), and AU07 (M diff = 0.62), (all $p < .01$). **Peak Intensity:** Main effects of Country, $F(1, 97) = 7.13$, $p = .009$, and AU, $F(3.427, 332.467) = 386.332$, $p < .001$, were subsumed by a Country x AU interaction, $F(3.427, 332.467) = 10.97$, $p < .001$. Post-hoc tests showed lower AU15 (M diff = 0.44) peak intensity in Singaporean than Brazilian infants ($p < .001$).

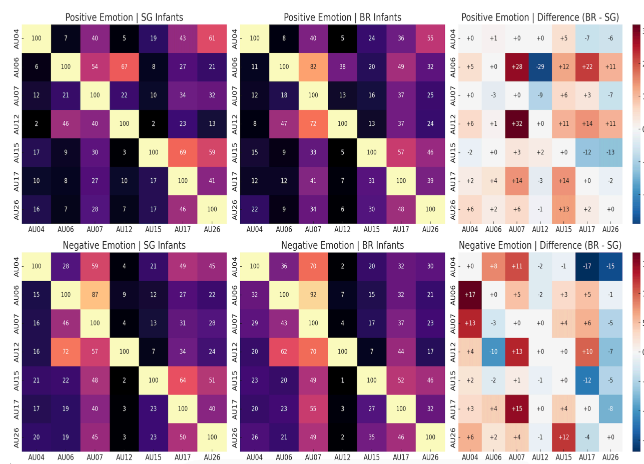

Figure 3. Left (SG Infants) and middle (BR Infants) panels illustrate the concurrency of AUs towards the positive and negative emotions. Lighter colours/higher percentages indicate higher co-occurrence. Right panels display the difference in percentage co-occurrence between groups (Brazilian - Singaporean); positive values (red) reflect higher co-occurrence in Brazilian infants; negative values (blue) reflect higher co-occurrence in Singaporean infants.

Table 2. Temporal dynamics of AUs. Mean duration (seconds) of full AU movement, apex, offset, and peak.

	Singaporean Infants				Brazilian Infants			
AU	Full (s)	Apex (s)	Offset (s)	Peak	Full (s)	Apex (s)	Offset (s)	Peak
AU04	1.4	0.7	0.7	1.6	2.8	1.4	1.4	1.6
AU06	2.2	1.0	1.2	1.4	3.8	1.9	1.9	1.4
AU07	1.6	0.8	0.9	1.9	2.9	1.4	1.5	1.9
AU12	1.6	0.8	0.8	1.4	2.0	1.1	0.9	1.3
AU15	1.0	0.5	0.6	2.5	1.2	0.5	0.7	2.9
AU17	2.1	1.0	1.1	2.2	2.5	1.3	1.3	2.3
AU26	1.7	0.8	0.9	2.2	2.0	1.0	1.1	2.2
Mean	1.7	0.8	0.9	1.9	2.4	1.2	1.2	2.0

Discussion

The current study investigated cross-cultural and gender differences in infant facial expression, using a novel infant emotion-detection model – Hierarchical Multimodal Attention Neural (HMAN) Network – for automated AU prediction.

Cross-Cultural Differences Brazilian infants were more likely than Singaporean infants to lower brows (AU04) during negative emotion. This aligns with theories on cultural differences in emotion expressivity: compared to Western contexts, East Asian collectivist cultures value emotion suppression to maintain social harmony (Leong, 2022; Matsumoto et al., 2008a, b; Matsumoto, 1993; Weiss et al., 2022). Although Latin American cultures are also collectivist, they emphasise emotional expressivity more than East Asian contexts (Chaplin, 2015; Halgunseth et al., 2006; Matsumoto et al., 2008a). Therefore, the current findings may reflect early-emerging cultural variation.

Compared to Singaporean infants, Brazilian infants took longer to fully develop a facial expression. This might appear to contrast the findings from older children and adults (Ip et al., 2021; Kelly et al., 2011; Liu et al., 2011), which predict that Singaporean infants would take longer to fully develop a facial expression, reflecting greater inhibition. Alternatively, the results may reflect cross-cultural differences in the expression of low arousal emotion (e.g., mild discomfort). If milder emotion is easier to inhibit, this may produce faster observable expression despite comparable underlying emotion experience. Further research should test this using physiological measures of arousal to distinguish between the duration of *experienced* (i.e. physiological arousal) and *expressed* emotion (facial AUs).

Gender Differences Male infants across countries were more likely than female infants to pull lip corners (AU12), and less likely to lower brows (AU04) when expressing positive emotion. This contradicts findings from adult samples, which suggest that females display more positive emotion (e.g., McDuff et al., 2017). The current results may reflect innate gender differences *specific to infant* temperament (e.g., Gagne et al., 2013). Alternatively, the observed gender differences may reflect gender-specific *caregiving practices*, with more verbal engagement typically observed with daughters (e.g., dyadic conversation and singing) despite no gender difference in infant vocalization, and more physical play with sons (e.g., bouncing, lifting) (Clearfield & Nelson, 2006; Fausto-Sterling et al., 2015; Johnson et al., 2014; Mascaro et al., 2017), These *socialisation* practices may reinforce gender-specific emotion tendencies. However, since the current study does not directly measure caregiver behaviour, it cannot determine the extent to which the observed differences reflect biological or socialisation explanations. Thus, while adult-typical emotion expression gender differences are not clearly evident in the present data, further longitudinal work may investigate how such patterns emerge over development.

Limitations and Real-World Applications Several limitations should be considered. First, the older age of Singaporean than Brazilian infants limits the ability to disentangle cultural from developmental effects. Additionally, the use of a structured A-not-B task may mean that the findings are context-specific and not generalizable to more naturalistic settings. Finally, the ML model generates fine-grained AU predictions, despite coarse valence-based training labels (positive/negative/neutral). Therefore, future work is needed to further validate these findings.

If gender and cross-cultural differences exist in infants' emotion expression, this may influence *how* people interpret and respond to their emotion. For example, Singaporean infants less brow lowering, may be perceived as more neutral by observers, despite potentially similar arousal as Brazilian infants. Similarly, male infants' greater lip corner activation may make their emotion more detectable than females'. Caregivers' differing responses according to infant gender are often attributed to caregivers' gender stereotypes, but may also partly reflect variation in infants' observable emotion signalling. Approaches like HMAN may help to quantify whether the magnitude of differences in facial expression correlates with the magnitude of difference in judgement of, or response to, emotion (McDuff et al., 2017).

ML for automated AU detection can be highly cost- and time-efficient, enabling scalability beyond what is possible with manual coding. Application of ML-based infant emotion recognition offers clinical potential. The findings suggest that identifying which facial regions are most diagnostic of neurodevelopmental (e.g., Attention Deficit (hyperactivity) Disorder, Autism Spectrum Disorder), or environmental differences (e.g., parent mental health) *within specific cultural and gender contexts*, may improve diagnostic approaches.

Acknowledgments

This research is supported by the RIE2025 Human Potential Programme Prenatal/Early Childhood Grants (H22P0M0002 & H24P2M0008), administered by A*STAR and by the Ministry of Education, Singapore, under its Academic Research Fund Tier 2 (MOE-T2EP40121-0001) and Social

Science & Humanities Research Fellowship to VL (MOE2020-SSHR-008).

References

Anisfeld, M. (1991). Neonatal imitation. In *Developmental Review* (Vol. 11, Issue 1). https://doi.org/10.1016/0273-2297(91)90003-7

Baltrusaitis, T., Robinson, P., & Morency, L. P. (2016). OpenFace: An open source facial behavior analysis toolkit. *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*. https://doi.org/10.1109/WACV.2016.7477553

Bandura, A. (1971). Social Learning Theory. *New York: General Learning Press.*

Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2006). Fully automatic facial action recognition in spontaneous behavior. *FGR 2006: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, 2006*. https://doi.org/10.1109/FGR.2006.55

Bennett, D. S., Bendersky, M., & Lewis, M. (2002). Facial expressivity at 4 months: A context by expression analysis. *Infancy*, *3*(1). https://doi.org/10.1207/S15327078IN0301_5

Bennett, D. S., Bendersky, M., & Lewis, M. (2005). Does the organization of emotional expression change over time? Facial expressivity from 4 to 12 months. In *Infancy* (Vol. 8, Issue 2). https://doi.org/10.1207/s15327078in0802_4

Brody, L. R., & Hall, J. A. (2008). Gender and emotion in context. *Handbook of Emotions*.

Campos, J. J., Stenberg, C. (1981). Perception, appraisal, and emotion: The onset of social referencing. In *Infant social cognition: Empirical and theoretical considerations*.

Camras, L. A. (1991). Conceptualizing early infant affect: View II & Reply. In *In K. Strongman (Ed.), International review of studies on emotion* (pp. 16–28). Wiley.

Camras, L. A. (2019). Facial Expressions Across the Life Span. In *Handbook of Emotional Development*. https://doi.org/10.1007/978-3-030-17332-6_5

Camras, L. A., & Fatani, S. S. (2008). The Development of Facial Expressions: Current Perspectives on Infant Emotions. In *Handbook of Emotions*.

Chan, K., Penner, K., Janet, W. T., & Charlotte, J. (2010). Assessing Parenting Behaviors in Euro–Canadian and East Asian Immigrant Mothers: Limitations to Observations of Responsiveness. *Http://Dx.Doi.Org/10.1080/07317101003776423*, *32*(2), 85–102. https://doi.org/10.1080/07317101003776423

Chao, R. K. (1994). Beyond Parental Control and Authoritarian Parenting Style: Understanding Chinese Parenting Through the Cultural Notion of Training. *Child Development*, *65*(4), 1111. https://doi.org/10.2307/1131308

Chaplin, T. M. (2015). Gender and emotion expression: A developmental contextual perspective. *Emotion Review*, *7*(1). https://doi.org/10.1177/1754073914544408

Cheah, C. S. L., Li, J., Zhou, N., Yamamoto, Y., & Leung, C. Y. Y. (2015). Understanding Chinese immigrant and European American Mothers' expressions of warmth. *Developmental Psychology*, *51*(12), 1802–1811. https://doi.org/10.1037/A0039855

Chen, S. H., Hua, M., Zhou, Q., Tao, A., Lee, E. H., Ly, J., & Main, A. (2014). Parent-child cultural orientations and child adjustment in Chinese American immigrant families. *Developmental Psychology*, *50*(1), 189–201. https://doi.org/10.1037/A0032473

Clearfield, M. W., & Nelson, N. M. (2006). Sex differences in mothers' speech and play behavior with 6-, 9-, and 14-month-old infants. *Sex Roles*, *54*(1–2). https://doi.org/10.1007/s11199-005-8874-1

Cohn, J., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. [References]. *Coan, James A [Ed], May 2014*.

Cohn, J. F., & Ekman, P. (2008). Measuring Facial Action. In *The New Handbook of Methods in Nonverbal Behavior Research*. https://doi.org/10.1093/acprof:oso/9780198529620.003.0002

Cuevas, K., & Bell, M. A. (2014). Infant Attention and Early Childhood Executive Function. *Child Development*, *85*(2). https://doi.org/10.1111/cdev.12126

de Klerk, C. C. J. M., Hamilton, A. F. de C., & Southgate, V. (2018). Eye contact modulates facial mimicry in 4-month-old infants: An EMG and fNIRS study. *Cortex*, *106*. https://doi.org/10.1016/j.cortex.2018.05.002

Del Giudice, M., & Colle, L. (2007). Differences between children and adults in the recognition of enjoyment smiles. *Developmental Psychology*, *43*(3). https://doi.org/10.1037/0012-1649.43.3.796

Doan, S. N., & Wang, Q. (2010). Maternal Discussions of Mental States and Behaviors: Relations to Emotion Situation Knowledge in European American and Immigrant Chinese Children. *Child Development*, *81*(5), 1490. https://doi.org/10.1111/J.1467-8624.2010.01487.X

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(15). https://doi.org/10.1073/pnas.1322355111

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. *Nebraska Symposium on Motivation*, *19*.

Ekman, P. (2003). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life.* Times Books/Henry Holt and Co. https://psycnet.apa.org/record/2003-88051-000

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne Smile: Emotional Expression and Brain Physiology II. *Journal of Personality and Social Psychology*, *58*(2). https://doi.org/10.1037/0022-3514.58.2.342

Ekman, P., & Friesen, W. V. (1971a). Constants across cultures in the face and emotion. *Journal of Personality*

and Social Psychology, *17*(2). https://doi.org/10.1037/h0030377
Ekman, P., & Friesen, W. V. (1971b). Facial action coding system: A technique for the measurement of facial movement. *Journal of Personality and Social Psychology*, *17*(2).
Ekman, P., & Friesen, W. V. (2019). Facial Action Coding System. In *PsycTESTS Dataset*. https://doi.org/10.1037/t27734-000
Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*(3875). https://doi.org/10.1126/science.164.3875.86
Fausto-Sterling, A., Crews, D., Sung, J., García-Coll, C., & Seifer, R. (2015). Multimodal sex-related differences in infant and in infant-directed maternal behaviors during months three through twelve of development. *Developmental Psychology*, *51*(10). https://doi.org/10.1037/dev0000033
Gagne, J. R., Miller, M. M., & Goldsmith, H. H. (2013). Early-but modest-gender differences in focal aspects of childhood temperament. *Personality and Individual Differences*, *55*(2). https://doi.org/10.1016/j.paid.2013.02.006
Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive Function in Preschoolers: A Review Using an Integrative Framework. *Psychological Bulletin*, *134*(1). https://doi.org/10.1037/0033-2909.134.1.31
Geangu, E., Benga, O., Stahl, D., & Striano, T. (2010). Contagious crying beyond the first days of life. *Infant Behavior and Development*, *33*(3). https://doi.org/10.1016/j.infbeh.2010.03.004
Geangu, E., Quadrelli, E., Conte, S., Croci, E., & Turati, C. (2016). Three-year-olds' rapid facial electromyographic responses to emotional facial expressions and body postures. *Journal of Experimental Child Psychology*, *144*. https://doi.org/10.1016/j.jecp.2015.11.001
Gross, J. J., & John, O. P. (2003). Individual Differences in Two Emotion Regulation Processes: Implications for Affect, Relationships, and Well-Being. *Journal of Personality and Social Psychology*, *85*(2), 348–362. https://doi.org/10.1037/0022-3514.85.2.348
Halgunseth, L. C., Ispa, J. M., & Rudy, D. (2006). Parental control in latino families: An integrated review of the literature. *Child Development*, *77*(5). https://doi.org/10.1111/j.1467-8624.2006.00934.x
Hammal, Z., Chu, W. S., Cohn, J. F., Heike, C., & Speltz, M. L. (2017). Automatic action unit detection in infants using convolutional neural network. *2017 7th International Conference on Affective Computing and Intelligent Interaction, ACII 2017*, *2018-January*. https://doi.org/10.1109/ACII.2017.8273603
Ip, K. I., Miller, A. L., Karasawa, M., Hirabayashi, H., Kazama, M., Wang, L., Olson, S. L., Kessler, D., & Tardif, T. (2021). Emotion expression and regulation in three cultures: Chinese, Japanese, and American preschoolers' reactions to disappointment. *Journal of Experimental Child Psychology*, *201*. https://doi.org/10.1016/j.jecp.2020.104972
Isomura, T., & Nakano, T. (2016). Automatic facial mimicry in response to dynamic emotional stimuli in five-monthold infants. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1844). https://doi.org/10.1098/rspb.2016.1948
Izard, C. E. (1971). The face of emotion. In *New York: Appleton-Century-Crofts.*
Izard, C. E. (1977). The Face, The Fundamental Emotions, and Affect Interactions. In *Human Emotions*. https://doi.org/10.1007/978-1-4899-2209-0_4
Izard, C. E. (1991). The psychology of emotions. In *New York: Plenum Press.*
Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, *115*(2). https://doi.org/10.1037/0033-2909.115.2.288
Johnson, K., Caskey, M., Rand, K., Tucker, R., & Vohr, B. (2014). Gender differences in adult-infant communication in the first months of life. *Pediatrics*, *134*(6). https://doi.org/10.1542/peds.2013-4289
Jones, S. S. (2009). The development of imitation in infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528). https://doi.org/10.1098/rstb.2009.0045
Kelly, D. J., Liu, S., Rodger, H., Miellet, S., Ge, L., & Caldara, R. (2011). Developing cultural differences in face processing. *Developmental Science*, *14*(5). https://doi.org/10.1111/j.1467-7687.2011.01067.x
Keltner, D., & Haidt, J. (2010). Social Functions of Emotions at Four Levels of Analysis. *Http://Dx.Doi.Org/10.1080/026999399379168*, *13*(5), 505–521. https://doi.org/10.1080/026999399379168
Leong, V. (2022). Neural Sociometrics: Toward Early Screening of Infant Psychosocial and Brain Health to Improve Lifelong Mental Well-Being. *Policy Insights from the Behavioral and Brain Sciences*, *9*(1). https://doi.org/10.1177/23727322211068020
Leong, V., & Schilbach, L. (2019). The promise of two-person neuroscience for developmental psychiatry: Using interaction-based sociometrics to identify disorders of social interaction. In *British Journal of Psychiatry* (Vol. 215, Issue 5). https://doi.org/10.1192/bjp.2019.73
Leppänen, J. M., & Nelson, C. A. (2009). Tuning the developing brain to social signals of emotions. In *Nature Reviews Neuroscience* (Vol. 10, Issue 1). https://doi.org/10.1038/nrn2554
Lewis, M. (2000). The Emergence of Human Emotions. *Handbook of Emotions 2*, *01*(01).
Lewis, M., Alessandri, S. M., & Sullivan, M. W. (1990). Violation of Expectancy, Loss of Control, and Anger Expressions in Young Infants. *Developmental Psychology*, *26*(5). https://doi.org/10.1037/0012-1649.26.5.745
Liu, S., Quinn, P. C., Wheeler, A., Xiao, N., Ge, L., & Lee, K. (2011). Similarity and difference in the processing of same- and other-race faces as revealed by eye tracking in

4- to 9-month-olds. *Journal of Experimental Child Psychology*, *108*(1). https://doi.org/10.1016/j.jecp.2010.06.00

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010*. https://doi.org/10.1109/CVPRW.2010.5543262

Lucey, P., Cohn, J. F., Prkachin, K. M., Solomon, P. E., & Matthews, I. (2011). Painful data: The UNBC-McMaster shoulder pain expression archive database. *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011*. https://doi.org/10.1109/FG.2011.5771462

Mascaro, J. S., Rentscher, K. E., Hackett, P. D., Mehl, M. R., & Rilling, J. K. (2017). Child gender influences paternal behavior, language, and brain function. *Behavioral Neuroscience*, *131*(3). https://doi.org/10.1037/bne0000199

Matsumoto, D., Yoo, S. H., Fontaine, J., Anguas-Wong, A. M., Arriola, M., Ataca, B., Bond, M. H., Boratav, H. B., Breugelmans, S. M., Cabecinhas, R., Chae, J., Chin, W. H., Comunian, A. L., Degere, D. N., Djunaidi, A., Fok, H. K., Friedlmeier, W., Ghosh, A., Glamcevski, M., … Grossi, E. (2008a). Mapping Expressive Differences Around the World: The Relationship Between Emotional Display Rules and Individualism Versus Collectivism. *Http://Dx.Doi.Org/10.1177/0022022107311854*, *39*(1), 55–74. https://doi.org/10.1177/0022022107311854

Matsumoto, D., Yoo, S. H., Nakagawa, S., Alexandre, J., Altarriba, J., Anguas-Wong, A. M., Arriola, M., Bauer, L. M., Bond, M. H., Cabecinhas, R., Chae, J., Comunian, A. L., DeGere, D. N., de Melo Garcia Bley, L., Fok, H. K., Friedlmeier, W., Garcia, F. M., Ghosh, A., Granskaya, J. v., … Zengeya, A. (2008b). Culture, Emotion Regulation, and Adjustment. *Journal of Personality and Social Psychology*, *94*(6), 925–937. https://doi.org/10.1037/0022-3514.94.6.925

Matsumoto, D. (1993). Ethnic Differences in Affect Intensity, Emotion Judgments, Display Rule Attitudes, and Self-Reported Emotional Expression in an American Samplel. *I Motivation and Emotion*, *17*(2).

Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., & Cohn, J. F. (2013). DISFA: A spontaneous facial action intensity database. *IEEE Transactions on Affective Computing*, *4*(2). https://doi.org/10.1109/T-AFFC.2013.4

McDuff, D., Kodra, E., Kaliouby, R. el, & LaFrance, M. (2017). A large-scale analysis of sex differences in facial expressions. *PLOS ONE*, *12*(4), e0173942. https://doi.org/10.1371/journal.pone.0173942

McKeown, G., Valstar, M., Cowie, R., Pantic, M., & Schröder, M. (2012). The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing*, *3*(1). https://doi.org/10.1109/T-AFFC.2011.20

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of Facial and Manual Gestures by Human Neonates. *Science*, *198*(4312). https://doi.org/10.1126/science.198.4312.75

Messinger, D. S., Mahoor, M. H., Chow, S. M., & Cohn, J. F. (2009). Automated measurement of facial expression in infant - Mother interaction: A pilot study. *Infancy*, *14*(3). https://doi.org/10.1080/15250000902839963

Mitsven, S., Messinger, D. S., Moffitt, J., & Ahn, Y. A. (2020). Infant Emotional Development. In *The Cambridge Handbook of Infant Development* (pp. 742–776). Cambridge University Press. https://doi.org/10.1017/9781108351959.027

Nelson, N. L., & Russell, J. A. (2013). Universality revisited. In *Emotion Review* (Vol. 5, Issue 1). https://doi.org/10.1177/1754073912457227

Okagaki, L., & Frensch, P. A. (1998). Parenting and Children's School Achievement: A Multiethnic Perspective. *American Educational Research Journal*, *35*(1), 123. https://doi.org/10.2307/1163454

Oster, H. (2007). BabyFACS: Facial Action Coding System for infants and young children. In *Unpublished monograph and coding manual, New York University*.

Oster, H., Hegley, D., & Nagel, L. (1992). Adult Judgments and Fine-Grained Analysis of Infant Facial Expressions: Testing the Validity of A Priori Coding Formulas. *Developmental Psychology*, *28*(6). https://doi.org/10.1037/0012-1649.28.6.1115

Piaget, J. (1957). *Construction of reality in the child. New York: Basic.*

Ray, E., & Heyes, C. (2011). Imitation in infancy: The wealth of the stimulus. *Developmental Science*, *14*(1). https://doi.org/10.1111/j.1467-7687.2010.00961.x

Sariyanidi, E., Gunes, H., & Cavallaro, A. (2015). Automatic analysis of facial affect: A survey of registration, representation, and recognition. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Vol. 37, Issue 6). https://doi.org/10.1109/TPAMI.2014.2366127

Savran, A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B., & Akarun, L. (2008). Bosphorus database for 3D face analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5372 LNCS*. https://doi.org/10.1007/978-3-540-89991-4_6

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2023). *YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors*. https://doi.org/10.1109/cvpr52729.2023.00721

Wang, Y., & de Hamilton, A. F. C. (2012). Social top-down response modulation (STORM): A model of the control of mimicry in social interaction. In *Frontiers in Human Neuroscience* (Issue JUNE 2012). https://doi.org/10.3389/fnhum.2012.00153

Weiss, N. H., Thomas, E. D., Schick, M. R., Reyes, M. E., & Contractor, A. A. (2022). Racial and ethnic differences in emotion regulation: A systematic review. In *Journal of Clinical Psychology* (Vol. 78, Issue 5, pp. 785–808). John Wiley and Sons Inc. https://doi.org/10.1002/jclp.23284

Widen, S. C., & Russell, J. A. (2003). A Closer Look at Preschoolers' Freely Produced Labels for Facial Expressions. *Developmental Psychology*, *39*(1). https://doi.org/10.1037/0012-1649.39.1.114
Widen, S. C., & Russell, J. A. (2008). Children acquire emotion categories gradually. *Cognitive Development*, *23*(2). https://doi.org/10.1016/j.cogdev.2008.01.002
Yuki, M., Maddux, W. W., & Masuda, T. (2007). Are the windows to the soul the same in the East and West? *Journal of Experimental Social Psychology*, *43*(2).
Zaharieva, M. S., Salvadori, E. A., Messinger, D. S., Visser, I., & Colonnesi, C. (2024). Automated facial expression measurement in a longitudinal sample of 4- and 8-month-olds: Baby FaceReader 9 and manual coding of affective expressions. *Behavior Research Methods*, *56*(6). https://doi.org/10.3758/s13428-023-02301-3
Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., Liu, P., & Girard, J. M. (2014). BP4D-Spontaneous: A high-resolution spontaneous 3D dynamic facial expression database. *Image and Vision Computing*, *32*(10). https://doi.org/10.1016/j.imavis.2014.06.002