Division of Psychology and Language Sciences

University College London

HUMAN AND MACHINE RECOGNITION OF SPONTANEOUS AND DYNAMIC FACIAL EXPRESSIONS OF EMOTIONS

Hyunwoo Kim

Primary supervisor:

Dr. Eva G. Krumhuber

Secondary Supervisor:

Prof. Nadia Berthouze

London, May 2024

This dissertation is submitted for the degree of Doctor of Philosophy

"…Facial movement of expression impresses us
through its changes, through its melody.
The characteristic of the person will always be the way
they move, the melody of the expression; this
can never be caught in snapshots…"

(Sir Ernst Gombrich, cited by Miller, 1983)

Hyunwoo Kim

Name: Hyunwoo Kim
Dissertation title: Human and Machine Recognition of spontaneous and Dynamic Facial Expressions of emotion

ABSTRACT

Over the last few decades, much research on facial expression recognition has predominantly focused on posed, static facial images, often overlooking the importance of dynamic and spontaneous information. This dissertation addresses these gaps by exploring the roles of dynamic and spontaneous aspects in emotion recognition through comprehensive reviews and empirical studies of both humans and automated systems. In the first set of studies, various expression formats - dynamic, target, and non-target static - are analysed to determine the conditions under which dynamic information significantly enhances recognisability of expressions. Results revealed that dynamic cues play a compensatory role, particularly aiding recognition when static expressions fail to represent target emotions adequately. Subsequently, Chapter 2 reviews the existing databases of spontaneous and dynamic facial expressions, detailing their conceptual, technical, and practical features, thereby providing a comprehensive benchmark for research on encoding and decoding facial expressions. Employing automated facial expression analysis tools, Chapter 3 presents an empirical cross-corpus evaluation of the databases reviewed in Chapter 2. Findings showed that, although recognition rates for spontaneous databases generally remain low, they vary significantly across databases, highlighting the inherent difficulty and variability in recognising spontaneous expressions. Furthermore, this work elucidates the critical roles of featural parameters – prototypicality, ambiguity, and complexity – in accurate emotion recognition. In sum, the findings demonstrate that dynamic properties and spontaneous aspects convey important information that significantly influences the human and machine recognition of facial expressions.

Hyunwoo Kim

IMPACT STATEMENT

This dissertation, “Human and Machine Recognition of Spontaneous and Dynamic Facial Expressions of Emotion,” addresses significant gaps in the understanding of how dynamic and spontaneous aspects of facial expression influence emotion recognition. The comprehensive analysis conducted across various facial expression databases not only enriches the academic knowledge by providing a critical assessment of existing corpora but also establishes a new benchmark for future empirical studies in emotion recognition. This work is crucial for scholars seeking to employ or develop methodologies that capture diverse ways emotions are expressed and recognised by both humans and machines.

Outside academia, the applications of this research are extensive, particularly in technological and clinical settings. Advances in affective computing improve device interactions by systematically adapting to user emotions in real-time. This directly benefits mental health diagnostics by improving the accuracy of tools that assess and identify conditions such as social anxiety, where understanding subtle emotional cues is essential.

In the context of public safety, this research enhances facial recognition technology used in security and surveillance, potentially contributing the machine capacity to detect emotional expressions quickly. This capability is vital in high-security environments and public spaces, where accurately interpreting emotional cues can pre-empt potential threats and improve crisis management.

Furthermore, the insights from this dissertation can contribute to the development of educational technologies that utilise emotional data to adapt learning experiences, potentially improving student outcomes by responding to emotional cues that indicate confusion of negative emotions.

In summary, this research not only advances academic discourse in emotion recognition but also catalyses significant developments in technology, mental health, education, and security. By improving how machines understand human emotions, it paves the way for broader societal impacts, including enhanced non-verbal communication and more empathetic interactions in an increasingly connected world.

Publication details

Chapter 2: Published in *Frontiers in Psychology*

Kim, H., Küster, D., Girard, J. M., & Krumhuber, E. G. (2023). Human and machine recognition of dynamic and static facial expressions: Prototypicality, ambiguity, and complexity. *Frontiers in Psychology*, *14*, 1221081. https://doi.org/10.3389/fpsyg.2023.1221081

Chapter 3 (partially): Under review for *Affective Science*

Kim, H., Bian, Y., & Krumhuber, E. G. (under review). A Review of 25 Spontaneous and Dynamic Facial Expression Databases for Basic Emotion Research.

DECLARATIONS

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature: [redacted]

Name: Hyunwoo Kim

Date: 10. 05. 2024

ACKNOWLEDGEMENTS

Navigating through the challenging journey of a PhD is akin to sailing through stormy seas; both are relentless and require careful steering and resilience. I have been incredibly fortunate to have the support and guidance of a group of exceptional individuals who have enriched this journey immensely.

First and foremost, I would like to express my deepest gratitude to my principal supervisor, Dr. Eva G. Krumhuber, for her unwavering support and expert guidance throughout the course of this research. Her mentorship was both a beacon and a compass, illuminating the path forward and directing my progress with empathy and academic rigour. Eva’s enthusiasm and commitment to research excellence have profoundly influenced my own aspirations and achievements in academia.

I am also immensely thankful to my secondary supervisor, Prof. Nadia Berthouze, whose insights and feedback have been invaluable. Nadia’s constructive critiques and encouragement have significantly enhanced my work, pushing me to refine my ideas and broaden my perspectives.

Special appreciation goes to my thesis committee members Dr. Dennis Küster and Dr. Jeffrey M. Girard, whose expertise and thoughtful feedback have been crucial in shaping the final outcome of this dissertation. Their rigorous reviews and detailed suggestions have not only strengthened this work but have also contributed significantly to my personal and professional growth.

I am grateful to the members of our research group and the broader academic community for their insightful discussions and for providing a supportive and collaborative environment. These interactions have been instrumental in refining my research and have added depth and clarity to my work.

Hyunwoo Kim

My thanks also extend to the administrative and support staff within our department. Their assistance and dedication have made a significant difference in my day-to-day academic activities, from maintaining research equipment to organising training sessions that have enhanced my skills and knowledge.

I cannot forget all the participants, research assistants, and colleagues who played a role, big or small, in the various stages of this research. Your involvement was essential to the progress of this endeavour.

Last but certainly not least, I must express my profound gratitude to my family, friends and partner. Your endless patience, encouragement, and love provided the strength and motivation needed to pursue and complete this challenging academic journey. You all have my deepest appreciation and love for your support and for believing in me even during the most demanding times.

Thank you all for being part of this significant phase of my life and career.

Hyunwoo Kim May 11th 2024

CONTENTS

CHAPTER 3
A Systematic Review of Spontaneous and Dynamic Databases for Facial Expression Research **85**

CHAPTER 4
Empirical Evaluation of Spontaneous and Dynamic Facial Expression Databases for Basic Emotions **140**

CHAPTER 5
General Discussion **174**

REFERENCES

TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 1

Facial Expression and Emotion Recognition: An Introduction

1.1 Facial Expression Recognition

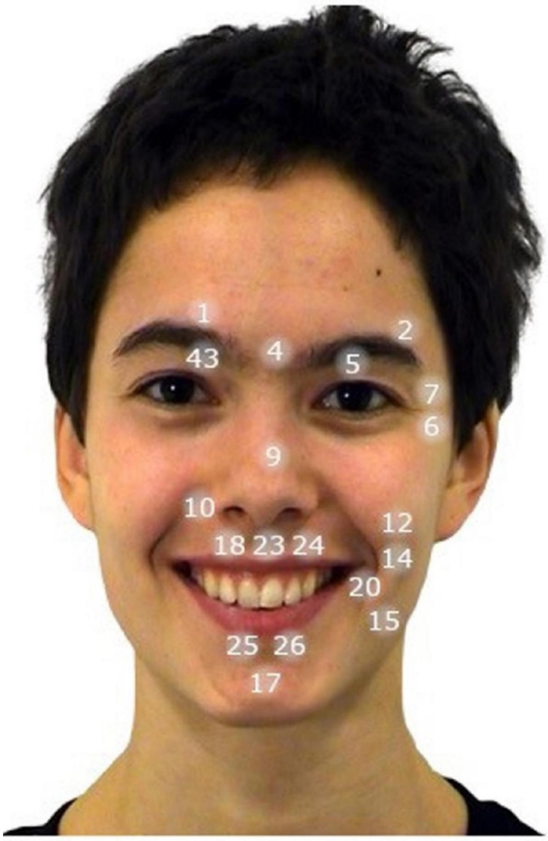
1.1.1 Emotions and Facial Expressions

Facial expressions (FEs) are not merely reflections of internal emotional states; they function as a primary medium of nonverbal communication, essential for conveying emotions and intentions in human interactions. The intricate relationship between emotions and FE has long been a focal point of psychological research, bridging various disciplines such as affective science, social psychology, developmental psychology, and even clinical research (Thomas et al., 2008). The importance of FEs is evident in their role in empathy, attachment, and mental health (Frith, 2009; McClure et al., 2000), with expressions offering a window into individuals' internal states and their social connections.

Historically, the prevailing assumption that facial expressions are reliable indicators of internal emotional states has been rooted in Charles Darwin's evolutionary perspective of emotion (Darwin, 1872). Darwin argued that facial expressions are adaptive, evolved functions that serve to communicate specific emotional states and intentions. For instance, an expression of anger might signal that the target of that expression is a potential threat, while and expression of fear could indicate imminent danger, thus triggering protective responses in observers (Erickson & Schulkin, 2003; Horstmann, 2003). This evolutionary view of expressions providing adaptive functions laid the foundation for much of the modern scientific understanding of facial expressions.

Building on Darwin's work, the Basic Emotion Theory (BET) posits that a set of core emotions – happiness, sadness, fear, anger, surprise, and disgust – are universally recognised and expressed through corresponding facial movements (Ekman, 1992; Izard, 1971). According to BET, these core emotions are biologically ingrained, each with distinct physiological patterns, thus forming a fundamental emotional lexicon that serves as the foundation of other emotions. BET has profoundly influenced facial expression research, particularly through its alignment with the Facial Action Coding System (FACS). FACS is an anatomically based system that delineates facial behaviour into discrete Action Units (AUs), each linked with specific muscle movement (Ekman & Friesen, 1978). These AUs operate in various combinations to express distinct emotions. For example, the expression of happiness is represented by a combination of AU12 (zygomatic major) and AU6 (orbicularis oculi), which together create what is commonly recognised as a smile (Figure 1.1). Recent studies have shown that even a single AU can be indicative of emotional meanings, highlighting the important role of AUs in facial displays (Namba et al., 2017).

FACS has contributed significantly to our understanding of facial expressions by offering a detailed, objective framework for examining the complex muscle movements involved in emotional expression. By breaking down facial expressions into individual AUs, researchers can analyse the subtle physical manifestation of emotion in a more structured way beyond limited number of emotion categories (Cohn et al., 2007). The level of detail in FACS has led to important findings in the field, such as the distinction between genuine and posed expressions (Namba et al., 2017; Valstar et al., 2006), the dynamics of spontaneous facial expressions (Bartlett et al., 2006; Park et al., 2020), and the exploration of cultural differences in cultural similarities and differences (Elfenbein & Ambady, 2002), subtle emotional nuances (Keltner et al., 2019), and facial indicators of depressive symptoms (Girard et al., 2013).



Facial Expression

AU01 Inner Brow Raiser
AU02 Outer Brow Raiser
AU04 Brow Lowerer
AU05 Upper Lid Raiser
AU06 Cheek Raiser
AU07 Lid Tightener
AU09 Nose Wrinkler
AU10 Upper Lid Raiser
AU12 Lip Corner Pull
AU14 Dimpler
AU15 Lip Corner Depressor
AU17 Chin Raiser
AU18 Lip Puckerer
AU20 Lip Stretcher
AU23 Lip Tightener
AU24 Lip Pressor
AU25 Lips Part
AU26 Jaw Drop
AU27 Mouth Stretch
AU43 Eyes Closed

Figure 1.1. Example FACS analysis (van der Schalk et al., 2011)

1.1.2 How Well Facial Expressions Are Recognised

While BET and FACS has significantly advanced our understanding of the facial expressions, the broader question of how accurately these expressions are recognised remains a critical area of study. Recognising facial expression is fundamental to human social interactions, as it serves a critical means through which individuals discern others’ emotional states. This capability is central to effective communication and social bonding (Cheung et al., 2015). The ability to accurately recognise facial expressions underpins various societal domains, from legal judgments and policy decisions to healthcare and education (Sun et al., 2016; Zloteanu et al., 2021). Moreover, the pivotal role of facial expression recognition in affective computing highlights its importance in enhancing human-computer interaction by enabling more intuitive and responsive technology (Valstar et al., 2015). Similarly, in the field of neuroscience, exploring how facial expressions are processed provides insights into neural substrates involved in emotion perception and social cognition (Morecraft et al., 2001).

From an information-processing perspective, facial expressions contain two major channels: the physical facial configurations and the affect they are presumed to convey (Calvo & Nummenmaa, 2016). Given that only the physical aspects of expressions are accessible to visual perception, the majority of prior studies on the recognition of facial expressions generally involve the visual processing of facial expressions (Calder & Young, 2005). Mostly, recognition tasks often operationalised expression identification through categorisation, matching observers' responses with predefined emotion categories of facial stimuli. This categorical view highly aligns with BET, arguing that viewers perceive discrete emotions from specific facial muscle changes, emphasising a direct link between morphological changes and internal emotional states (Ekman 1992).

Experiments assessing recognition performance across basic emotions report accuracy rates significantly above chance, generally exceeding 70% for most expressions (though posed; Haidt & Keltner, 1999). Such recognition performance is seemingly modulated by emotion category, with happiness often recognised with the greatest accuracy, likely due to its distinct facial cues that are more easily identifiable than other emotions, followed by surprise, anger, sadness and disgust, and fear (Calder et al., 2000; Calvo & Lundqvist, 2008). Such recognition patterns are consistent across different stimulus sets (Ekman & Friesen, 1978; Lundqvist et al., 1998; Tottenham et al., 2009). Interestingly, facial expressions are distinguishable under constrained visual conditions, including very brief exposures or when visual details are obscured. Such findings highlight the human capability to extract emotional information from facial expressions, even with limited visual input (Calvo & Lundqvist, 2008; Milders et al., 2008).

While earlier studies have demonstrated remarkable recognition performance from observers, a consistent pattern of confusion across various expressions sheds light on the complexities inherent in this process. Specifically, fear and surprise are often confused with each other, reflecting a substantial morphological and perceptual overlap in how these expressions are processed and interpreted (Palermo & Coltheart, 2004; Tottenham et al., 2009). Similarly, instances of mistaking disgust for anger or sadness – and vice versa – further illustrate the challenges in distinguishing between these emotions. The confusion rate for these misidentifications ranges from 10% to 42% (Palermo & Coltheart, 2004; Recio et al., 2013), suggesting a significant level of ambiguity in recognising certain expressions. These patterns of confusion highlight a critical aspect of facial expression recognition, revealing that the visual and affective cues theoretically distinguishing basic emotions are not always clear-cut to observers (Jack et al., 2014). This blurring of emotional boundaries is particularly prevalent in spontaneous expressions, suggesting a more intricate interplay between perceptual mechanisms and affective interpretation than previously understood.

In FACS AU recognition, the proficiency of trained coders is notable, with an accuracy level of around 85%, evidencing the effectiveness of human expertise in identifying specific facial actions (for posed expressions; Frank et al., 1993). These accuracies, however, are contingent on extensive training, posing a challenge for wider practical application. Surprisingly, untrained individuals also revealed a natural ability to recognise action unit patterns for posed micro-expressions, albeit with lower accuracy and consistency, typically between 60 to 70%, compared to their trained counterparts (Matsumoto & Hwang, 2011). The simplicity and exaggerated (although brief) appearance of posed expressions may aid the recognition. Despite the low effectiveness, the capability in untrained individuals suggests an inherent human skill in AU recognition, which can be refined through systematic training. It is important to note that most studies described above have employed posed, static expressions for testing recognition performance.

1.1.3 Existing Limitations of Basic Emotion Research

BET and FACS have made substantial contributions to the study of facial expressions and their recognition, yet they are not without limitations. One major critique of BET concerns its oversimplification of the emotional spectrum, particularly in its reliance on categorical distinctions between basic emotions. Compressing the wide spectrum of human emotional experiences into a limited set of fixed categories may inevitably risks loss of meaningful emotional information (Barrett, 2019; Russell, 1994), often neglecting to consider the roles of cultural, contextual, and individual differences (Elfenbein et al., 2007).

More importantly, much of past research supporting BET has relied heavily on posed static expressions. While these expressions are useful in controlled settings, they frequently exaggerate emotional intensity and fail to capture the fluidity and spontaneity of facial expressions in real-world contexts (Krumhuber et al., 2023). This over-reliance on static, exaggerated facial expressions has significant implications for how facial expressions are understood and studied. By focusing primarily on static frames captured at peak emotional intensity, researchers tend to emphasise the morphological aspects of facial expressions more than the temporal changes – the way expressions unfold over time – as those documented in FACS (Krumhuber et al., 2021b). Given that FACS itself was developed primarily based on posed, static expressions, which are often designed to convey specific emotional meanings (Krumhuber et al., 2021a), this approach overlook the importance of other facial areas that may not be well-documented in FACS. This reliance raises questions about the ecological validity of the findings, prompting ongoing debates and calls for more comprehensive research that incorporate the dynamic and spontaneous aspects of emotional experiences.

Specifically, focusing solely on the features of static expressions risks failing to account for the broader temporal dynamics that naturally occur in everyday interactions. Research has shown that dynamic information, such as how quickly a smile forms or the gradual fading of a frown, provides crucial context that enhances the recognition and interpretation of emotions (Krumhuber et al., 2013; Cunningham & Wallraven, 2009). By neglecting these dynamic cues, studies relying on static images may provide an incomplete or even misleading picture of how emotions are communicated through facial expressions.

Similarly, limiting the analysis to highly controlled, posed expressions may encourage observers to focus too heavily on distinct facial features, such as the mouth or eyes, while ignoring the gestalt processing of the face that occurs in real-life context (Krumhuber et al., 2013). In dynamic, spontaneous expressions, the meaning of an emotion is conveyed not just through isolated features but through the interplay of multiple facial areas and the overall movement of the face. This integrated approach is essential for understanding more ambiguous or blended emotions, which often require observers to combine information from different parts of the face and the temporal progressions of the expression (Jack et al., 2014).

Additionally, BET fail to account for the full range of variability observed in spontaneous and natural expressions. Ekman's neurocultural theory suggests that facial expressions of basic emotions have universal biological basis, yet their manifestation is shaped by cultural norms that determine the frequency, intensity, and appropriateness of these expressions across different contexts (Ekman, 1972). This cultural modulation may result in notable differences in facial morphology across cultures, as supported by findings from Elfenbein and colleagues (2007), who showed how cultural background influence the expression and perception of emotions. Moreover, research indicates that frequency of exposure to certain FEs can influence recognition accuracy, with commonly encountered expressions being more easily recognised than less frequent ones (Calvo et al., 2014).

Furthermore, studies by Durán and Fernández-Dols (2021) and Reisenzein and colleagues (2006) suggest that many spontaneous expressions deviate from the prototypical facial configurations defined by FACS, revealing a broader range of variability in how emotions are expressed. This supports Fridlund's (1994) behavioural ecology view, which posits that FEs function more as social signals communicating intentions than as mere reflection of internal emotional states. Nevertheless, previous research suggests that spontaneous expressions aligning with prototypical patterns can still be rapidly and accurately recognised (Sauter & Fischer, 2018) supports the idea that the fundamental components of basic emotions remain influential in the recognition of emotions. Together, these findings emphasise the interplay between universal emotional signals and culturally specific patterns of expressions.

In summary, BET and FACS have significantly advanced our understanding of facial expressions. However, their reliance on static, posed expressions limits their ability to capture the dynamic and spontaneous nature of real-world emotional communication. Existing research has shown that spontaneous expressions often deviate from the conservative FACS prototypes, raising questions about the role of such prototypes in spontaneous expression recognition. To improve the ecological validity of facial expression research, future studies must move beyond static frames and when it comes to capturing the dynamic and spontaneous nature of real-world emotional communication. The current project aims to refine and expand BET through the integration of spontaneous and dynamic expressions into empirical research, providing more comprehensive approach to studying facial expressions. In the following sections, I will discuss the distinctions between static versus dynamic and posed versus spontaneous expressions, both in terms of their production and recognition.

1.2 Static versus Dynamic Facial Expressions

Human FEs are inherently dynamic, evolving over time through distinct phases of onset, peak, and offset. This progression furnishes a detailed temporal structure that conveys different nuances of FE (Nishiyama et al., 2005). Despite this, the study of FE recognition has historically been dominated by the use of static images, capturing only fleeting moments often at peak intensity (Krumhuber et al., 2021b). Although such snapshots can effectively differentiate between different emotions (Ekman & Friesen, 1978), they inadequately represent the movement and fluidity of facial expressions as they occur in daily experiences. Beyond simple facial muscle activities, FEs encompass a series of micro-movements that articulate the narrative of emotional states (Bould & Morris, 2008; Morishima et al., 2001). This dynamic interplay is evident not only in observable changes but also in the subtle dynamics of timing and intensity, elements that are absent in static portrayals.

1.2.1 Static Facial Expression Recognition

Having acknowledged the dynamic nature of FEs, it is important to have a closer look at why traditional reliance on static portrayals has prevailed in emotion recognition research. Static expressions have long been a cornerstone in the exploration of human emotions, offering a detailed view of facial behaviour typically at its most expressive moments. By focusing on the peak of emotional portrayals, researchers glean insights into the muscle-driven morphological changes that underlie specific facial behaviours (Krumhuber et al., 2023). For example, FACS has traditionally been focused on the analysis of static expressions, allowing for the precise dissection of the constituents of emotional displays at fixed points in time. Consequently, the presence and frequency of AUs have been prioritised over their duration in representing emotional displays.

Static displays, especially those posed ones, can be tightly controlled, thereby minimising extraneous sources of image variation such as cultural and individual differences (e.g., age & gender; Dawel et al., 2021). Not surprisingly, these images have been widely used in studies exploring the encoding and decoding of FEs (Calvo & Nummenmaa, 2016; Barrett et al., 2019). The intricate interplay of facial muscles in these static cues, though devoid of temporal dynamics, provides crucial emotional information, particularly for basic emotions (Gold et al., 2013). For example, the lift of an eyebrow or the curve of a lip can articulate a spectrum of feelings, from happiness to sadness, and anger to surprise (Ekman & Friesen, 2002).

Given the controlled and clear depiction of static displays in conveying emotional information, these frozen snapshots were found to be sufficient in representing basic emotions with recognition rates ranging from 70 to 90% (Goeleven et al., 2008; Palermo & Coltheart, 2004; Tottenham et al., 2009). The peak intensity of static expressions contributes significantly to their recognisability, creating distinct, prototypical facial configurations that are readily identifiable (Hess & Kleck, 1990).

Empirical investigations into static FEs revealed that specific facial regions yield informative value in conveying distinct basic emotions. For example, areas around the eyes and upper half of the face, are pivotal for recognising emotions such as fear, anger, and sadness, while the mouth region primarily conveys signals of happiness and disgust (Blais et al., 2017; Yitzhak et al., 2021). The necessity for integral processing arises when the same AUs are involved across multiple emotions, necessitating the consideration of additional cues for accurate interpretation. For example, both fear and surprise involve raised eyebrows, but the presence of brow furrowing suggests fear, whereas an open jaw indicates surprise (Krumhuber et al., 2023). This detail illustrates the reliance on both localised facial features and their collective, gestalt processing for accurate emotion inference, reflecting the intricacies of facial muscle coordination in emotional expression (Calder et al., 2000a).

Moreover, studies employing simplified representations, such as line-drawn or point-light faces, have illuminated the fundamental role of basic visual properties in emotion recognition (Krumhuber et al., 2023). These studies have shown that individuals are capable of recognising distinct emotions even through minimal visual cues (Atkinson et al., 2004, 2012; Bidet-lldei et al., 2020). Importantly, accentuating key differences in these two-dimensional shapes increases their distinctiveness and perceived intensity, which in turn facilitates faster and more accurate recognition (Calder et al., 1997, 2000). Thus, research on static expressions can approximate aspects of human facial recognition, including prevalent errors and confusion, suggesting that some aspects of human emotion recognition may be grounded in fundamental visual properties.

Despite their utility, static expressions have inherent limitations. They lack the temporal dynamics of real-life emotional expressions, potentially leading to an inaccurate understanding of emotional expressions. The static nature fails to capture the fluidity and progression of emotional responses. Consequently, there is a growing, yet limited consensus in the field towards integrating dynamic elements into the study of emotion recognition, to ensure a more valid approach for studying lifelike facial behaviour. A recent review showed a significant albeit modest shift, with only 13% of articles published in psychology between 2000 and 2020 incorporating dynamic stimuli for emotion-related questions (Dawel et al., 2021; Krumhuber et al., 2023).

1.2.2 Dynamic Advantage in Facial Expression Recognition

When evaluating the utility of facial motion in emotion recognition, it is imperative to ascertain whether dynamic information adds unique value beyond that provided by static representations and to pinpoint the nature of these additional insights. Unlike static expressions, dynamic expressions provide detailed information on both the structure and movement of the face. Specifically, they include spatial details about the positioning and arrangement of facial features, as well as temporal information, such as how quickly a smile forms or a frown deepens (Krumhuber et al., 2013). The displacement, velocity, and acceleration of movement closely mirrors the expressions encountered in everyday life (Barrett et al., 2019; Krumhuber et al., 2023). Converging evidence showed that movement facilitate the recognition of emotion with which higher recognition rate than that of static point-light displays (Atkinson et al., 2012; Bidet-lldei et al., 2020). In this vein, movements seem to provide additional benefits that static expressions fail to provide.

Individuals seem to be sensitive to this spatiotemporal information as people recognise dynamic faces with greater accuracy and confidence than static displays of facial expressions (Lederman et al., 2007). Such sensitivity towards dynamic information may be automatically ingrained as people are found to accurately reproduce the progress of expressions from a scrambled set of image sequences (Edwards, 1998). These elements are integral not just for understanding emotional states but also for making social inferences – judgements about individuals' intentions, relationships and social context (Arsalidou et al., 2011; Krumhuber et al., 2013; Marian & Shimamura, 2013).

Dynamic expressions comprise multiple images over time, thereby providing a larger number of static cues than a single image (Recio et al., 2011; Krumhuber et al., 2023). Research indicates that higher frame rates in dynamic sequences facilitate a more effective extraction and recognition of emotional meanings (Bould & Morris, 2008; Calvo & Nummenmaa, 2016). However, this advantage is not due to the mere increased quantity of static cues but is attributed to indicating the direction of change. Such unfolding of expressions heightens sensitivity to changes in facial features and their trajectories (Cunningham & Wallraven, 2009; Krumhuber et al., 2013). The sequential order of these changes is crucial since disrupting the natural temporal progression was found to significantly impair emotion recognition (Edward, 1998).

Accordingly, individuals seem to be attuned to detecting the temporal progression of expressions which can alert emotional meaning behind the faces. This integration over temporal sequences captures the transitions from onset through peak to resolution, thereby allowing observers to discern the unfolding of emotional expressions. This transitional shift lacks in the momentary presentation of static images (Cunningham & Wallraven, 2009; Jack et al., 2014).

Considering the additional benefits provided by dynamic cues, studies demonstrate that facial movements lead to higher classification rates for emotion recognition, often outperforming static expressions. For example, hit rates for dynamic expressions range from 48 to 98% (Dupré et al., 2020), surpassing the rates achieved by static expressions. This advantage is particularly salient in instances of degraded or subtly expressed emotions, where static representations struggle to convey emotion effectively. Studies using point-light displays highlight the effectiveness of movement in conveying emotional information, demonstrating that dynamic cues aid emotion recognition where static point-light displays fall short (Bassili, 1978; Valentine & Bruce, 1988). Additionally, research leveraging synthesised facial animations corroborates the facilitative effect of dynamic presentations in emotion recognition by showing higher recognition accuracy for FE with movements (Kätsyri & Sams, 2008; Wehrle et al., 2000).

Furthermore, dynamic expressions are not only perceived as more genuine and intense (Zloteanu & Krumhuber, 2020) but also instrumental in detecting complex emotional states where several basic emotions are compounded or blended (Adams et al., 2015; Bassili, 1978). Particularly notable is the contribution of mouth movements in conveying a spectrum of emotions (Eisenbarth & Alpers, 2011). The way a smile gradually transitions into an expression of surprise, or how tension around the lips intensifies into a display of anger, offers detailed insights into how multiple emotional states are communicated within a dynamic sequence.

These findings collectively emphasise the added value of incorporating dynamic cues in the study of emotion recognition, highlighting the additional information provided by movement beyond what static images can provide.

1.2.3 Conditional Advantage of Dynamic Expression

While there is considerable evidence supporting the advantage of dynamic expressions in emotion recognition, their advantage is not absolute and can vary depending on the clarity and distinctiveness of static images. For example, Kätsyri and Sams (2008) elucidated that the benefit of movement facilitated emotion recognition for synthesised facial animations but not for natural (albeit posed) expressions. Their findings indicate that the dynamic advantage may be condition-specific. Similarly, Kamachi (2001) and Gold and colleagues (2013) found that when static images are highly distinctive, with near-perfect identification accuracy, the superiority of dynamic expressions tends to wane. This pattern suggests that dynamic expressions predominantly offer supplementary information where static cues are ambiguous, degraded or subtly expressed (Ambadar et al., 2005; Harwood et al., 1999; Wehrle et al., 2000).

In support of this notion, studies suggest that the necessity of movement for accurate emotion recognition may diminish when static expressions are presented clearly in full intensity (Ambadar et al., 2005; Bould & Morris, 2008; Tobin et al., 2016; Blais et al., 2017). This conditional dynamic advantage suggests a compensatory role for motion, particularly in filling informational gaps left by static cues. The divergent findings across past studies emphasise the importance of considering specific conditions in determining the relative benefits of dynamic versus static representation in emotion recognition and suggest that the effectiveness of dynamic cues is contingent upon the limitations of their static counterparts.

In summary, recent research in emotion recognition highlights the unique contribution of dynamic facial expressions in conveying emotional information. These dynamic expressions provide spatiotemporal cues that static images lack, significantly aiding in the recognition of emotions, especially under conditions where static cues are insufficient or ambiguous to represent emotions. The dynamic advantage stems not simply from the increased number of frames, but from the directional changes of facial movements, which observers process intuitively. However, this advantage appears to be condition-driven in terms of visual quality and intensity. Counterevidence suggests that dynamics mainly serve a compensatory role, particularly when static information is either unclear or unavailable. This raises important questions about the specific conditions under which dynamic facial expressions offer a recognitional advantage. The specific conditions in which dynamic cues offer the most significant benefits remain to be fully delineated, particularly with regard to the expressive features.

In addition to dynamic-static comparison, much of the previous research on FEs has primarily focused on highly controlled, posed expressions. While these expressions are useful for isolating key features, they lack the spontaneous, fluid nature of real-world emotional displays. This reliance may also have limited the ecological validity of previous findings, as genuine expressions are far less structured and predictable. This gap highlights the importance of incorporating spontaneous expressions in research to better reflect those encountered in daily interactions. In the next section, we will further explore the distinctions between posed and spontaneous expressions and discuss their implications for facial expression recognition.

1.3 Posed versus Spontaneous Facial Expressions

1.3.1 Encoding Posed and Spontaneous Expressions

Studies have demonstrated that individuals have the ability to distinguish between posed and spontaneous FEs (Dawel et al., 2017). Despite this, past research on FEs has been predominantly focused on the analysis of posed expressions, preferred for their experimental control and easier recognisability. Distinguishing between posed and spontaneous expression is crucial, not just in psychological research but also in practical applications, such as security screening, legal context, or customer service, where accurately interpreting facial expressions can significantly impact decision-making and outcomes.

In FE research, posed expressions are typically elicited through direct instruction for facial movements, often guided by the FACS manual (Ekman et al., 2002). An alternative approach involves instructing participants to show facial expressions they associate with specific emotional states. Actors are also asked to mimic the example face depicting a target emotion (Aifanti et al., 2010). Such controlled approaches result in stylised displays of emotion consistent across individuals and cultures (Krumhuber et al., 2021). These expressions, albeit effective for distinguishing discrete emotions (Ekman et al., 1987), may not adequately capture the subtlety and complexity inherent in the FEs encountered in daily life. This limitation brings into question the extent to which findings derived from posed expressions may, in fact, have little to do with spontaneous nonverbal behaviours (Motley & Camden, 1988).

On the other hand, spontaneous expressions are elicited via induction (e.g., watching emotion-evocative videos, hearing jokes, or specific tasks) or simulation (e.g., by recalling emotion-relevant memories; Gross & Levenson, 1997), thereby presenting facial displays that better resonate with the genuine emotional states while maintaining necessary controls. Unlike posed expressions, different elicitation techniques employed to elicit spontaneous ones introduce heterogeneous variability, not confined to standardised fixed signals (Gross &

Levenson, 1995). This diversity likely reflects the natural variability across expressions encountered in everyday life, offering an ecologically valid perspective on human expression. Therefore, it is imperative to acknowledge the fundamental difference between posed and spontaneous expressions to facilitate the lifelike assessment of FEs.

1.3.2 Distinctive Morphological Appearance

In their very nature, there are distinct neural pathways underpinning posed and spontaneous FEs, each originating from separate brain regions. While deliberate facial movements arise from the cortical motor strip, involuntary emotional facial actions are rooted in the subcortical areas of the brain (Meihlke et al., 1973). These pathways not only innervate different facial muscles but can also influence the dynamics and muscular involvement of the expressions (Morecraft et al., 2001). For instance, when individuals are instructed to simulate an expression like fear, the resultant expression differs from one that emerges spontaneously (Bartlet et al., 2006; Ekman & O'Sullivan, 1991). This distinction highlights the inherent difference in how our brain processes and produces posed versus spontaneous expressions.

The different elicitation and neural processes lead to different morphological patterns between the two types of expressions. Spontaneous FEs often display varied configurations, contrasting with the uniform, stylised activation pattern seen in posed expression (for a review see Calvo & Nummenmaa, 2016). A mere fraction (0% to 11%) of spontaneous expressions strictly adhere to prototypical AU patterns as outlined in FACS (Durán & Fernández-Dols, 2021; Reisenzein et al. 2006; Wang et al. 2010), with many exhibiting variations including additional or missing AUs that are not accounted for by FACS criteria (Bartlett et al., 2006; Smith et al., 1986). These morphological nuances are especially pronounced in emotions like surprise or disgust requiring a larger number of facial muscle involvements (Namba et al., 2017). However, it is important to note that these morphological patterns are not always reliable indicators of expression types. While simple emotions like happiness are often associated with genuine expressions indicated by specific AUs (e.g., AU6: orbicularis oculi + AU12: zygomaticus major; Figure 1.2), studies revealed that even untrained individuals can deliberately contract these muscles to simulate genuine expressions (Gunnery et al, 2013). Some studies also reported overlaps in facial configuration between spontaneous and posed expressions, (Carroll & Russell, 1997; Gosselin et al., 1995; Scherer & Ellgring, 2007).

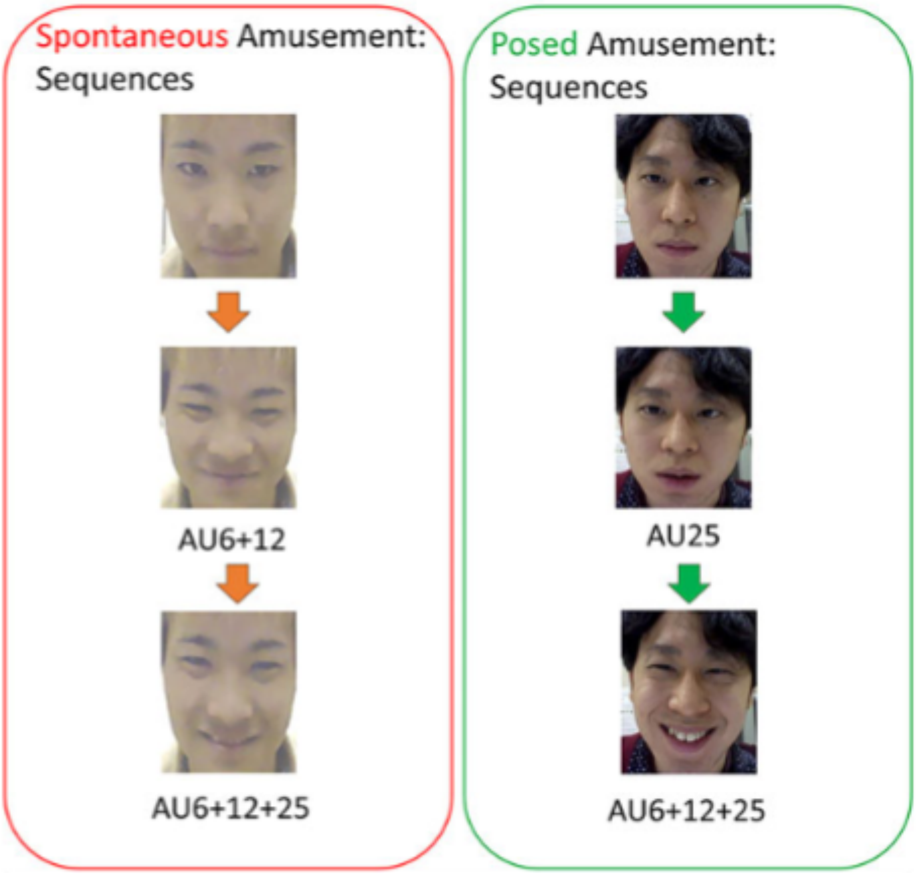


Figure 1.2. Morphological difference between posed and spontaneous smiles (Namba et al., 2017)

Building on this, posed and spontaneous expressions are significantly different in their respective intensity. Typically, spontaneous FEs are subtler in their AU and overall expression intensities compared to their posed counterparts (Saumure et al., 2018). The posed expressions often display exaggerated intensities, especially in diagnostic AUs (Park et al., 2020; Zeng et al., 2009), resulting in more stereotypical facial displays. On the other hand, the transient movement in spontaneous expressions can obscure their defining features, aligning with findings on the fleeting nature of micro-expressions (Davison et al., 2018). Such variation in intensity has been linked to asymmetry patterns between posed and spontaneous FEs (Ekman et al., 1981; Powell & Schirillo, 2009). For instance, genuine emotional indicators tend to emerge more frequently and intensively in the upper half of the face, whereas the lower half often reflects the intended, posed emotions (Costantini et al., 2005; Park et al., 2020).

1.3.3 Distinctive Temporal Dynamics

The distinctions between posed and spontaneous expressions are further elucidated by examining spatial characteristics alongside temporal dynamics. Posed expressions are often associated with rapid onsets and brief duration (Lander & Butcher, 2020; Schmidt et al., 2006), a feature that contrasts sharply with smoother trajectories with less abrupt dynamics of spontaneous expressions (Ekman, 2003; Ekman & Friesen, 1982). This progression often (but not always; Namba et al., 2017) results in a more gradual onset and offset, with extended durations that mirror the natural progressions of emotions (Cohn & Schmidt, 2004; Hess & Kleck, 1990; Schmidt et al., 2009). Particularly during the offset phase, spontaneous expressions diminish gradually rather than disappear abruptly in the post-elicitation phase (Guo et al., 2018; but see Schmidt et al., 2006). Yet, intriguingly, a prolonged apex duration in expressions has been associated with the perception of reduced genuineness (Krumhuber & Kappas, 2005; Guo et al., 2018), hinting towards the distinction between expression recognition and perception. Further studies have highlighted the different timing and small amplitude of spontaneous expressions, distinguishing them from their often-amplified posed counterparts (Cohn & Schmidt, 2004; Hess & Kleck, 1990).

1.3.4 Decoding Posed and Spontaneous Expression

Accordingly, empirical evidence showed observers' discernment capabilities between deliberate and genuine expressions, emphasising the crucial role of both morphological and temporal differences between types of expressions (Hess & Kleck, 1994; Gunnery & Ruben, 2016; McLellan et al., 2010). Yet, this capacity for differentiation is not consistent across all emotions. For instance, research by Dawel and colleagues (2017) reveals that observers face particular challenges in distinguishing between deliberate and genuine expressions of fear. Notably, the ability to accurately differentiate between types of expression significantly improves with dynamic presentation, supporting the notion that movement enhances the perception of authenticity (Zloteanu et al., 2018). This pivotal role of morphological and temporal differences is supported by the performance of automated classifiers, which, by capitalising on these distinctions, demonstrate exceptional accuracy in distinguishing types of expressions (for review, see Jia et al., 2021).

The differences between posed and spontaneous FEs further impact their recognition. Specifically, research consistently demonstrates a marked decline in recognition accuracy for spontaneous expressions, often falling below 40% (Kayyal & Russell, 2013; Naab & Russell, 2007), with some instances even falling below chance levels (Wagner, 1990). Such recognition rate is notably lower than those reported for posed expressions, which generally exceed >70% (Calvo & Nummenmaa, 2016). Furthermore, automated classifiers also exhibit diminished performance in recognising spontaneous expressions (Krumhuber et al., 2021b). While posed expressions are suggested to have a degree of universality in recognition, spontaneous expressions show substantial variability across cultures (Matsumoto et al., 2009).

Direct comparisons between the recognition and perception of spontaneous and posed FEs are scarce. However, those that exist generally support the notion that posed expressions are typically better recognised than spontaneous ones (Mortley & Camden, 1998; Russell, 1994). This trend extends to perceivers' ratings of emotional valence, which are more often accurate for posed expressions, highlighting the interpretive challenges associated with spontaneous expressions (Zuckerman et al., 1976). Interestingly, the discrepancy in recognition accuracy varies dramatically across different emotion categories, with posed expressions of anger being recognised more accurately than spontaneous ones, whereas spontaneous expressions of sadness achieve higher recognition accuracy than posed ones (Jürgens et al., 2015). Nevertheless, research suggest that these findings may be contingent on the specific stimuli used. Depending on how the expressions were elicited and the context in which expressions occur, spontaneous expressions can sometimes be intense and prototypical enough to surpass posed expressions in terms of recognisability (Sauter & Fischer, 2018). This finding highlights that spontaneous expressions, though typically more subtle and variable, can at times closely resemble prototypical forms, leading to better recognition under certain conditions and for certain emotions.

The challenge of recognising spontaneous expressions is further complicated by increased confusion among emotion categories. Observers frequently assign multiple emotion labels to a single spontaneous expression (Calvo & Nummenmaa, 2016), reflecting the inherent ambiguity in these displays. Unlike posed expressions, which typically present exaggerated and clear-cut signals of basic emotions, spontaneous expressions do not always involve fixed signals with greater subtlety. As a result, the accurate interpretation of spontaneous expressions demands an understanding beyond mere facial muscle configurations, requiring social knowledge and contextual information (Hassin et al., 2013; Parkinson, 2013).

In summary, the literature delineates clear distinctions between posed and spontaneous expressions, highlighting differences in their elicitation methods, morphological characteristics, and dynamic properties. In particular, spontaneous expressions exhibit a greater degree of morphological flexibility, diverging from the more rigid, prototypical patterns typically observed in posed expressions. In dynamic properties, spontaneous expressions tend to follow smoother trajectories, a stark contrast to the shorter onset and offset durations of posed expressions, which often display irregular and abrupt movement patterns. These differences profoundly influence the accuracy with which these expressions are recognised and their perceived authenticity. Notably, spontaneous expressions typically show poorer recognition rates compared to their posed counterparts.

While research on posed expressions has been extensive, studies investigating spontaneous expressions remain relatively limited. Given the natural variability and subtlety of spontaneous expressions, more research is needed to understand how these expressions are produced and recognised. Exploring spontaneous expressions could provide valuable insights, particularly in context where emotional authenticity is critical, such as human-computer interaction, social communication, and affective computing. Having established the key difference between posed and spontaneous expressions, it is crucial to further explore the specific features that contribute to their recognisability. In the following section, we will examine how expression features such as prototypicality, intensity, and ambiguity play a role in shaping the recognition of facial expressions.

1.4 Prototypicality, Intensity and Ambiguity

Despite consistent findings that spontaneous expressions are recognised with lower accuracy compared to posed ones, there has been limited investigation into the specific features that influence recognition performance. Considering the differences between posed and spontaneous expressions collectively, it becomes apparent that spontaneous expressions tend to be less prototypical and intense, yet more ambiguous in their presentation.

1.4.1 Prototypicality

Drawing upon the BET, it is assumed that the recognisability of FEs may largely depend on their alignment with stereotypical emotional displays. Prototypicality refers to the degree to which a combination of AU within expressions matches the classical depictions of basic emotions (Ekman et al., 2002). Empirical evidence showed the importance of such prototypical configurations in facilitating emotion recognition. By manipulating the presence of prototypical AUs, Matsumoto and Hwang (2014) found that expressions featuring highly prototypical cues were recognised with greater accuracy. Moreover, those expressions achieved faster reaction time with higher perceived intensity (Young et al., 1997; Matsumoto et al., 2009). Conversely, expressions deviating from these prototypical configurations – through missing or altered AUs - are still recognisable but perhaps less distinctly representative (Cabeza et al., 1999; Gaspar et al., 2014). The distinctiveness of prototypical expressions, therefore, lies in the human ability to detect emotion with relative ease based on facial components.

In addition to the basic six emotions, research has increasingly recognised a broader range of emotions that may have distinct facial prototypes. Keltner and colleagues (2019) introduced the concept of “new basic emotions,” including more complex social emotions like pride, embarrassment, and love. These emotions, while not traditionally part of Ekman’s framework, show consistent facial expressions across cultures. For example, pride is often expressed with a slight smile and an upward tilt of the head (Tracy & Robins, 2007), whereas embarrassment may involve a downward gaze, a modest smile (Keltner, 1995). These findings suggest that additional emotions, beyond the traditional six, may also have consistent signals, although further research is needed to confirm these patterns.

1.4.2 Intensity

Nevertheless, the mere presence of prototypical cues does not solely influence the recognisability of expressions; these cues must be intense enough to be discernible to observers. Expression intensity refers to the strength or degree of activation of facial muscles, reflected on how pronounced the facial expression appears. This characteristic plays a crucial role in how effectively expressions are recognised. In general, intense expressions with pronounced facial cues are typically identified more accurately and rapidly (Hess et al., 1997). This effectiveness is likely attributed to the heightened visibility of prototypical configurations in intense expressions (Calvo & Nummenmaa, 2016). Conversely, subtler expressions, including micro-expressions, pose greater challenges in recognition due to their low intensity and fleeting nature, making them less conspicuous (Stanciu & Albu, 2019).

The significance of intensity in facial expression recognition is particularly evident in studies employing morphing techniques. These techniques present dynamic sequences that show gradual progression of expression from neutral to peak intensity. Rodger and colleagues (2018) utilised morphing techniques to demonstrate that expressions at their peak intensity are recognised more accurately than those in their onset or offset phases. Furthermore, studies found that observers can identify emotions more quickly and with greater confidence as the expressions near their maximum intensity (Young et al., 1997). This body of evidence suggests the important role of intensity in enhancing the recognisability of expressions.

Interestingly, highly intense expressions are sometimes perceived as less authentic, possibly due to their exaggerated nature (Zloteanu & Krumhuber, 2021). In everyday interactions, FEs are more commonly found at low to medium intensities, as such expressions in varying intensity may offer a more realistic representation of emotional displays (Adolphs & Tranel, 2004; Motley & Camden, 1988). Consequently, varying the intensity within emotion recognition tasks can enhance sensitivity to these subtle differences (Calder et al., 1997). This suggests that subtle expressions are not necessarily a disadvantage in perception; rather they demand more focused attention to discern their emotional content.

1.4.3 Ambiguity

While prototypical facial expressions in high intensity provide clarity in emotion recognition for singular emotions, their utility becomes problematic when expressions blend multiple emotions simultaneously or present only partial prototypes, such as happy crying or fake smiles. Ambiguity in facial expressions refers to the presence of multiple, overlapping emotional signals, which complicates the recognition of a single emotion. Diverging from the simplicity of singular-emotion prototypes, the intersections of contradictory emotional signals introduce categorical uncertainty in facial expressions. Such ambiguous expressions are more reflective of real-life expressions, where emotions are often interwoven, not isolated depending on the context (Du et al., 2014). These ambiguous expressions also introduce different emotional nuances beyond basic emotions (Du & Martinez, 2015).

Past research has typically manipulated face stimuli to present contradictory emotional cues, such as combining angry eyes with a smiling mouth (Kinchella & Guo, 2021), complicating the task of accurate emotion identification. Such studies highlight how human perception is swayed by biases, including a tendency towards negative interpretations when faced with ambiguous emotional signals (Ito et al., 2017). As such, the recognition accuracy for these expressions is typically lower than those with singular emotional cues (Neta & Whalen, 2010). Ambiguity in expression also tends to diminish perceived intensity (Kinchella & Guo, 2021), possibly because the human visual system has a limited capacity to process multiple emotions simultaneously (Ito et al., 2017). Additionally, viewing conditions like image resolution and spatial frequency can impact the perceived ambiguity (Kinchella & Guo, 2021), suggesting a complex interplay between the expression itself and the viewing quality.

Our study also showed that ambiguity is one of the key indicators in predicting recognition performance, both for human and automated recognition tools (Kim et al., 2023).

While previous studies often emphasise that real-life expressions are ambiguous, most studies (if not all) have tended to rely on morphing techniques or varied image qualities to manipulate expressions. Our study addresses this gap by being one of the first to measure ambiguity in both posed and spontaneous FEs. We highlighted the challenges arising from the lack of a common metric and varying definitions of expression ambiguity (degree of closeness in categorical boundaries or the omission of key emotional configurations).

It has also been argued that the role of characteristics might be influenced by common methodological designs in emotion recognition studies. Often, forced-choice response tasks do not include a neutral option (Rotshtein et al., 2010), potentially leading participants to choose emotions they might otherwise perceive as non-expressive. This trend could inflate recognition rates for more prototypical expressions, while artificially deflate accuracy rates for ambiguous and subtle expressions. Particularly for ambiguous expressions, observers may focus on the most salient emotional signals (Limbrecht-Ecklundt, 2013), potentially overlooking subtler cues, as these tend to deviate from stereotypical emotional configurations. In support of this notion, research has shown that moderate-intensity expressions often require longer reaction time compared to subtle expressions, particularly when tasks are equipped with neutral options (Wells et al., 2016). Additionally, the relationship between prototypicality, intensity, and recognition accuracy is not consistent across all emotions. For some, like fear and surprise, changes in intensity or the presence of ambiguous cues may not substantially alter recognition accuracy (Hoffmann et al., 2010). This complexity highlights the intricate balance between expression prototypicality, intensity, and ambiguity, and how each characteristic contributes to the cognitive mechanisms that underlie emotion recognition.

Putting things together, the investigation into prototypicality, intensity and ambiguity illuminates the distinct challenges and advantages in recognising posed and spontaneous facial expression recognition. Prototypicality and intensity emerge as key elements that enhance the accurate and swift recognising of expressions, offering clear and strong emotional signals. In contrast, the ambiguity inherent in expressions, especially spontaneous ones, introduces significant challenges, often resulting in diminished recognition accuracy. This exploration suggests the intricate interplay between these factors, emphasising their significant impact on our perception and recognition of emotional cues. While these characteristics tap into morphological features it is also important to consider the impact of dynamic aspects influencing facial expression recognition.

Building on the discussion of prototypicality, intensity, and ambiguity in facial expression recognition, it is crucial to now explore how these characteristics are addressed by both human observers and automated systems. With the advancements in technology, Automated Facial Expression Analysis (AFEA) tools have emerged as a powerful alternative to human observers, particularly for recognising prototypical expressions. However, just as humans face challenges in interpreting subtle or ambiguous expressions, machines exhibit their own set of limitations, particularly in the recognition of non-prototypical spontaneous facial displays. In the following section, I compare the capabilities of human observers and AFEA systems, examining their respective strengths and limitations in facial expression recognition.

1.5 Human versus Machine Facial Expression Recognition

1.5.1 AFEA: A General Introduction

The integration of affective computing into emotion research has led to significant utilisations of AFEA tools, fostering a multidisciplinary approach to studying emotional communication across computational and psychological domains (D’Mello et al., 2018). By leveraging cutting-edge technology to analyse facial cues, AFEA contributes to various applications, from enhancing user interface design to improving mental health diagnostics, embodying a significant stride towards machines that can understand and interact with humans (Calvo & D’Mello, 2010; Dupré et al., 2019). Given that AFEA is nowadays widely accessible, emotion classification using commercially available software (e.g., AFFDEX, FACET, FaceReader, OpenFace) is of increasing research interest. Predominantly trained on the foundational principles of the FACS, these tools rely on morphological analysis of FEs to discern emotional states from human faces (Calvo et al., 2018; Ekman et al., 2002).

AFEA tools typically employ a structured three-step process to recognise facial expressions (Martinez et al., 2017; Sariyandi et al., 2017). This initial phase involves detecting faces within images or video, identifying them based on shape, morphological features, and configurations. The second phase focuses on detecting and localising facial landmarks – specific points defined by their geometric properties – and monitoring their changes over time. The final phase involves analysing the movement patterns of these facial landmarks, classifying them into pre-defined emotion categories or dimensions based on their configurations. This detailed methodological process suggests the technical sophistication of AFEA tools.

These classifiers extend the scope of emotion recognition, offering standardised and efficient data processing beyond human capabilities. They excel in eliminating several sources of noise-related variance like participant fatigue, inherent in human assessment (Pantic & Rothkrantz, 2000). The Facial Expression Recognition and Analysis (FERA) challenges showcase the capabilities of these algorithms, where the top-performing algorithmic models have reached impressive recognition rates of 84% for basic emotion recognition (Valstar et al., 2012), and 94% accuracy in FACS recognition for specific AU (Valstar et al., 2017). However, the past training and refinement of these tools have been heavily reliant on posed expressions in controlled settings (Pantic & Bartlett, 2007). When it comes to FACS recognition, their efficacy is often limited to a select number of AUs, varying from as few as 2 (Jian-zheng et al., 2011) to a maximum of 7 (Baltrusaitis et al., 2015). It is important to note that many of these in-house algorithmic models are proprietary, which may not be easily accessible for cross-laboratory research (Dupré et al., 2020).

1.5.2 Commercial AFEA Recognition

Recent advancements in AFEA have led to a surge in both commercial and open-source algorithms, making AFEA more accessible (Cohn & Sayette, 2010). As documented by Littlewort and colleague (2011), these classifiers have been proficient in accurately identifying basic emotions and AUs simultaneously across various stimulus sets, particularly for posed expressions. The comparison of the human ability to discern facial cues with the precision of AFEA tools also reveals a compelling narrative that highlights distinctive strengths and limitations inherent in human and machine recognition. For example, Lewinski and colleagues (2014) showed the impressive performance of FaceReader which correctly recognised emotions with 89% accuracy, surpassing the human recognition rate of 85% on the same tasks. Others studies also consistently showed that FACET outperformed human observers for posed expressions, but worse or comparable performance for subtle and spontaneous expressions (Krumhuber et al., 2021; Yitzhak et al., 2017).

Given that most AFEA tools are trained on highly prototypical expressions (Calvo et al., 2018), the exaggerated intensity of standardised expressions typically aids in the featural analysis by machines (Pantic & Bartlett, 2007). Interestingly, by comparing three machine classifiers (Azure Face API, FaceReadear, and Face++) with human observers, Küntzler and colleagues (2021) showed that machines outperformed humans for both posed and spontaneous expressions, making AFEA classifiers an attractive alternative to human observers. The recognition capability of AFEA tools has also been compared with other physiological measures such as EMG, showcasing the comparable performance of automated tools compared to EMG for measuring facial movements (Beringer et al., 2019; Kulke et al., 2020; Höfling et al., 2021).

However, the performance of AFEA tools is not uniformly high across all classifiers. Cross-classifier evaluations, such as those conducted by Dupré et al. (2021), indicate a variance in accuracy rates ranging from 43% (AFFDEX) to 68% (FaceReader), highlighting the heterogeneity in algorithmic efficiency and training methodologies among different AFEA tools. Stöckli and colleagues (2018) also showed varied performance where FACET consistently outperformed AFFDEX on both valence and categorisation tasks. This variability becomes even more pronounced when assessing the recognition of different emotion categories, revealing certain emotions as consistently more challenging for AFEA systems to accurately classify (Dupré, 2021; Küntzler et al., 2021; Stöckli et al., 2018). Among these, happiness has been consistently recognised with the highest accuracy, whereas fear and disgust often being more challenging to recognise (Lewinski et al., 2014; Skiendziel et al., 2019). These recognition patterns are more pronounced in spontaneous expressions (Calvo et al., 2018). The variability in recognition performance across classifiers may be attributed to the different datasets used for testing. While most studies used existing databases (both pose and spontaneous) for testing (Dupré et al., 2020; Krumhuber et al., 2021), some utilised their own dataset (Stöckli et al., 2018; Tcherkassof & Dupré, 2020). The number of testing datasets also varies across studies.

Commercial tools have also demonstrated proficiency in FACS AU recognition, occasionally exceeding human-level accuracy. For instance, classifiers achieved indices of 0.69 and 0.66 in various static databases, just below the FACS certification threshold for human coders (0.7; Lewinski et al., 2014). Advanced systems like CERT (a precursor FACET commercial software) have shown remarkable accuracy, with an average recognition performance of 90.1%, highlighting their efficacy for real-time FACS analysis (Littlewort et al., 2011). Comparative evaluations of various classifiers indicated diverse performance strengths, with some excelling in static emotion recognition and others in dynamic context (Dupré et al., 2020; Lampropoulos et al., 2009). Notably, all systems effectively detected AUs above chance levels (Namba et al., 2021). Yet, their accuracy varied across different AUs, with some expressions being more challenging to recognise than others (Skiendziel et al., 2019). This discrepancy may arise from AFEA performance which is more effective with clear-cut facial actions but less adept at detecting subtler variations. Collectively, these findings position AFEA software as a capable alternative to human observers, at least for prototypical expressions. Similarities and differences between AFEA and human observers need further attention, with studies showing similar recognition patterns on non-prototypical, subtle, and dynamic expressions.

1.5.3 Similarities between Machines and Humans

Both machines and humans exhibit proficiency in identifying clear, prototypical expressions (Yitzhak et al., 2017), yet accuracy declines with non-prototypical expressions (e.g., spontaneous and naturalistic expressions; Sato et al., 2019) that deviate from basic emotion prototypes (Küntzler et al., 2021; Pantic, 2009; Stöckli et al., 2018). This decrease is more notable in AFEA, particularly when handling low-intensity expressions (Calvo et al., 2018; Küntzler et al., 2021), resulting in lower recognition rates compared to humans (Yitzhak et al., 2017).

Interestingly, similar confusion patterns emerge in both humans and machines, like mistaking fear for surprise (Calvo et al., 2018), which suggests that shared facial actions between certain emotions complicate recognition of both humans and machines (Lewinski et al., 2014). Considering that most AFEA algorithms are developed and trained using human-annotated data (Chen & Joo, 2021), this resemblance is perhaps expected. Specifically, the labelling process for training data highly relies on human perception. Furthermore, the methodology underlying machine-based recognition discerns specific patterns of facial movements or AUs and links them with corresponding emotion categories, a process heavily influenced by the perceptual interpretation inherent to human observers (Matsumoto et al., 2009; Tcherkassof & Dupré, 2020).

Additionally, technical factors such as illumination and image resolutions play a crucial role in further influencing the recognition ability of both groups, further complicating facial expression recognition (Khan, 2017; O'Toole et al., 2012). Several studies showed that stable lighting conditions significantly enhance recognition performance (Stratou et al., 2011), whereas inconsistent or fluctuating lights can hinder recognition (Wang et al., 2013; Nguyen et al., 2014). Furthermore, the complex backgrounds have been found to divert focus away from the face (particularly for human observers), thereby negatively affecting recognition (Righart & de Gelder, 2008; Sannikov et al., 2017).

1.5.4 Differences between Machines and Humans

In assessing the difference between human observers and AFEA classifiers, a critical focus emerges on their respective adaptability to dynamic expressions. As discussed, converging evidence suggests that humans often benefit from expressions incorporating movements (Ambadar et al., 2005; Krumhuber et al., 2013, 2023). This contrasts with AFEA systems, which historically have faced challenges in accurately accounting for expressions in motion (Namba et al., 2021; Tcherkassof & Dupré, 2020), possibly due to the large segments of frames displaying comparatively subtle intensity. In consequence, machine accuracy has been shown to drop for dynamic compared to static stimuli commonly taken at the peak of the emotional display (Stöckli et al., 2018; Skiendziel et al., 2019; Onal Ertugrul et al., 2023). This decline in machine accuracy may stem from AFEA systems inadequately integrating sequential facial movements into a cohesive emotional interpretation (Dupré et al., 2018). To date, the role of dynamic information in AFEA is still poorly understood, with performance varying substantially across stimulus conditions (Yitzhak et al., 2017; Dupré et al., 2019; Krumhuber et al., 2021). These findings highlight the need for further investigation into AFEA performance on dynamic expressions, particularly how these systems process and integrate temporal information.

In summary, the comparison between human and automated recognition of facial expressions has illuminated several key insights. Primarily trained on posed expression, AFEA tools are adept at recognising basic emotions from prototypical, posed expressions surpassing human levels yet face difficulties with the subtleties and ambiguities of spontaneous expressions. While substantial progress has been made, there are significant gaps, particularly in the development of automated systems capable of interpreting complex, dynamic expressions in real-world contexts as proficiently as humans. Bridging these gaps is essential for the development of more empathetic artificial intelligence systems, with broad implications for psychology, technology, and beyond. This area of research draws attention to the importance of considering the inherent complexity and ambiguity of human expressions when developing and training AFEA systems.

1.6 Unanswered Questions

First, the discourse surrounding the dynamic advantage in facial expression recognition provides supportive yet inconsistent effects of movement. The core of this debate centres on the conditions under which facial movements confer their facilitative effects. A significant body of research advocating for the dynamic advantage has frequently employed stimuli that are artificially degraded or distorted faces (Atkinson et al., 2012; Calder et al., 2000). This approach manipulates stimulus materials to make them more challenging to recognise – a context that markedly deviates from everyday facial expressions. This divergence prompts critical examinations of the applicability of findings on dynamic advantage to scenarios involving non-degraded faces. A key question, therefore, is whether facial dynamics consistently provide facilitative benefits on expression recognition when the static faces are clearly visible and undistorted, accurately representing emotion. Moreover, existing counterarguments suggest a negligible impact of facial movements on recognition under certain circumstances (Gold et al., 2013; Kamachi et al, 2001), hinting at a more complex interaction between facial movements and recognition processes than previously assumed. Thus, it becomes crucial to validate the conditions under which dynamic information significantly enhances expression recognition.

Second, previous research has typically focused on contrasting the recognition accuracy between static and dynamic expression based on a singular static frame, often neglecting the variability in representativeness across different static frames extracted from a dynamic sequence. Dynamic expressions inherently comprise a multitude of static frames, each potentially varying in its expressive clarity. The point when static expressions are extracted within the dynamic sequence could critically influence their representativeness. If a static frame captures a peak moment of expression, it may inherently convey enough emotional information, rendering additional dynamic cues superfluous. This oversight raises pivotal questions about the optimal conditions under which static expressions, whether at target or non-target moments, are comparable or fall short of the informational value provided by dynamic expressions in terms of recognisability.

Third, past methodologies in static and dynamic facial expression recognition typically tasked participants with classifying expressions under varied formats (i.e., static vs dynamic) without delving into the finer details of factors that specifically enhance recognisability. While this strategy seemingly ascertains the existence (and counterevidence) of dynamic advantage, it falls short of elucidating the features that make an expression recognisable. Featural parameters like prototypicality, intensity, and ambiguity have been identified as influential factors in recognition accuracy. However, their examination has often been hampered by a lack of consistent measures across studies. This inconsistency poses a challenge in drawing definitive conclusions about how each parameter individually, or in combination, contributes to the accuracy of static and dynamic expression recognition. In this sense, a more refined approach is needed, one that incorporates standardised definitions and methodologies for assessing these featural parameters.

Fourth, earlier research on dynamic advantage has predominantly been focused on human perceptual analysis. With rapidly increasing interest in AFEA analysis, such focus leaves a significant gap in understanding how motion impacts machine-based recognition systems. Although several investigations have been made showing that machines exhibit reduced accuracy in recognising dynamic expressions, these studies do not directly compare the performance of dynamic expressions against static snapshots derived from the same video sequences. Specifically, it remains to be determined whether machines, like humans, exhibit a dynamic advantage in recognising facial expressions and, if so, under what conditions this advantage is most pronounced.

Fifth, previous studies have yielded inconsistent results regarding the comparative performance of human versus machine recognition, with some research indicating humans outperform machines, while other studies suggest the opposite. This inconsistency points to the presence of specific conditions under which humans excel in recognising expressions and scenarios where machines demonstrate superior performance, especially depending on their dynamic properties and featural characteristics (i.e., degrees of parameters including prototypicality, intensity, ambiguity etc.). This dichotomy requires further explorations of the conditions that differently affect human and machine recognition capabilities.

Sixth, with the burgeoning interest in achieving greater ecological validity in facial expression research, the availability of spontaneous facial expression databases has significantly increased. However, there appears to be a lack of a systematic review concerning these resources, particularly those cataloguing spontaneous and dynamic facial expressions. Understanding the characteristics of these databases is crucial for several reasons. Firstly, it would provide insights into the range and diversity of spontaneous expressions recorded across different elicitation techniques, demographics and emotional categories. Secondly, a detailed review could elucidate the methodological approaches employed in developing these databases, including the techniques for eliciting spontaneous emotions and the criteria for categorising and annotating expressions. It remains to be seen how these characteristics influence the encoding and decoding of facial expressions.

Seventh, a critical observation in the field of facial expression research is that many spontaneous facial expression databases (FEDBs) have not undergone rigorous empirical testing to assess their reliability and validity, particularly for those featuring basic emotions. Furthermore, there is a conspicuous absence of cross-corpus evaluations. This gap raises important questions about the recognition rates achievable with spontaneous FEDBs, and whether there exists significant variability in these rates. In conjunction with a comprehensive review, an in-depth exploration of the prototypicality, complexity, and ambiguity influencing the recognition rate is also a question that needs to be discussed.

1.7 Overview of the Present Dissertation

The present dissertation aims to systematically investigate the dynamic and spontaneous aspects of facial expressions and their contribution to emotion recognition. Exploring dynamic and spontaneous facial expressions is pivotal in facial expression research as they provide a balanced trade-off between experimental control and authenticity of expressions (Zhang et al., 2014). Specifically, the current work investigates the recognisability of dynamic versus static expressions, with a keen focus on the challenges posed by spontaneous expressions. To this end, this dissertation comprises three experimental studies and one extensive review, each designed to systematically address the key questions highlighted earlier.

Chapter 2 delves into whether movement in facial expression confers a recognition advantage for both human and machine observers. This investigation utilised facial stimuli under three distinct conditions: target static, non-target static, and dynamic. Target static refers to the frame displaying the peak intensity of the target emotion, while non-target static captures frames form the same sequence showing non-target emotions. Dynamic stimuli encompass the entire expression sequence (please find Chapter 2 for details). This selection of stimuli aims to determine the specific circumstances under which facial movement either enhances or fails to aid recognition. The chapter also assess the role of featural parameters – namely prototypicality, ambiguity, and complexity – on the recognisability of facial expressions, aiming to uncover the underlying factors that influence how expressions are recognised. Although posed and spontaneous expressions are conceptually distinct, they were treated as a unified stimulus material in Chapter 2 to focus on the dynamic aspects of facial expressions across all types. By combining posed and spontaneous expressions, the study aims to explore how movement influences facial expression recognition in general, aligning with the thesis' aim to examine the broader role of dynamic information in facial expression recognition. This chapter also evaluates the performance of both human and machine observers to not only validate the dynamic advantage hypothesis across different observers but also to highlight the comparative strengths and weaknesses in their recognition abilities. This comparative analysis offers detailed insights into the similarities and differences in recognition performance between humans and machines.

In Chapter 3, a comprehensive review of existing spontaneous and dynamic facial expression databases is undertaken, showcasing a wide spectrum of available datasets in the field. Most past reviews failed to provide a systematic understanding of existing databases, encompassing a limited number of spontaneous databases. This chapter details the unique characteristics of these databases, including the diversity of emotion categories and demographics, methodologies employed for eliciting spontaneous expressions, the technical frameworks utilised for recording and annotating these expressions, and the accessibility of these databases. By assessing their conceptual, technical, and practical aspects, the review highlights their strengths and identifies gaps within existing databases that may limit their effectiveness. Through the critical discussion, this chapter aims to guide researchers in making informed decisions when selecting a database, thereby enhancing the quality and applicability of facial expression research.

Chapter 4 conducts an empirical evaluation of selected databases from Chapter 3, with a focus on those capturing basic emotions. Utilising cross-corpus evaluation through the AFEA tool AFFDEX, this chapter assesses to what extent spontaneous databases accurately represent emotional states and the AUs that underlie these expressions. This evaluation includes an examination of how well these databases are recognised and the role of AUs in this process.

Considering the featural parameters such as prototypicality, ambiguity and complexity discussed in Chapter 2, their impact on AFEA classification is re-evaluated in this chapter. Additionally, informed by the reviews in Chapter 3, this chapter discusses potential features that may influence machine recognition performance.

Finally, Chapter 5 serves as a general discussion that synthesizes the findings from the preceding chapters. This chapter critically evaluates the theoretical and practical implications of the research conducted and outlines how the present work fills the knowledge gap, thereby encapsulating the contributions and limitations of the present dissertation. It highlights how the exploration of spontaneous and dynamic facial expressions, alongside the assessment of featural parameters and the efficacy of human and machine recognition, enriches our understanding of emotion recognition. Acknowledging the challenges encountered, the chapter also reflects on the scope of the studies and methodological constraints. Additionally, it suggests directions for future research.

CHAPTER 2

Human and Machine Recognition of Dynamic and Static Facial Expressions: Prototypicality, Ambiguity and Complexity

2.1 Introduction

Much of our understanding of facial expressions of emotions has come from studies of static displays typically captured at their peak (Dawel et al., 2021). Static expressions have the advantage that they can be strictly controlled, allowing observers to focus on the key features of interest. Not surprisingly, static images have been widely used in studies exploring the recognition of the basic six emotions (Barrett et al., 2019; Calvo & Nummenmaa, 2016). Due to their lower ecological validity, however, the last two decades have seen increased questioning and criticism of this type of stimulus. Given that facial expressions evolve over time, they are intrinsically dynamic events. Accordingly, facial movement has been shown to aid expression recognition (e.g., Ambadar et al., 2005; Cunningham & Wallraven, 2009; Wehrle et al., 2000) and facilitate the extraction of emotion-relevant content from faces (for reviews, see Dobs et al., 2018; Krumhuber et al., 2013; Krumhuber et al., 2023; Krumhuber & Skora, 2016; Lander et al., 1999), such as expression authenticity (Zloteanu et al., 2018; Krumhuber et al., 2013), naturalness (Sato & Yoshikawa, 2004) and intensity (Biele & Grabowska, 2006, Widen & Russell, 2015). Nonetheless, the effects of movement are not uncontested, with some studies showing little or no benefits of dynamic information (e.g., Fiorentini & Viviani, 2011; Gold et al., 2013; Kamachi et al., 2001; Knight & Johnston, 1997; Lander et al., 1999). The present research aims to compare static versus dynamic expressions in human and machine analysis, thereby exploring the role of featural parameters in emotion recognition.

Despite substantial evidence showing a dynamic advantage, several studies have failed to find the respective benefits of movement. For example, the advantage was found to disappear when identification was already close to perfect, with static stimuli that were highly distinctive in expression (Gold et al., 2013; Kamachi et al., 2001 experiment 2; Kätsyri & Sams, 2008). Also, the effect of movement diminished for static displays presented for more than 1000 ms, which naturally allows for a deeper exploration of the facial stimulus (Bould & Morris, 2008; Kätsyri & Sams, 2008). Finally, movement of the face may not always be necessary for non-degraded or full-intensity expressions (Ambadar et al., 2005; Blais et al., 2017; Bould & Morris, 2008; Tobin et al., 2016). In those cases, static snapshots can be sufficient to recognise emotions. Such counterevidence aligns with arguments proposing a compensatory role of dynamic information, particularly when static cues are inaccessible or insufficient (Ambadar et al., 2005; Atkinson et al., 2004; Ehrlich et al., 2000; Wehrle et al., 2000). For example, dynamic expressions aid the recognition of degraded or distorted stimuli such as in point-light displays, synthetic displays, or ```ffled morphed sequences (e.g., Cunningham & Wallraven, 2009; Dobs et al., 2018; Plouffe-Demers et al., 2019; Wallraven et al., 2008). Similarly, facial movement facilitates the recognition of weakly expressed and non-basic emotions (guilt, shame), which may be more subtle and nuanced in their appearance (Ambadar et al., 2005; Bould & Morris, 2008; Cassidy et al., 2015; Yitzhak et al., 2020).

While attempts have been made to specify the conditions under which the dynamic advantage occurs, it is still unclear when dynamic information matters and when it does not. In most past studies, static displays were used to depict the peak of the target emotion (Bould & Morris, 2008; Gold et al., 2013; Harwood et al., 1999; Kamachi et al., 2001). Such high-intensity features, with their specific shapes and spatial arrangement, may leave little scope for the additional benefits offered by movement. The present research is the first to compare dynamic expressions with static images extracted from various time points of the facial display. In particular, we explore whether peak frames of the target emotion (e.g., the image frame with the highest surprise evidence within a surprise video; see Dente et al., 2017) achieve recognition rates that are similar to dynamic stimuli (e.g., a full-length surprise video) and higher compared to those of non-target emotions (e.g., image frames with the highest anger, fear, disgust, happiness or sadness evidence within a surprise video).

Beyond this comparison of dynamic expressions to automatically extracted single images, the present work examines three key featural parameters and their contribution to emotion recognition. According to Basic Emotion Theory (BET), a small number of fundamental emotions are characterised by *prototypical* patterns of facial actions (Ekman, 1982, 1992). That is, when an emotion is elicited a particular set of action units is triggered by specific muscular movements (Ekman et al., 2002). These unique configurations of prototypical facial displays offer a quick and accurate feature-based categorisation of expressions as they are unambiguously linked with discrete emotion categories (see Calvo & Nummenmaa, 2016; Ekman, 2003). Such categorical distinctiveness makes them perceptually salient, thereby providing a shortcut to emotion recognition (Calvo et al., 2013). Hence, facial displays closely resembling those prototypes are more easily and rapidly classified (Matsumoto et al., 2009; Matsumoto & Hwang, 2014; Young et al., 1997). Conversely, accuracy is thought to drop for non-prototypical expressions (Barrett et al., 2019; Motley & Camden, 1988; Naab & Russell, 2007; Wagner et al., 1986).

While prototypicality crucially functions as a perceptual indicator of emotion category, most of the facial expressions seen in everyday life are likely to be ambiguous, fractional, and/or blended (Calvo et al., 2014; Scherer & Ellgring, 2007). That is, they often convey a mixture of emotions (Halberstadt et al., 2009; Hassin et al., 2013; Parkinson, 2013) or partial versions of configurations, with a great amount of idiosyncrasy and variability beyond uniform configurations of a single emotion (Du et al., 2014; Du & Martinez, 2015). To capture these deviations, it is therefore important to define a second featural parameter.

Ambiguity arises when an expression displays multiple basic emotions (i.e., when facial expressions are categorically unclear), thereby containing contradictory emotional information. Given that classification decisions typically rely on the most distinctive facial features (Calvo et al., 2012; Du et al., 2014; Fiorentini & Viviani, 2009; Tanaka et al., 2012), ambiguous expressions are often subject to misclassification and interpretation biases (Calvo et al., 2012; Ito et al., 2017; Kinchella & Guo, 2021). In turn, recognition accuracy is reduced (Calder et al., 2000b; Neta & Whalen, 2010) because people are perceptually less able to identify several emotions at once (Ito et al., 2017; Kinchella & Guo, 2021). Neuroscientific evidence points toward the role of the amygdala, which encodes not only the intensity but also the categorical ambiguity of an expression (Ito et al., 2017). Since the processing of ambiguous displays requires more cognitive effort, confidence ratings tend to be lower and reaction times are prolonged (Calvo et al., 2012; Wang et al., 2017).

Notwithstanding its importance, empirical evidence regarding expression ambiguity remains elusive mainly due to the lack of a common metric. While some studies define it as the degree of closeness to categorical boundaries (Halberstadt et al., 2009; Kinchella & Guo, 2021; Wang et al., 2017), others conceptualise it as the omission of core emotional cues (Matsumoto & Hwang, 2014). This could be problematic as both definitions indicate different expression characteristics. Additionally, most prior research has manipulated (rather than measured) ambiguity by creating blended, morphed, or composite face stimuli (Nummenmaa, 1988; Calder et al., 2000a, 2000b). Such an approach may result in unnaturalistic displays which are not representative of the type of expressions seen in real-life situations. The present work therefore introduces a new ambiguity measure that is based on the perceived presence of two or more emotions.

Finally, expression *intensity* has been consistently shown to influence emotion recognition. Specifically, intense displays enhance accurate classification and response times (e.g., Ambadar et al., 2005; Jones et al., 2018; Matsumoto et al., 1999, 2002; Palermo & Coltheart, 2004; Young et al., 1997). Also, they lead to higher intensity and confidence ratings (Calder et al.,2000a; Recio et al., 2013), as well as agreement ratings between viewers (Matsumoto et al., 2002; Matsumoto & Hwang, 2014). In contrast, weak expressions tend to be less accurately categorised (although above chance level; Matsumoto & Hwang, 2014) and are subject to greater confusion and uncertainty in emotion judgements (Bould & Morris, 2008; Ichikawa & Yamaguchi, 2014; Matsumoto et al., 2002).

The intensity of expressions may play a crucial role in detecting individual facial configurations because intense expressions often contain diagnostic features of facial prototypes. Expression prototypicality is therefore likely to co-occur with higher expressive intensity. Only a few studies to date have tried to identify their relative influence, suggesting that prototypicality is a more important feature for emotion classification than intensity (Matsumoto & Hwang, 2014; Matsumoto et al., 2002). Nonetheless, both parameters are likely to be confounded as expression intensity usually concerns emotion-relevant facial actions such as those predicted by BET. This makes intensity not representative of the overall expressivity of the face but of the degree of emotion in a facial expression. More intense emotional expressions (especially when they are posed) are likely to be more prototypical and vice versa. In order to conceptualise expression intensity as a measure that is independent from its emotional connotation, we therefore introduce a new metric called ‘complexity’ which captures the intensity of all action units in the face.

While traditional measures of intensity consider the strength of Action Unit (AU) contractions, our measure of 'complexity' quantifies the number of contracting AUs, irrespective of their individual intensities. This approach captures the richness of facial actions without being influenced by the strength of individual AU contractions. Although the probabilities of AU-occurrences are positively (albeit weakly) correlated with their respective intensities (Girard, Cohn, & De la Torre, 2015), complexity provides a comprehensive representation of facial expressivity. This distinction is crucial as facial expressions often involve a mixture of AUs and may not strictly adhere to the prototypical expressions of basic emotions. As such, our measure of complexity offers a unique perspective that is distinct from traditional measures of intensity, which are typically tied to the intensity of emotion-specific AUs.

Quantifying featural parameters necessitates an objective classification of facial expressions, which is a time-consuming and resource-intensive process for human coders (De la Torre & Cohn, 2011). With rapid advances in the field of affective computing, commercial and open-source algorithms for automated facial expression analysis (AFEA) are now widely available (Cohn & Sayette, 2010). These can reliably classify discrete emotions as well as facial actions (Lewinski et al., 2014; Littlewort et al., 2011). Given that most classifiers have been trained based on the theoretical principle proposed by the Facial Action Coding System (FACS, Ekman et al., 2002; Calvo et al., 2018), recognition performance is found to be comparable to human coders (Krumhuber et al., 2021a; Skiendziel et al., 2019) and other physiological measurements (Höfling et al., 2021; Kulke et al., 2020), sometimes even outperforming human raters (Krumhuber et al., 2021b). In most cases, the distinctive appearance of highly standardised expressions benefits the featural analysis by machines (Pantic & Barrett, 2007).

Despite several attempts to validate AFEA, its performance on non-prototypical, subtle, and dynamic expressions needs further attention, with studies showing substantial variation in recognition success. For example, hit rates drop remarkably when an expression moves farther away from basic emotion prototypes (Küntzler et al., 2021; Stöckli et al., 2017). Likewise, machines frequently misclassify expressions that are weak in intensity (Calvo et al., 2018; Küntzler et al., 2021), resulting in recognition rates often lower than those of humans (Mandal et al., 2015; Yitzhak et al., 2017). Since machines rely heavily on physical features of an expression (Del Líbano et al., 2018), less prototypical and more subtle displays of emotion pose a greater challenge for AFEA (Calvo et al., 2018). This is particularly evident for dynamic expressions, which often include large segments of frames with comparatively subtle features. In consequence, machine accuracy has been shown to drop for dynamic compared to static stimuli commonly taken at the peak of the emotional display (Onal Ertugrul et al., 2022; Skiendziel et al., 2019; Stöckli et al., 2017). To date, the role of dynamic information in AFEA is still poorly understood, with performance varying substantially across stimulus conditions (Dupré et al., 2019; Krumhuber et al., 2021b; Yitzhak et al., 2017).

There is suggestive albeit ambivalent evidence for the dynamic advantage with inconclusive findings on why and when facial movements offer benefits for recognition. The present research aims to fill this knowledge gap by investigating the conditions under which dynamic information exerts its facilitative effects on emotion classification. It does so by comparing dynamic stimuli with static peak images that show either the target or non-target emotion (thereafter referred to as 'target-images' and 'non-target images'). In line with previous research on the dynamic advantage (Ambadar et al., 2005; Cunningham & Wallraven, 2009; Wehrle et al., 2000), we predicted superior recognition rates for dynamic displays when compared to static (non-target) images consisting of peak frames that are unreflective of the target emotion. In other words, images taken from any time point of the expression may show minimal benefits, resulting in recognition rates lower than those of dynamic expressions. However, the opposite pattern was expected for static images showing the peak frame of the target emotion (target-images). Given that these are highly distinctive and intense displays of the relevant emotion (Gold et al., 2013; Kamachi et al., 2001; Kätsyri & Sams, 2008), they should be easier to recognise, with performance rates exceeding those of dynamic expressions. To investigate what makes the expression recognisable, we tested the relative contribution of three featural parameters - prototypicality, ambiguity and complexity - to emotion recognition. If the stimuli closely resemble discrete emotion categories as proposed by BET, they should be more prototypical and intense as well as less ambiguous in appearance (Neta & Whalen, 2010; Matsumoto & Hwang, 2014; Jones et al., 2018). Stimuli that show well-recognisable discrete emotions should also be more complex than most other patterns of facial actions. Furthermore, prototypicality and ambiguity as its counterpart should predict emotion recognition, particularly in machines which have often been trained on posed/acted datasets (Pantic & Bartlett, 2007), making them potentially superior to human observers in classification accuracy (Krumhuber et al., 2021b).

Two studies were conducted to test the above hypotheses. Study 2.1 focused on AFEA to compare video (dynamic), target and non-target images (static), and define measures of prototypicality, ambiguity, and complexity. As a way of validating the machine data, we also obtained ratings from human observers on target and non-target images. Study 2.2 focused on human observers with the aim to replicate the findings from the first study with a subset of the stimuli and a larger sample of participants.

2.2 Experiment 1

The first study aimed to test for the dynamic advantage in AFEA, thereby comparing recognition rates of video (dynamic), target and non-target images (static). Human observer ratings were also obtained for target and non-target images as a source of machine validation.

In addition, we explored the relative contribution of prototypicality, ambiguity and complexity to image and video recognition, and whether video recognition can be predicted based on six images that represent the respective peak expressions for the basic emotions.

2.2.1 Method

Figure 2.1. Example of the image selection procedure, showing the highest FACET evidence values for each of the six basic emotions as extracted from a surprise video (A). The surprise image (bottom right) is the target image for the surprise video (as labelled by the dataset authors), whereas the other five mages are non-target images (B).

Stimulus material

162 facial expression (85 females, 77 males) videos portraying the six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) were obtained from Krumhuber and colleagues (2021b). Stimuli originated from a range of databases showcasing a mixture of

emotion elicitation procedures (e.g., instruction to perform an expression, scenario enactment, and emotion-eliciting tasks).

In this study, posed and spontaneous expressions were not distinguished as separate categories but treated collectively as a unified stimulus set. This methodological choice was made to focus on how facial movement affects recognition accuracy, regardless of whether they were posed or spontaneous. While this approach deviates from distinguishing between posed and spontaneous expressions, it builds on previous research (Krumhuber et al., 2020), which has already established differences in recognition rates between the two, providing a foundation for this methodological decision.

Across all emotion categories, the encoders were predominantly white/Caucasian, young to middle-aged adults. Stimuli were presented in a frontal view of the face. The videos had an average duration of 5 seconds and were displayed in colour. Portrayals that lasted longer than 15 seconds were segmented to display the onset, apex, and offset of expression (if applicable), in line with other portrayals. None of the facial stimuli exceeded 10 seconds in duration.

For each video, machine analysis was performed using a commercial software called FACET (Littlewort et al., 2011), which provides estimates for facial expressions of the six basic emotions (anger, disgust, fear, happiness, sadness, surprise) and 20 Action Units (AU1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 18, 20, 23, 24, 25, 26, 28, and 43; Ekman, Friesen & Hager, 2002). Predominantly trained on posed expressions, FACET outputs evidence scores on a frame-by-frame basis, estimating the likelihood that a human observer would code the frame as containing each emotion and action unit. Evidence values are shown on a decimal logarithmic scale centred around zero, with zero indicating 50% probability, negative values indicating that an expression is likely not present, and positive values indicating that an expression is likely to be present (Dente, Küster, Skora, & Krumhuber, 2017).

Within each video, six frames with the highest individual evidence value for the six basic emotions were identified based on the raw FACET output. Extractions were performed automatically via Python and FFmpeg. Among the six frames, one image was indicative of the "target" emotion (e.g., the frame with the highest surprise evidence score from a video that was labelled by the dataset authors as surprise), and five images were indicative of "non-target" emotions (e.g., frames with the highest anger, disgust, fear, happiness, and sadness evidence scores from a surprise video; see Figure 2.1). To this end, a total of 972 static facial images (162 videos × 6 images) were extracted. The number of portrayals was equally balanced across disgust, fear, happiness, and surprise (168 images each), except for anger (144 images) and sadness (156 images) which had fewer portrayals because they were not available in some of the databases. All image stimuli were rendered in colour and had an approximate resolution of 550 × 440 pixels.

To achieve comparability with the confidence ratings provided by human observers, the raw FACET evidence values for each of the six basic emotions and 20 AUs were initially converted into probabilities by using the formula provided in the FACET documentation (iMotions, 2016) and then into confidence odds scores (for a similar procedure see Krumhuber et al., 2021a). Let x_{ijk} represent the evidence value for emotion or AU k in image j from video i. This value can be converted into probability (p_{ijk}) and odds (o_{ijk}) units using Equations 1 and 2, respectively:

$$p_{ijk} = \frac{1}{1 + 10^{-x_{ijk}}} \tag{1}$$

$$o_{ijk} = \frac{1}{1/p_{ijk} - 1} \tag{2}$$

Human observers

Power analysis. A simulation-based power analysis was conducted using the "simr" package in R to determine the required sample size for detecting the effects of stimulus types (target vs non-target) and rater (humans vs machine) on recognition accuracy in a multilevel logistic regression model. The analysis revealed that a sample size of 142 is required to achieve 80% power at an alpha level of 0.05, based on 1000 simulations.

Participants. One hundred and fifty-four participants (76 females), aged between 18-60 years (*M* = 29.78, *SD* = 11.85), volunteered to take part in the study. This sample size was calculated using G*Power to ensure 85% power to detect Participants were recruited face-to-face or online via the departmental subject pool and Prolific Academic’s digital recruitment platform. Participants received course credits or £10 for taking part in the study. All participants were White/Caucasian and identified as British or European and ordinary residents in the UK. Ethical approval was granted by the Department of Experimental Psychology at University College London, UK.

Procedure. To reduce participation time, a subset of 162 facial images portraying the six basic emotions were extracted from the 972 static expression stimuli and were randomly presented. As such, every participant viewed one image from each video. The number of portrayals was balanced across the six emotions. Each facial expression was presented for 15 seconds using the Qualtrics software (Provo, UT). Participants could provide their ratings anytime during or after the 15-second exposure. For each facial stimulus, participants rated the extent (from 0% to 100%) to which each of the six emotions (anger, disgust, fear, happiness, sadness, and surprise) is recognisably expressed in the face. At least one emotion rating per image (greater than 1% for any emotion) had to be given. Participants could respond using multiple sliders (if applicable) to choose the exact confidence levels for each response category. After providing their ratings, participants had to click the “next” button to move on to the next stimulus, with no imposed time pressure.

Parameters

Prototypicality. We defined expression "prototypicality" as the degree to which the combination of AUs estimated to be present in a facial expression matches the prototypical facial expression configuration proposed by Basic Emotion Theory (Ekman, 1992). The FACS manual (Ekman et al., 2002) was used to define the full prototype and major variants of each basic emotion. According to the FACS manual, the full prototypes indicate the complete set of AUs associated with a target emotion (e.g., e.g., AU1+2+5+26 for surprise), while major variants refer to commonly observed deviations from the full prototypes that still convey the target emotion (e.g., AU1+2+5 for surprise). These variants typically involve the omission of one or more AUs that are less critical to the recognition of the emotion. The odds of FACET AU scores for the target emotion were summed up and weighted by a factor of 1 (full prototype) or 0.75 (major variant). This resulted in an estimated prototypicality score for each image, with higher scores indicating greater prototypicality of the expressed emotion (for a similar procedure, see Krumhuber et al., 2021a). Prototypicality for emotion k in image j from video i was calculated as:

$$PRO_{ijk} = \sum_{l=1}^{v} O_{ijkl} w_{kl} \tag{3}$$

where O_{ijkl} is the FACET-estimated odds that image j from video i contains prototype l from emotion k and w_{kl} is the weight of prototype l from emotion k (i.e., 1 if a full prototype and 0.75 if a major variant). To calculate the prototypicality for emotion k in video i (across all m images), we averaged the prototypicality for that emotion across all m images (i.e., $m = 6$).

$$PRO_{ik} = \frac{1}{m} \sum_{j=1}^{m} PRO_{ijk} \tag{4}$$

Ambiguity. We defined expression "ambiguity" as the degree to which the facial expression is classified as containing multiple basic emotions, which makes the expression categorically unclear (Kinchella & Guo, 2021). To this end, we used normalised entropy as a metric to represent the amount of uncertainty in emotion classification for each image (Shannon, 1948). Entropy is high when multiple emotions have high estimated probabilities and low when only a single emotion has a high estimated probability. The ambiguity of image j from video i (in terms of the q different emotions) was calculated using the following equation:

$$AMB_{ij} = -\sum_{k=1}^{q} \frac{p_{ijk}}{\log(q)} \tag{5}$$

where p_{ijk} is the FACET-estimated probability that image j from video i contains emotion k. (Note that the logarithm bases do not matter due to their division.) To calculate the ambiguity for video i (across all m images), we averaged the ambiguity across all m images (i.e., $m = 6$).

$$AMB_i = \frac{1}{m}\sum_{j=1}^{m} AMB_{ij} \tag{6}$$

Complexity. We defined expression "complexity" as the average probability of evidence across all 20 FACET AU estimates in each image. This resulted in an estimated complexity score for each image, with higher scores indicating more complex expressions (with evidence of more AUs present). This complexity measure therefore differs from other conceptualisations of "intensity" by taking all FACET AUs into account and using their probability of occurrence rather than their estimated intensity. The complexity for image j from video i was calculated as:

$$COM_{ij} = \frac{1}{m}\sum_{l=1}^{f} P_{ijl} \tag{7}$$

where p_{ijl} is the FACET-estimated probability that image j from video i contains AU l and $f = 20$ (i.e., the superset of all estimated AUs). To calculate the complexity for video i (across all m images), we averaged the complexity across all m images (i.e., $m = 6$).

$$COM_i = \frac{1}{m}\sum_{j=1}^{m} COM_{ij} \tag{8}$$

Data preparation

FACET recognition accuracy for both video and image was calculated by determining whether the emotion with the highest recognition score matched the target emotion label given by the database authors. As FACET is an algorithm-based classifier that provides the same values across trials, recognition accuracy was binary in the form of either 0 (incorrect) or 1 (correct). To compare FACET and human performance, the recognition scores by human observers were also converted into this binary format as a function of whether the majority (> 50%) of participants correctly recognised the target emotion of video. Specifically, human observers were randomly assigned to different sets of images (162 images out of 972), meaning not all participants rated every images. For each image, the number of participants who observed it was counted, and if the majority of these participants correctly recognised the emotion (i.e., highest rating for the target emotion of video). This approach aligns the representative human rating across observers with the consistent output provide by the machine.

2.2.2 Results

6-images as predictor of video recognition

We first tested whether the emotion classification accuracy of the video can be predicted from the recognition of the 6 extracted images. For this, a multilevel logistic regression model predicting video-level emotion classification accuracy (by FACET) was estimated with a random intercept for each video and a fixed slope for the sum of correct image-level emotion classification accuracy (per video). The results revealed a significant main effect ($\exp(\beta) = 2.86$, Wald = 35.63, $p < .001$, exp(95%CI) [2.10, 4.22]), indicating that the odds of correct video-level emotion classification increased by 186% for each additional correct image-level emotion classification.

Video vs target image vs non-target images

In general, recognition accuracy varied across stimulus types and raters. Target images achieved highest recognition rates, with humans reaching 65.4% accuracy and FACET at 82.7%, likely due to their clear representation of the target emotion at peak moment. In contrast, non-target expressions, designed to signal competing emotions, presented significant ambiguity. This was reflected in lower recognition accuracy for both humans (34.4%) and FACET (42.2%), highlighting the inherent challenges in classifying these images. The consistently low performance of non-target images suggests that these expressions contained subtle, overlapping or multiple emotional cues, making them difficult to categorise. For videos, FACET achieved an intermediate accuracy rate of 65.4%, possibly because the dynamic nature of the stimuli included subtle, non-expressive moments that diluted the overall recognition process compared to the more consistent peak emotional representation in target images.

To statistically examine whether recognition accuracy differs as a function of stimulus type (video vs. target image vs. non-target images), a multilevel logistic regression analysis with a random intercept by video was conducted on the FACET accuracy data. The odds of correct emotion classification were significantly higher for target images than for non-target images (exp(β) = 40.66, Wald = 99.48, $p < .001$, exp(95%CI) [20.40, 88.10] and were significantly higher for the video (exp(β) = 6.47, Wald = 48.37, $p < .001$ exp(95%CI) [3.87, 11.12]) than for non-target images (see Figure 2.2). Interestingly, the odds of correct emotion classification were significantly lower for the video than for target images (exp(β) = 0.16, Wald = 21.98, $p < .001$ exp(95%CI) [0.07, 0.34]). As such, the dynamic advantage only occurred for non-target images, but not target images. Overall, recognition accuracy was highest for the target image, followed by the video and non-target images (see Figure 2.2).

We conducted another multilevel logistic regression analysis with stimulus type (target vs. non-target images) and rater type (FACET vs. human observers) as predictors and with a random intercept for each video. The results revealed significant main effects of stimulus type, (exp(β) = 7.05, Wald = 74.47, p < .001 exp(95%CI) [4.52, 10.98]) and rater type (exp(β) = 1.65, Wald = 16.16, p < .001 exp(95%CI) [1.29, 2.11]), as well as a significant interaction between the two (exp(β) = 2.38, Wald = 6.23, p = .035 95%CI [1.20, 4.70]). For both FACET and humans, target images were better recognised than non-target images (*ps* < .001). Thus, the target peak image seemed to be a better exemplar of the expression in human and machine analysis. Results also revealed that recognition accuracy of FACET was significantly higher than that of humans for both target and non-target images (*ps* < .001). Additionally, the interaction effect showed that the difference in accuracy between machine and human observes was greater for target images than for non-target images.

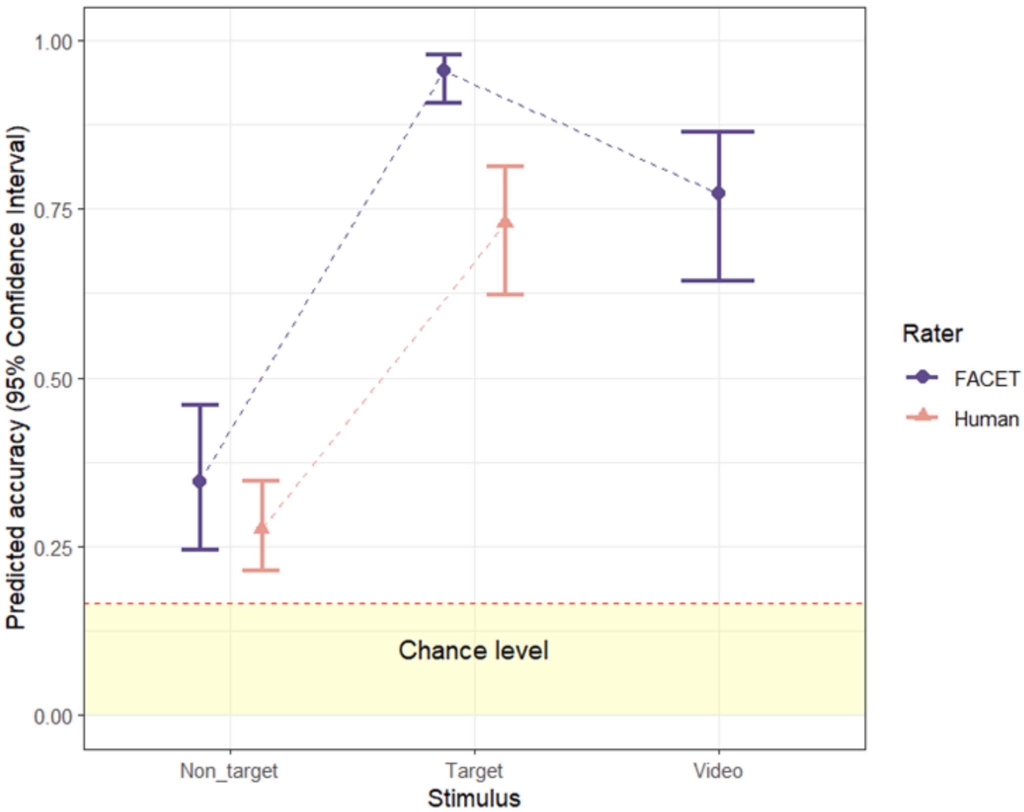


Figure 2.2. FACET and human recognition accuracy for video, target- and non-target images.
Note. Error bars represent the upper and lower bounds of the 95% confidence interval. The dashed red line indicates a 1/6 conservative chance level (Krumhuber et al., 2020)

Prototypicality, ambiguity, and complexity of expression

To investigate what makes the expression recognisable, separate Welch's t-tests were conducted to compare stimulus types (target vs non-target images) in terms of prototypicality, ambiguity, and complexity. As expected, target images were significantly more prototypical (M_{target} = 64.08, *SD* = 34.11 vs $M_{non\text{-}target}$ = 37.18, *SD* = 33.16), *t*(226.03) = 9.21, *p* < .001, *d* = 0.81), less ambiguous (M_{target} = 29.79, *SD* = 25.60 vs. $M_{non\text{-}target}$ = 46.99, *SD* = 22.20), *t*(212.16) = 5.18, *p* < .001, *d* = 0.75), and more complex (M_{target} = 28.22, *SD* = 7.76 vs. $M_{non\text{-}target}$ = 24.60, *SD* = 9.72), *t*(272.75) = 5.18, *p* < .001, *d* = 0.38, than non-target images.

To ensure that featural parameters do not violate the issue of multicollinearity, we tested the bivariate correlations across scaled predictors prior to building the models. No pair showed a high correlation (all < 0.4), suggesting no issue with multicollinearity. Additionally, we checked the variance inflation factors (VIFs) to further confirm the absence of multicollinearity. VIFs across the three parameters were close to 1 (ranging between 1.01 and 1.07) for both images and videos, indicating no multicollinearity concerns. Typically, VIFs greater than 5 or 10 are taken as indicative of problematic collinearity, which was not the case in our data.

Next, we examined the relative contribution of each parameter to emotion classification accuracy. For this, a multilevel logistic regression model predicting each image's classification accuracy was estimated with random intercepts for each video and fixed slopes for prototypicality, ambiguity, complexity, rater type, and the interaction of rater type with the other three measures. Results revealed a significant main effect of prototypicality (exp(β) = 1.05, Wald = 135.06, *p* < .001, exp(95%CI) [1.04, 1.05]), ambiguity (exp(β) = 0.99, Wald = 9.63, *p* = .002, exp(95%CI) [0.98, 0.99]), and complexity (exp(β) = 1.04, Wald = 8.36, *p* = .004, exp(95%CI) [1.01, 1.06]). All three parameters showed a significant interaction effect with rater type (*ps* < .01). Post-hoc tests revealed that the effects of prototypicality (exp(β) = 1.02, Wald = 32.14, *p* < .001, exp(95%CI) [1.01, 1.03]) and ambiguity (exp(β) = 1.01, Wald = 7.90, *p* = .005, exp(95%CI) [1.00, 1.02]) were significantly greater for FACET than for humans. In contrast, the effect of complexity (exp(β) = 1.03, Wald = 10.91, p < .001, exp(95%CI) [0.95, 0.99]) was significantly greater for humans than FACET (see Figure 2.3 & Table 2.1).

Finally, we explored the partial association of each parameter with video-level recognition accuracy. For this, a multilevel logistic regression model predicting video-level emotion classification accuracy (by FACET) was estimated with random intercepts for each source database and fixed slopes for video-level prototypicality, ambiguity, and complexity. Results revealed a significant main effect of prototypicality (exp(β) = 1.01, Wald = 7.54, p = .006, exp(95%CI) [1.00, 1.02]), and ambiguity (exp(β) = 0.97, Wald = 26.12, p < .001, exp(95%CI) [0.96, 0.98]). The main effect of complexity was marginally significant (exp(β) = 0.98, Wald = 3.81, p = 0.051, exp(95%CI) [0.95, 1.00]). In general, the odds of recognition accuracy increased by 1% for each unit increase in prototypicality, while it decreased by 3% for each unit increase in ambiguity (see Table 2.2).

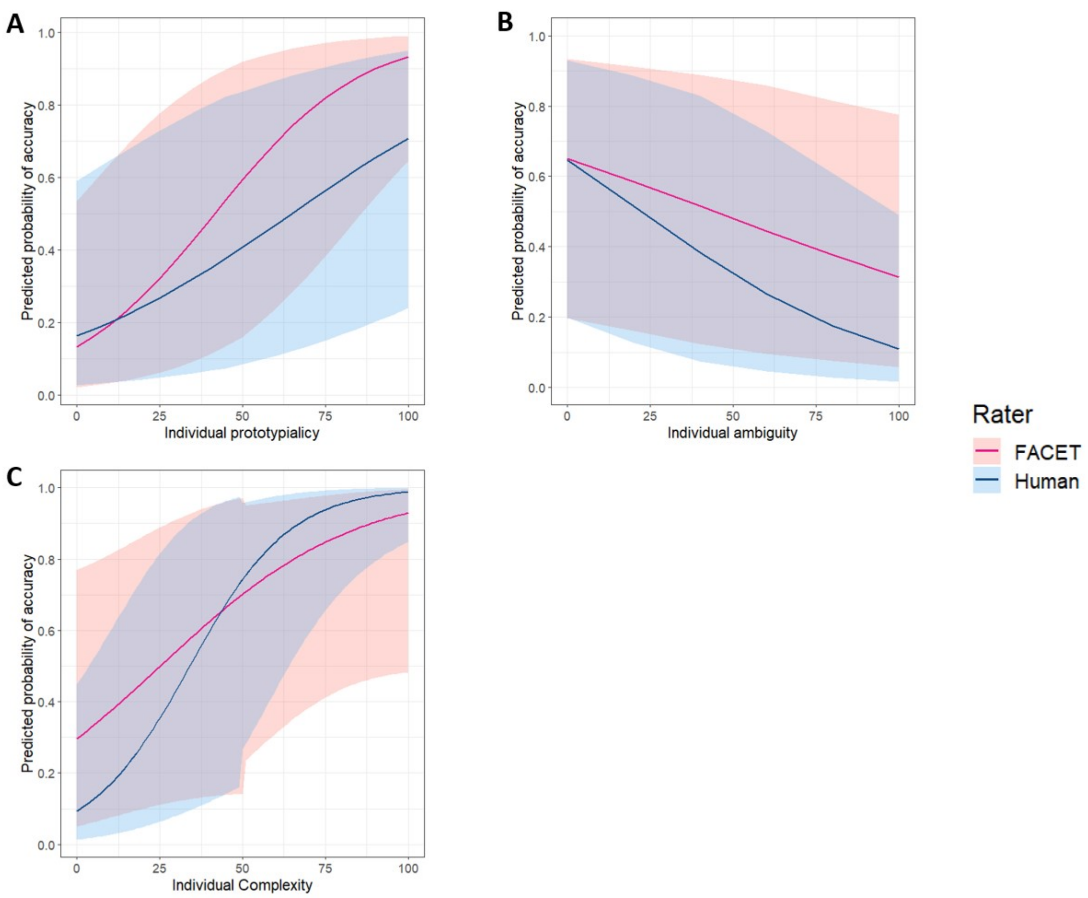


Figure 2.3. predicted power of prototypicality, ambiguity and complexity for image recognition accuracy in FACET and humans. *Note.* Regression line indicates the relationship between image recognition accuracy (red:

FACET, blue: Human) and individual scores of (A) prototypicality, (B) ambiguity, and (C) complexity. The line shades represent upper and lower bounds 95% confidence interval at each predictor score point.

Predictor	exp(b)	Wald	L95%CI	H95%CI	p
Prototypicality	1.05	135.06	1.04	1.05	>.001***
Ambiguity	0.99	9.63	0.98	0.99	.002**
Complexity	1.04	8.36	1.01	1.06	.004**
Prototypicality:Rater	0.99	23.06	0.98	0.99	>.001***
Ambiguity:Rater	0.99	6.70	0.98	1.00	.010*
Complexity:Rater	1.02	8.68	1.01	1.04	.003**

Table 2.1. Model estimates for FACET and human image recognition accuracy, showing main and interaction effect estimates in logits, upper and lower bounds of exponentiated 95% confidence intervals, and significance of each predictor.

Predictor	exp(b)	Wald	L95%CI	H95%CI	p
Prototypicality	1.01	7.54	1.00	1.02	.006**
Ambiguity	0.97	26.12	0.96	0.98	>.001***
Complexity	0.98	3.81	0.95	1.00	.051

Table 2.2. Model estimates for FACET video recognition accuracy, showing main effect estimates in logits, upper and lower bounds of exponentiated 95% confidence intervals, and significance of each predictor.

Predictor	exp(b)	Wald	L95%CI	H95%CI	p
Prototypicality	1.01	3.14	1.00	1.03	.076
Ambiguity	0.96	7.60	0.94	0.99	.006**
Complexity	1.16	13.63	1.07	1.25	>.001***

Table 2.3. Model estimates for FACET image recognition accuracy, showing main effect estimates in logits, upper and lower bounds of exponentiated 95% confidence intervals, and significance of each predictor.

2.2.3 Discussion

The results of the first study demonstrated considerable variation in recognition accuracy as a function of stimulus type. On average, recognition accuracy was highest for target images, followed by the video and non-target images. In accordance with previous findings (Ambadar et al., 2005; Bould & Morris, 2008; Gepner et al., 2001; Harwood et al., 1999), movement (in the form of videos) aided emotion classification over non-target images that were generally less prototypical and complex but more ambiguous than target images. Such a dynamic advantage was absent in comparison to static images which showed the expression at its peak intensity of the target emotion. Additionally, accurate recognition of the video was successfully predicted by the six images, pointing towards the usefulness of single images in video prediction.

Regarding featural parameters, higher prototypicality and complexity but lower ambiguity encouraged correct recognition in both humans and machines. While prototypicality and ambiguity were better predictors of machine performance, complexity (as a reflection of overall expressivity) was more effective in predicting human accuracy. These findings are in line with prior works suggesting that AFEA relies heavily on specific facial configurations (Krumhuber et al., 2021a; Zeng et al., 2009) due to its training on a few – often posed/acted – datasets (Pantic & Bartlett, 2007) while humans tend to process expressions more holistically including all facial actions (Calvo et al., 2012). When comparing human and machine performance, a similar pattern was observed in the sense that accuracy decreased for non-target (vs target) images. Interestingly, the machine outperformed humans on both types of static stimuli, thereby extending previous findings on target emotion recognition (Krumhuber et al., 2021a). With the absence of video ratings from human observers, however, no firm conclusion can be drawn regarding the role of movement versus static information in human emotion classification. To rectify this shortcoming, a second study was conducted in which human observers rated all three types of stimuli: video (dynamic), target and non-target images (static).

2.3 Experiment 2

The second study aimed to replicate and extend the findings of the first study with solely human observers, thereby using a subset of the stimuli and a larger sample of participants. For this purpose, we obtained human ratings of three stimulus types (video, target and non-target images) and analysed the relative contribution of prototypicality, ambiguity and complexity to emotion classification. We further explored the extent to which video recognition can be predicted based on performance for single images.

2.3.1 Method

Stimulus material

To select a diverse set of stimuli, 8 videos per emotion were randomly selected from Study 2.1. This resulted in a total of 48 videos (8 videos × 6 emotions) and 288 images (48 videos × 6 images). To ensure balanced representation, an equal number of encoders from each gender were selected (24 females, 24 males). Other demographics, such as age and race, were similar to study 2.1, with the majority of encoders being white/Caucasian, young to middle-aged adults. Each emotion was portrayed by encoders with similar demographics. The size of the image and video stimuli was approximately 550 × 440 pixels.

Human observers

Power analysis. A simulation-based power analysis was conducted using the "simr" package in R to determine the required sample size for detecting the effects of stimulus types (target vs non-target vs video) and rater (humans vs machine) on recognition accuracy in a multilevel logistic regression model. The analysis revealed that a sample size of 288 is required to achieve 80% power at an alpha level of 0.05, based on 1000 simulations.

Participants. Three hundred and three participants (141 females), aged between 18-60 years (M = 35.99, SD = 10.84), volunteered to take part in the study. Participants were recruited online via a digital recruitment platform (Academic Prolific). Participants were compensated £7 for taking part in the study. All participants were White/Caucasian who identified themselves as British or European and were ordinary residents in the UK. Ethical approval was granted by the Department of Experimental Psychology at University College London, UK.

Procedure. The experiment was programmed using the Qualtrics software (Provo, UT). In the first block, participants were randomly presented with one of the six images extracted from each video, yielding 48 images showing each of the six basic emotions. In the second block, 48 videos displaying each of the six basic emotions in dynamic form were presented in a randomised order. Measures of emotion recognition were the same as in Study 2.1.

2.3.2 Results

6-images as predictor of video recognition

We first tested whether the 6 images can predict how well the video is recognised. For this, a multilevel logistic regression model predicting video-level emotion classification accuracy (by humans) was estimated with a random intercept for each video and a fixed slope for the sum of correct image-level emotion classification accuracy (per video). The results revealed a significant main effect (exp(β) = 2.43, Wald = 11.99, $p < .001$, exp(95% CI) [1.47, 4.03]), indicating that the odds of correct video emotion classification increased by143% for each additional correctly classified image.

Video vs target image vs non-target images

Again, recognition accuracy varied across stimulus types and raters, showing a similar pattern to experiment 1. Target images achieved the highest recognition rates, with humans at 72.9% and FACET at 93.8%. Non-target images resulted in lower accuracy, particularly for humans (48.8%) compared to FACET (76.2%). Video stimuli showed moderate accuracy rates, with humans at 70.8% and FACET at 89.6%. These findings reflect the challenges in recognising non-target images, consistent with the results of experiment 1.

To examine whether recognition accuracy differs as a function of stimulus type (video vs. target image vs. non-target images) and rater type (FACET vs. human observers), a multilevel logistic regression analysis with a random intercept by video was conducted on the accuracy data. The results revealed significant main effects of stimulus type. The odds of correct emotion classification were significantly higher for target images than for non-target images (exp(β) = 5.89, Wald = 7.67, p = .006, exp(95%CI) [1.68, 20.63]), and significantly higher for the video than for non-target images (exp(β) = 3.19, Wald = 4.87, p = .027 exp(95%CI) [1.14, 8.95]). However, the odds of correct emotion classification were not significantly different between the target image and the video (exp(β) = 0.54, Wald = 0.60, p = .439 exp(95%CI) [0.11, 2.56]). Similar to Study 2.1, the dynamic advantage only occurred when the video was compared to non-target images, but not target images.

The results also reveal a significant main effect of rater type (exp(β) = 7.33, Wald = 7.74, p = .005 exp(95%CI) [1.80, 29.82]. Across all three stimulus types, FACET consistently outperformed human observers. The model did not show a significant interaction effect between stimulus and rater types (*ps* > .05), indicating that the differences in accuracy between FACET and human were relatively consistent across stimulus types (see Figure 2.4).

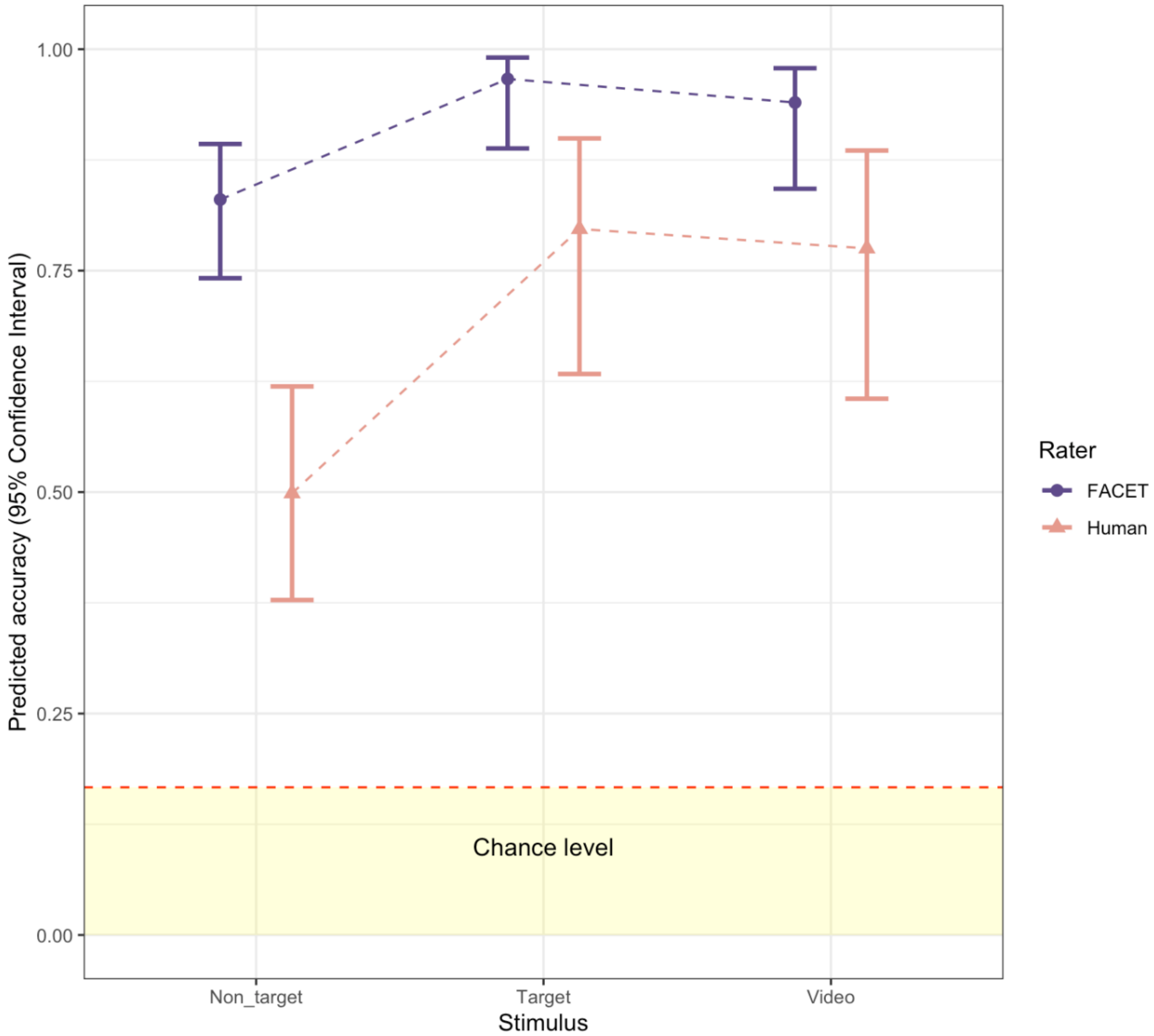


Figure 2.4. Human and machine recognition accuracy for video, target- and non-target images.

Note. Error bars represent upper and lower 95% confidence interval. Dashed red line indicates 1/6 conservative chance level (Krumhuber et al., 2020).

Prototypicality, ambiguity, and complexity of expression

Using the machine data, we assessed prototypicality, ambiguity, and complexity of the stimulus types (target and non-target images). Overall, Welch's t-tests showed that target images were significantly more prototypical (M_{target} = 80.82, *SD* = 27.17 vs $M_{non\text{-}target}$ = 56.36, *SD* = 32.84), $t(77.18) = 5.49$, $p < .001$, $d = 0.76$,) less ambiguous (M_{target} = 14.53, *SD* = 14.25 vs $M_{non\text{-}target}$ = 33.71, *SD* = 21.22), $t(94.25) = -7.76$, $p < .001$, $d = 0.95$), and more complex (M_{target} = 27.38, *SD* = 6.73 vs $M_{non\text{-}target}$ = 22.36, *SD* = 8.96), $t(84.17) = 4.44$, $p < .001$, $d = 0.58$) than non-target images. As such, the subset of 48 stimuli was sufficiently representative of the larger sample analysed in Study 2.1.

Next, we examined the partial contribution of each parameter to human and machine emotion classification accuracy of images. For this, a multilevel logistic regression model predicting each image's classification accuracy was estimated with random intercepts for each video and fixed slopes for prototypicality, ambiguity, and complexity, and the interaction of rater type with the other three measures. Results revealed a significant main effect of prototypicality (exp(β) = 1.03, Wald = 24.23, $p < .001$, exp(95%CI) [1.02, 1.05]), ambiguity (exp(β) = 0.98, Wald = 6.81, $p = .009$, exp(95%CI) [0.96, 0.99]), and complexity (exp(β) = 1.06, Wald = 6.22, $p = .013$, exp(95%CI) [1.01, 1.11]). In general, the odds of recognition accuracy increased by 3% and 6% for a unit increase in prototypicality and complexity respectively, while they decreased by 2% for a unit increase in ambiguity. Among parameters, only prototypicality showed a significant interaction with raters (exp(β) = 0.98, Wald = 9.72, $p = .002$, exp(95%CI) [0.96, 0.99]). Post-hoc tests revealed that the effects of prototypicality was significantly lower for humans than for FACET.

Finally, we explored the predictive power of each parameter for human and machine video recognition. For this, a multilevel logistic regression model predicting human video-level emotion classification accuracy was developed with random intercepts for each source database and fixed slopes for video-level prototypicality, ambiguity, and complexity and their interaction with rater. The results revealed a significant main effect of ambiguity (exp(β) = 0.96, Wald = 4.76, p = .029, exp(95%CI) [0.92, 0.99]), indicating that the odds of recognition accuracy decreased by 4% for each unit increase in ambiguity. The main effects of prototypicality (exp(β) = 0.99, Wald = 0.35, p = .552, exp(95%CI) [0.97, 1.02]) and complexity (exp(β) = 1.03, Wald = 0.42, p = .515, exp(95%CI) [0.94, 1.12]) were not significant. The model did not show significant interaction between predictors (*ps* > .05) (see Table 2.4).

Predictor	exp(b)	Wald	L95%CI	H95%CI	P
Prototypicality	0.99	0.35	0.97	1.02	0.552
Ambiguity	0.96	4.76	0.92	0.99	0.029*
Complexity	1.03	0.42	0.94	1.12	0.515
Prototypicality:Rater	0.98	0.28	0.93	1.04	0.597
Ambiguity:Rater	0.94	2.42	0.86	1.02	0.120
Complexity:Rater	1.05	0.30	0.88	1.25	0.585

Table 2.4. Model estimates for human video recognition accuracy, showing main effect estimates in logits, upper and lower bounds of exponentiated 95% confidence intervals, and significance of each predictor

2.3.3 Discussion

Similar to the first study, there were substantial differences in emotion recognition accuracy across stimulus types. While target images and videos were similarly well recognised, accuracy for non-target images was significantly reduced. As such, movement may function as a facilitative factor particularly when static information fails to convey the target peak emotion. Correct classification of the extracted images was predictive of human recognition performance for the full video, suggesting that single images may be useful for conveying a given expression. As in Study 2.1, higher complexity but lower ambiguity contributed to classification accuracy. Furthermore, the effect of prototypicality was only marginally significant, with facial expressions likely to be processed by humans more holistically and in an integrated fashion (Calder et al., 2000b, Calvo et al., 2012). Together, these findings suggest that categorical ambiguity and complexity (overall expressivity) play an important role in human emotion recognition which seems to rely on features other than prototypicality.

2.4 General Discussion

Past research has been inconclusive with regard to the conditions in which dynamic information matters. In two studies, dynamic expressions were more accurately classified than non-target images, with temporal information aiding emotion recognition. The results partially replicate previous findings on the dynamic advantage (Ambadar et al., 2005; Bould & Morris, 2008; Cassidy et al., 2015), showing that facial expressions are temporally structured in a way that is both meaningful and beneficial to observers. However, these movement-related benefits disappeared in comparison to static peak expressions of the target emotion. Insofar as target images represented static snapshots of a fully expressed emotion, they may have provided sufficient information for emotion classification. This was not the case for non-target images captured at various time points and indicative of peak expressions other than the target emotion. Together, these findings suggest a compensatory role of dynamic information, facilitating emotion recognition when static emotional cues are suboptimal or insufficient (Atkinson et al., 2004; Ehrlich et al., 2000; Wehrle et al., 2000).

Despite both human and machine recognition achieving higher accuracy for target images compared to videos, this does not necessarily imply that target images provide more information. Dynamic expressions offer a wealth of spatiotemporal information includes the progression and transitions between different facial actions (speed, rhythm and velocity; Krumhuber et at., 2013). Such information is crucial for perceptual judgement of expression authenticity and trustworthiness (Krumhuber et al., 2007; Zloteanu et al., 2018). However, dynamic expressions also include larger segments where the face may be less expressive or even neutral. For FACET, which operates by analysing each frame individually and then aggregating these frame-by-frame results, these subtle or non-expressive moments can dilute the overall confidence in emotion classification. In contrast, a target image represents the apex of the emotional display, offers a clear, consistent cue. This consistency over time in the target image allows for deeper exploration of the expressive moment, particularly for systems like FACET that might struggle with integrating varying levels of prototypicality and intensity across multiple frames.

Furthermore, the classification process for dynamic expressions requires a more complex analysis by machines, as it involves not only recognising the spatial cues but also determining how the expression changes over time (Sariyandi, Gunes & Cavallaro, 2017). This added confusion can lead to more errors, especially when the system encounters frames that are ambiguous or do not clearly represent the target emotion (Kuntzler et al., 2021; Yitzhak et al., 2017). On the other hand, the static nature of target images eliminates this complexity, allowing the system to focus solely on the clear, prototypical expression captured at a highly expressive moment. Therefore, the higher accuracy observed for target images reflects the advantage of presenting a single, highly expressive and consistent moment in time, which aligns more closely with the operational strengths of frame-by-frame analysis systems.

In support of this notion, non-target images were found to be less prototypical and complex, as well as more ambiguous. Similar to past research (Matsumoto et al., 2009; Matsumoto & Hwang, 2014) prototypicality played a crucial role, with expressions that more closely resemble BET predictions (Ekman et al., 2002) enhancing recognition. This applied particularly to the machine due to its history of training on posed/stylised expressions. For human observers, complexity was more important for emotion recognition. Consistent with previous work (Jones et al., 2018; Matsumoto et al., 2002), expression intensity (as measured by our new complexity metric) notably improved performance. Here, we showed for the first time that complexity can explain recognition performance without having to confound intensity with prototypicality and its BET-based assumptions. In the future, this allows for subtle expressions to be coded separately from non-prototypical expressions as both metrics tap into different characteristics. As predicted, ambiguous expressions were often subject to misclassification, with the simultaneous presentation of contradictory emotional cues increasing human and machine difficulty in recognising discrete emotions (Calder et al., 2000b; Neta & Whalen, 2010). While previous studies mainly relied on techniques to create ambiguous stimuli, the present research introduced a new metric for *quantifying* ambiguity. This metric can be applied to any emotion rating data in future research that provides a probability for a closed set of emotion categories.

Machine recognition exceeded human performance for both types of static images. The finding extends prior work (Krumhuber et al., 2021a, 2021b) by demonstrating a machine advantage for classifying expressions at the peak of the target emotion as well as other time points of the facial display (non-target images). In contrast to earlier studies showing a reduction in machine performance for low-intensity expressions (Calvo et al. 2018, Küntzler et al. 2021), we found that non-target images were better recognised by the machine than human observers despite their substantially lower prototypicality, greater ambiguity, and lower complexity. It should be noted, however, that stimuli were drawn from standardized datasets, which may benefit machine analysis (Pantic & Barret, 2007). Furthermore, our extraction procedure was designed to select peak images for other emotions to examine the underlying featural parameters. Therefore, the non-target images primarily differed from the target images in ambiguity and prototypicality, and less in complexity or intensity. Here, future work could systematically manipulate all three parameters to better understand their impact on human and machine recognition performance.

There is no doubt that video rating studies are costly and resource-intensive. Automatic peak extraction may be an economic choice for addressing certain research questions by reducing the required presentation time of each stimulus. After accounting for potential fatigue effects in our human sample, we could present three times as many image stimuli in Study 2.1 than video stimuli in Study 2.2. This was the case even though our videos were relatively short and standardised. As is now widely recognised in the field, there is a need to study more ecological behaviours such as those observed in the wild (Küster et al., 2020, 2022; Krumhuber et al., 2017). However, naturalistic stimuli tend to be considerably longer, less standardised, and less well-annotated (Benitez-Quiroz et al., 2016; Cowie et al., 2005; Girard et al., 2015). Here, algorithmic approaches could help by allowing thin slices of stimulus materials to be presented to participants. These could be static peak images or frame sequences extracted on the basis of machine parameters. As such, AFEA may provide a valuable tool to systematically define and extract appropriate research materials from otherwise seemingly “unwieldy” naturalistic datasets.

While present methods for identifying peak images vary between studies (Onal Ertugrul et al., 2022; Skiendziel et al., 2019; Stöckli et al., 2017), both expert-based and algorithmic selection may be subject to biases (e.g., human experts might discard images that appear too ambiguous due to the presence of additional action units). Here, an algorithmic may be more objective because each action unit is assessed separately. However, algorithmic peak selection may suffer from other types of biases. For example, variable lighting during a video might result in the machine missing certain peaks that a trained human expert could have recognised.

Thus, although algorithmic approaches might be particularly helpful for studying naturalistic datasets, further research will still be required to assess the reliability of these tools for more “in the wild” recordings.

The present work has taken the first steps to blend AFEA with psychological research on human emotion recognition. The results extend previous work by introducing complexity as a novel metric of intensity that is largely decoupled from prototypicality and BET. We argue that featural parameters such as prototypicality, ambiguity, and complexity reveal important new insights into human vs. machine differences. Specifically, complexity is a defining feature for humans who are likely to process expressions in a more integrated fashion. In contrast, machine algorithms such as FACET still mainly rely on prototypicality, achieving better performance on peak images than videos, especially if those are highly prototypical and complex, and low in ambiguity. The present research helps inform psychological studies into the mechanisms that underlie the dynamic advantage. Closing this knowledge might be particularly fruitful for future work on dynamic spontaneous expressions.

CHAPTER 3

A Systematic Review of Spontaneous and Dynamic Databases for Facial Expression Research

3.1 Introduction

Most faces we encounter and interact with are inevitably spontaneous, containing dynamic movements. The facial patterns and dynamic quality make spontaneous facial expressions (FE) a communicative source in conveying various emotional meanings (Schmidt, Cohn & Tian, 2003). Many review papers have highlighted the role of spontaneous FE in social-cognitive and emotional processes, such as emotion recognition, empathy, and perception (Dawel et al., 2021; Fabricio et al., 2022). Yet, much of the past research has predominantly relied on standardised images – typically captured at expression apex (Dawel et al., 2021). Such over-reliance may stem from methodological challenges in stimulus generation and the limited number of publicly available datasets (Kanade, Cohn & Tian, 2000; Mavadati et al., 2013), yielding the need for stimuli that more accurately reflect real-life facial behaviours. Given the rapidly growing interest in spontaneous FE, numerous spontaneous facial expression databases (hereafter referred to as FEDB) have been developed to alleviate past limitations. Overall, the increased use of spontaneous FE has established rigorous empirical milestones/bases of understanding real-life expressions - paired with high ecological validity (Dawel et al., 2021). This paper aims to provide a comprehensive overview of existing spontaneous and dynamic FEDBs, highlighting their conceptual and technical features. By doing so, we hope to assist prospective researchers in making well-informed decisions regarding stimulus selection.

3.1.1 Past literature on spontaneous facial expressions

It is important to differentiate spontaneous expressions from other types of expressions, which have been conjunctively used in previous research (Sneddon et al., 2012). In this review, we defined "spontaneous" displays as facial expressions that have been elicited via induction (e.g., by watching a video, hearing jokes, or playing games) or simulation (e.g., by recalling autobiographical memories; Rosenberg & Ekman, 2020; Zhang et al., 2014). Spontaneous expressions encompass a wide range of involuntary facial displays elicited through experimental manipulations that resemble emotionally evoking situations in the real world (Weber et al., 2018). Accordingly, spontaneous expressions depict (unlike posed displays) more accurate and ecologically valid representations of facial behaviour.

While preserving the time course of emotional episodes, spontaneous expressions allow for sufficient experimental control over contextual variations. In terms of facial features, they are less prototypical and more ambiguous due to the high degree of idiosyncrasy and variability between senders (Barrett et al., 2019). This flexibility do not strictly adhere to fixed signals and therefore allows for a range of unique appearances, such as blended expressions (Calvo et al., 2016; Calvo & Nummenmaa, 2016), or varied morphological patterns that go beyond prototypical patterns of facial configurations (Hassin, Aviezer & Bentin, 2013; Parkinson, 2013). Such complex actions in the form of co-occurring activation of different facial muscles, coupled with varying levels of intensity, are often an indicator of spontaneous expressions (e.g., Duchenne smile; Sheldon et al., 2021, but see Krumhuber & Kappas, 2022).

Spontaneous FEs are marked by distinctive temporal properties, such as relatively slow onset and offset timing (Krumhuber et al., 2013; Schmidt, 2006). The apex of these expressions is more fluid and less controlled than those in deliberate expressions, with variable durations and possibly multiple apexes within a single expression (Ekman & Rosenberg, 2005; Pantic & Patras, 2006). Subtle spontaneous FE, such as micro-expressions, last only a fraction of a second, indicating emotional leakage (Ekman, 2006). Such temporal characteristics highlight the rapid, automatic nature of spontaneous FEs, which manifest without much conscious deliberation (Schmidt et al., 2006).

The unique temporal and morphological characteristics of spontaneous FEs influence how they are perceived, affecting attributes like genuineness, trustworthiness, and intensity (Kocsor et al., 2019; Sauter & Fischer, 2018; Zloteanu et al., 2021). However, these factors also present challenges for emotion recognition. A vast number of studies have shown generally weak to moderate recognition accuracy for spontaneous expressions, whether in static (26-38%: Motley & Camden, 1988; Naab & Russell, 2007; Yik, et al., 1998) and dynamic formats (15-63%: Wagner, 1990; Wagner et al, 1986; Hess & Blairy, 2001). Critically, these rates are noticeably lower than those reported for posed expressions (generally above 70%; Calvo & Nummenmaa, 2016). Machine recognition also typically fares worse with spontaneous as opposed to posed expressions (Yitzhak et al., 2017; Krumhuber et al., 2021). Importantly, dynamic motion benefits the recognition of spontaneous expressions by offering distinct temporal information for emotion discrimination (Krumhuber et al., 2023).

The complexity of spontaneous FEs may be at the root of these difficulties, as intricate facial patterns and varying intensities introduce uncertainty in emotion recognition (Calvo et al., 2016; Cohn et al., 2007; Ito et al., 2017). Further research has examined the complex relationship between the morphological and temporal aspects of spontaneous FE, uncovering how these factors interplay to shape the perception of expression (Cohn & Schmidt, 2004; Namba et al., 2017). For example, while specific facial configurations might signal a particular emotion, the accompanying speed and rhythm of movements can add layers of nuance to how the expression is ultimately perceived (Pollick et al., 2003). In light of these findings, the challenges in recognising spontaneous FE illuminate areas for further exploration. By enriching our understanding of the subtle dynamics of temporal and morphological characteristics, they not only enrich our insight into human emotions but also prompt further exploration into the mechanisms behind the perception and recognition of spontaneous FEs.

3.1.2 Past literature on database reviews

The rising interest in spontaneous FE has led to various attempts to provide a comprehensive review of FEDBs. These efforts have illuminated a wide variety of available databases, each presenting unique features and characteristics. Such detailed reviews have led to significant progress in cataloguing and analysing spontaneous FEDBs, yielding valuable insights into the current landscape of the field. Some reviews have specifically targeted different facets of expressions, focusing on specific contexts (such as learning or driving; Li et al., 2023; Li et al., 2022), micro-expression (Li et al., 2013; Yan et al., 2013), or encoder demographics (e.g., Chinese; Cheng et al., 2014).

Reviews of databases embracing spontaneous FE have emerged as a critical area of inquiry. To grasp the intricate methodological and technical details of database construction, reviews typically focused on the size of the database (Cheng et al., 2017), encoder demographics (Ben et al., 2022), data acquisition methods (i.e., elicitation techniques; Levenson, 2007; Li et al., 2020, Li et al., 2022), included emotions (Qu et al., 2016), recording qualities (e.g., resolution, frame rate; Wang et al., 2021), and annotations (Wang et al., 2010). On the recognition side, emphasis has predominantly been focused on automated algorithmic models used for emotion recognition (Dupré et al., 2020), with identified challenges in constructing spontaneous FEDBs and a noticeable absence of unified evaluation standards (Jia et al., 2021).

Despite these efforts, the unique characteristics of spontaneous FEDBs warrant a more attentive and detailed discourse. While many reviews provided descriptive overviews (with varying levels of technical details), they often fall short in addressing the implication of these features for end-users and prospective database authors. Identifying commonalities across different groups of FEDBs is vital in guiding appropriate database selection, fostering new development, and promoting standardised practice across the field. A systematic review of spontaneous FEDBs can reveal unresolved gaps in the literature, contributing to a richer understanding of everyday facial expressions. Nevertheless, an in-depth discussion of the broader conceptual and practical implications, such as the efficacy of elicitation methods, sample representativeness, or the comprehensiveness of included emotion categories, has often been overlooked in prior reviews.

Additionally, while the availability of spontaneous databases continues to grow, existing reviews (Dawel et al., 2021; Fabricio et al., 2022) still typically encompass both posed and spontaneous FEDBs, including only a limited number of spontaneous FEDBs (between 4 to 9). The included databases are commonly centred on the most well-known databases, such as CK+, DISFA, MMI, and BP4D. Shifting focus towards spontaneous FEDBs offers a unique opportunity to test various predominant theories, such as appraisal or regulation theory of emotion (Gross, 2015; Scherer, 2005), and delve deeper into the interplay between expressed behaviour and underlying affects (Schmidt et al., 2003).

Spontaneous FEDBs, in contrast to posed ones, offer more fine-grained forms of expressions and enable the exploration of reliable display rules (Dawel et al., 2017; Ekman et al., 1990; Frank et al., 1993). These rules conceivably contribute to differences in a vast array of expression perceptions (Mehu et al., 2007; Sheldon et al., 2021) and facilitate an understanding of distinctions between posed and spontaneous expressions (Jia et al., 2021). Furthermore, spontaneous FEDBs provide context-specific information that varies across encoders and settings (Bänziger & Scherer, 2007; Sneddon et al., 2012).

Moreover, many reviews continue to encompass both static (typically JAFFE and Genki; Happy et al., 2017; Wang et al., 2021) and dynamic spontaneous FEDBs, which may inadvertently detract from the thorough exploration of dynamic and spontaneous FEDBs. Despite the fact that static databases have contributed to wide applications in affective science, computer vision, and psychology, the emerging prominence of dynamic FEDBs highlights their additional benefits.

Several studies incorporating these dynamic FEDBs emphasised a more extensive range of spatiotemporal information, including subtle changes in facial features over time and the spatial configuration of these features (Dawel et al., 2021; Krumhuber et al., 2017; Weber et al., 2018). As a result, spontaneous FEDBs in video format have become increasingly common in emotion recognition and perception research, exploring aspects like emotion judgement, perceived intensity and genuineness (Xiao et al., 2014; Zloteanu et al., 2020). Furthermore, dynamic stimulus sets play an essential role in advancing computer algorithms designed to recognise and respond to emotional signals (Sandbach et al., 2012). Given the growing interest in dynamic FEDBs, there is a clear imperative to focus on this area, facilitating a richer comprehension of spontaneous expressions.

3.1.3 Aims of present research

Facial expression research has steadily progressed toward more realistic stimuli, resulting in an increasing number of spontaneous and dynamic FEDBs. Despite this trend, no paper exists to date that would provide a systematic review of relevant stimulus sets. Such an attempt requires meticulous consideration of multiple aspects such as the number of emotions, elicitation techniques, annotation method, and type of validation. The present paper aims to address this knowledge gap by providing a comprehensive overview of existing spontaneous and dynamic FEDBs, identifying key dimensions and properties of the available sets: (a) *conceptual features* (Table 3.1), which reflect the thematic approaches in database construction and validation (b) *technical features* (Table 3.2), which include methodological considerations, such as experimental controls and recording setup), and (c) *practical features* (Table 3.3), which entail information about access to each database. In addition to providing details on each dataset, the discussion addresses the strengths and limitations of different approaches and techniques, as well as conceivable directions for future research on spontaneous expressions. By doing so, this review serves as a useful guide for researchers, assisting in the informed selection and tailoring of the stimuli suitable for specific research needs.

3.2 Methods

3.2.1 Literature search

The search for relevant articles entailed an exploration of pertinent literature from June 2019 to September 2022. The search procedure was built on a refined syntax which aligned with the strategies frequently employed in emotion research. The search terms were curated to focus exclusively on the studies of human faces published after 1990. This timeframe was chosen in light of the recent surge of interest in spontaneous and dynamic FEs.

The syntax for the search was crafted using a composition of keywords synonymous with spontaneous and dynamic facial expression databases. Keywords were interlaced using the Boolean operators "AND" and "OR", thereby enhancing the scope and depth of the search. To concentrate our efforts on highly relevant materials, we constrained our search to titles and abstracts, represented by "*[tiab]*", while the truncation symbol "*" was deployed to accommodate all derivatives of a keyword (e.g., mov*: move, moving, movement).

The composite of this meticulous process resulted in the following search syntax: *(spon* OR natural OR genuine OR authentic* OR real* OR involuntary OR induc*) [tiab] AND (dynamic* OR mov* OR motion OR action OR video) [tiab] AND (face OR facial OR emotion* OR affect* OR nonverbal* OR physio*) [tiab] AND (express* OR behaviour* OR* *display* OR visage OR present* OR manifest* OR feature OR communication) [tiab] AND (data* OR corp* OR collection) [tiab]*.

Inclusion in this systematic review was predicated on stringent selection criteria: (a) publications dating from 1990 to 2022, (b) peer-reviewed articles or conference proceedings (not abstracts), (c) public accessibility of database, (d) sufficient provision of database details (e.g. elicitation method, emotion, modality etc.), (e) recordings of real human encoders, (f) facial stimuli either with or without body gestures in visual or audio-visual modality, (g) at least one emotion depiction (either basic or non-basic emotion) and, (h) dynamic stimuli in the form of videos or image sequences. Conversely, papers that featured a non-publicly available stimulus sets that were not written in the English language or showcased artificial faces or those that were partially obscured/displayed under suboptimal conditions (e.g., point-light displays) were excluded from the current review.

3.2.2 Data sources and selection

The search procedure was primarily centred around three focal search engines/databases: *PsychInfo, PubMed, and Web of Science*, spanning the period from 1990 to September 2022. To broaden the search spectrum, supplementary platforms such as Google Scholar and AI-integrated search tools (including Elicit.org) were also used. We further undertook a manual examination of previous review papers and reference lists pertaining to FEDBs for potential inclusion in this research.

All retrieved articles were subsequently imported into *Zotero* reference management software to systematically handle the search results and streamline the removal of duplicates across various sources. The literature selection protocol adhered strictly to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA; Moher et al., 2009) guidelines, a structured approach offering a robust framework for segregating pertinent literature. Buffering a stepwise approach for segregating appropriate literature for review, we conducted the following four main steps: (1) sourcing literature that described spontaneous FEDBs from diverse sources, (2) a preliminary screening of titles and abstracts to gauge their relevance, (3) an exhaustive full-text review to confirm eligibility, and (4) finalising the selection of databases to be included in this review.

The initial search yielded 1,414 records, of which 355 were derived from PsychINFO, 589 from Web of Science, 394 from PubMed and 92 from other sources (see Figure 3.1). Post-review of tables and reference lists led to an additional inclusion of 20 studies. Duplicate entries were systematically removed, leaving 991 papers for title and abstract screening. Among those, 786 papers were deemed irrelevant, reducing the pool to 205 papers for a thorough full-text review. Following this in-depth analysis, 143 articles were disqualified for not meeting the predetermined inclusion criteria. The refined list comprised 62 papers that qualified as publicly available spontaneous and dynamic facial expression stimulus sets (Figure 3.1).

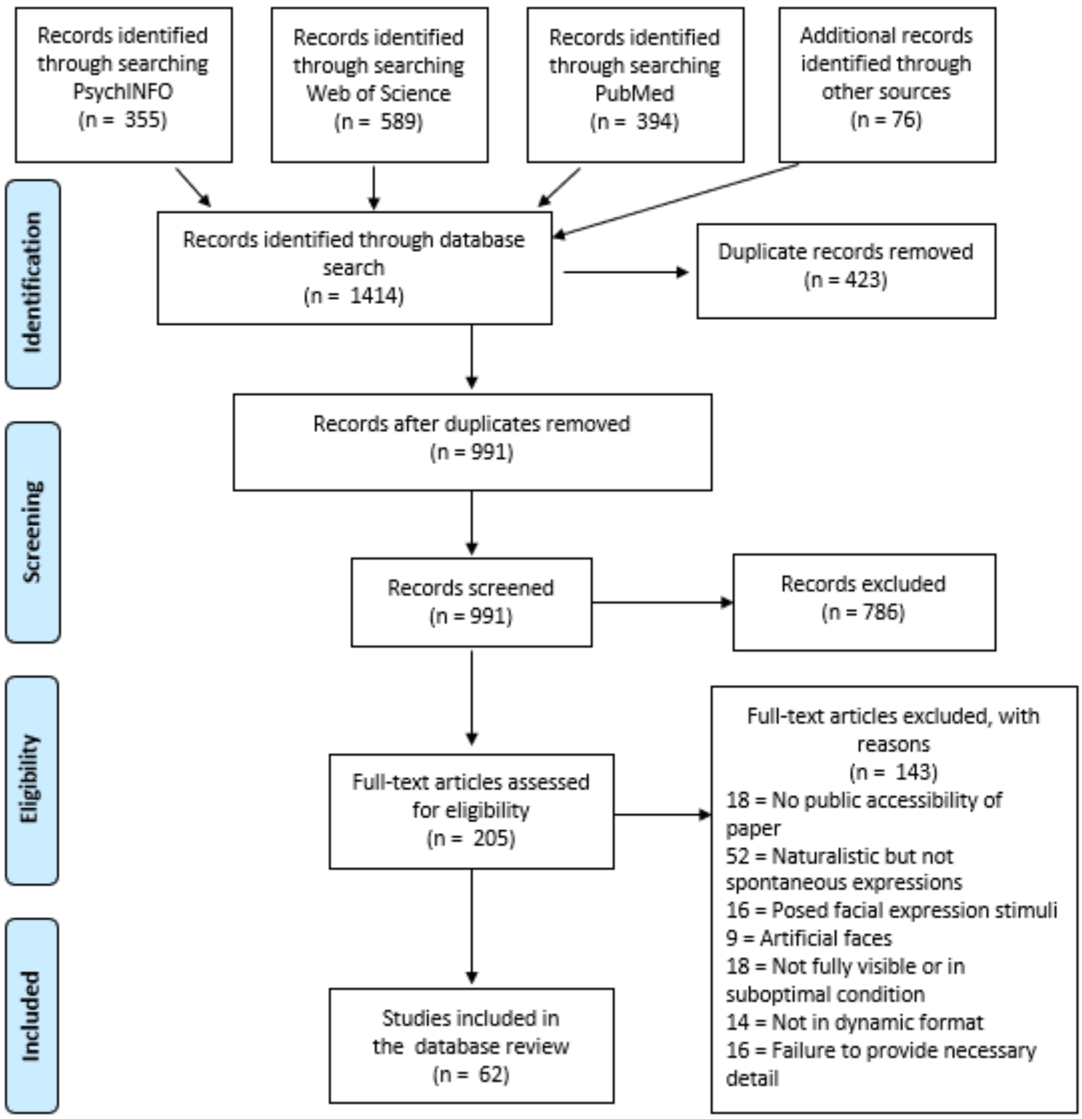


Figure 3.1. PRISMA flow diagram used to conduct the systematic literature search.

3.3 Conceptual features

This section reviews the thematic approaches that inform the development and validation of spontaneous and dynamic FEDBs. It aims to elucidate the conceptual features that characterise each database, i.e., emotional content and elicitation technique, encoder demographics, and measurement. To guide the selection process, Table 3.1 summarises key conceptual aspects that demonstrate the scope and potential applicability of each FEDB.

3.3.1 Emotional content and elicitation techniques

When selecting a suitable FEDB, it is important to ascertain whether the database contains relevant information for specific research needs (Wagner, 1997). The choice of spontaneous FEDBs typically depends on how well they convey information that can represent real-life expressions and whether the emotion can be detected by observers (Calvo & Nummenmaa, 2016). Depending on the aim of the research, researchers may find that some FEDBs may be more or less suitable to meet specific goals or requirements.

A review of prevailing emotional concepts indicated a marked preference for the categorical paradigm, with approximately 70% of FEDBs adopting this approach. It operates under the presumption that each expression signifies a distinct emotion that is mutually exclusive from others (Ekman & Cordaro, 2011). Within this paradigm, a few discrete categories represent a broader family of emotions, demarcated by stringent criteria (Cowie et al., 2001). Typically, these categories consist of three to six basic emotions as outlined by Basic Emotion Theory (BET; Ekman, 2005; Keltner et al., 2019) and often feature a neutral face as a baseline facial display. In some databases, emotions such as anger and fear are omitted (e.g., DISFA, ISED, LIRIS-CSE) likely due to their difficulty of induction in the laboratory, and/or replaced by contempt (SAFE-FE). The categorical approach ensures that each emotion stands out clearly (Calvo & Fernández-Martín, 2013), making this type of stimulus set particularly suitable for emotion recognition studies (Krumhuber et al., 2017). Despite BET paradigms remain influential (Cordaro et al., 2018), they face growing criticisms for oversimplifying the diverse nature of human emotions (Barrett et al., 2019; Krumhuber et al., 2013).

Acknowledging these constraints, subsequent FEDBs have broadened their scope beyond the basic emotion categories (e.g., BP4D+, DECAF, DynEmo), integrating additional states like boredom and confusion to indicate various levels of engagement during expression elicitation. Some databases also include subtypes of emotions such as amusement, enthusiasm, and liking, thereby representing different degrees of arousal that might be overlooked if generic labels (i.e., happiness) alone were used (Russell, 1980). This approach enriches the emotional content of FEDBs, offering more diverse portrayals of emotions similar to those encountered in daily life (Calvo & D’Mello, 2010; Krumhuber et al., 2017). For example, emotions like satisfied, excitement or being moved are bundled under the generic label of ‘happiness’ in BET, although they may represent different emotional states (Del Líbano et al., 2018). The inclusion of nuanced emotions also facilitates a deeper examination of composite emotion blends, such as joyful surprise or happy cry (Zhang et al., 2014).

In total, existing FEDBs capture 46 different discrete categories. While this diversity includes various non-basic emotions prevalent in real-life social interaction (e.g., ‘new basic emotions’; Cordaro et al., 2018; Keltner et al., 2019), some categories may not necessarily reflect affective sates but rather cognitive or engagement states (e.g., concentration, puzzlement, nervousness, thinking, fatigue). This expansion brings its challenges. Amplifying the range of emotional categories inherently introduces complexity to the annotation process due to the lack of universally agreed-upon criteria. Furthermore, automated systems processing these FEDBs may necessitate substantial computational power, which could pose challenges in terms of storage and system performance (Küntzler et al., 2021).

On the other hand, around 20% of FEDBs are highly specialised by focusing on one to three emotion categories in alignment with specific research interests (e.g., EmoPain, Vinereactor, PDSTD). These emotions were most often happiness/amusement, sadness, and pain. The prevalence of these three emotions in specialised databases may stem from their distinctiveness and easy recognisability, which are often used as benchmarks in emotion research. Happiness/amusement, for example, is typically associated with clear (often universally recognised; Leppänen & Hietanen, 2004) facial cues such as smiling or laughter.

Sadness and pain, while more complex, are also associated with specific facial expressions that are relatively consistent across individuals (Boucher, 1969; Kunz et al, 2008). These emotions represent fundamental aspects of human experiences, encompassing a range of positive and negative states, which may be why they are frequently the focus of specialised emotion research.

The narrow scope of these databases enables a more detailed exploration of the subtleties and nuances associated with each targeted emotion. Additionally, concentrating on fewer emotions may facilitate the collection of a larger number of portrayals for specific emotions (AM-FED+, MAHNOB-laughter). However, the specialised nature of these databases may not provide a comprehensive understanding of the wide array of human expressions and may limit opportunities for comparative research across diverse emotional states. Furthermore, such databases may not be ideal when the research objective involves understanding the relationship or co-occurrence of different emotions, given their limited emotional scope.

While specialised and large-scale FEDBs provide fundamental data for emotion research, the richness and quality of expressions within databases are significantly shaped by the elicitation methods employed to induce emotions. Given the methods used to elicit emotions are crucial in determining the utility of databases, striking a balance between capturing spontaneous behaviour and maintaining data quality necessitates meticulous elicitation protocols. These protocols must carefully manage the trade-off between experimental control and the naturalness of the encoder's responses (Fanelli et al., 2010). Overly stringent control in the recording environment can inhibit the encoder's natural expression, while an overemphasis on naturalness can introduce noise into the data. A variety of emotion elicitation techniques (i.e., passive and active inductions) have been proposed to strike this balance, each offering unique advantages.

As can be seen in Table 3.1, over 90% of the FEDBs utilised passive induction techniques, thereby capturing encoders' responses to predetermined emotional stimuli designed to provoke an instant reaction. Multimedia stimuli, including sound, still images, videos or even texts, have demonstrated effectiveness in inducing emotions (Brave & Nass, 2002). Of these, emotional images or video clips remain the most prevalent stimuli across FEDBs (Gross & Levenson, 1995; Schaefer et al., 2010).

Images are often sourced from the International Affective Picture System (IAPS; Lang et al., 2005), a comprehensive collection of emotionally provocative photographs. IAPS covers a broad spectrum of emotional categories and intensities, facilitating the elicitation of diverse emotional responses. Notably, while images offer a snapshot of affect-relevant moments, they may not encapsulate the dynamic progression of emotional events (Devilly et al., 2021; but see Uhrig et al., 2016).

In comparison, videos provide a multisensory experience through moving scenes, auditory cues, and emotional context, thereby conveying more immersive and holistic emotional narratives (Gross & Levenson, 1995). Various sources, such as YouTube, TV shows and films, are often used to obtain emotionally connoted video materials. Consequently, they often evoke more intense and pronounced emotional responses that are found to be stronger than those elicited by images (Horvat et al., 2015).

These visual stimuli, tailored to provoke an instant reaction from encoders, have garnered empirical support for their efficacy in emotion induction (Brave & Nass, 2002; Gross & Levenson, 1995; Schaefer et al., 2010), showing a robust alignment with encoders' self-reported emotions, physiological responses, and neural correlates (for a review, see Siedlecka & Denson, 2019). FEDBs employing this approach capture expressions that closely resemble natural facial behaviours while maintaining the requisite control over recording environments (Coan & Allen, 2007). As a result, these databases often stand as the preferred method of choice for encoding studies (Scherer & Ellgring, 2007) by exhibiting observable facial manifestations that tightly reflect authentic affective experiences (Zloteanu & Krumhuber, 2021). This advantage carries over to decoding studies by furnishing an objective and reliable emotion label with consistency, which can serve as a robust benchmark for testing recognition rates. This precision enables researchers to assess how accurately individuals or machines can discern the displayed expression, thus advancing our understanding of emotion perception and recognition.

Despite those advantages, paradigms that rely solely on the viewing of visual stimuli can constrain the range of facial responses, overlooking the variability of real-world emotional experiences (Zupan & Eskritt, 2020). To address this potential limitation, several databases (e.g., BINED, BioVid, BP4D+, DynEmo) have diversified their elicitation techniques, incorporating interactive and actively engaging tasks (e.g., touching unknown objects in a box, smelling unpleasant odours, playing games). For specific emotions like pain, tasks inducing discomfort (such as the cold pressor task or arm rotations) have been used (Aung et al., 2016; Littlewort et al., 2007). These innovative approaches allow researchers to better capture complex emotional states such as secondary or self-conscious emotions (e.g., embarrassment, pride, pain), resulting in a wider spectrum of emotions beyond the basic six emotions. Such expansion also facilitates a context-sensitive analysis of emotional reactions, where individuals may respond differently as a function of the situation, the environment, or the presence of other people (Hamann & Canli, 2004; Koval & Kuppens, 2012). These databases are particularly valuable for formulating and advancing deep-learning models of emotion recognition (for a review see Bian et al., 2023), which seek to enhance human-computer interaction through metadata about the emotion-inducing context (Sneddon et al., 2012).

Beyond traditional induction methods, some researchers are exploring alternative techniques, such as autobiographical recall. While the majority of FEDBs exclusively employ induction-based elicitation techniques, such an approach restricts our understanding of FEs to predetermined contexts, primarily within laboratory settings. To broaden the scope, autobiographical recall of personal memories offers a well-validated method that bypasses the constraints of induction tasks (Levenson, 2007; Siedlecka & Denson, 2019). The imagery-based nature of this technique allows for flexible contextual representations, which may be challenging to elicit under laboratory conditions (El Haj et al., 2018). Moreover, while induction-elicited expressions often exhibit specific behavioural patterns (e.g., eye fixation on screen), self-appraised emotional experiences can enhance the diversity of expressions, potentially offering greater ecological validity (Philippot, Schaefer & Herbette, 2003).

These insights into the methods and techniques of emotion elicitation demonstrate the multifaceted approach required to study human emotions. The balance between passive and active techniques, laboratory control, and real-world applicability can provide a landscape for future research.

3.3.2 Encoder demographics

Among the databases reviewed, the number of encoders varies considerably, from a minimum of 7 (SPOS) to a maximum of 416 (AM-FED+). This significant variation reflects the diverse research priorities and resource constraints inherent in different databases. Datasets with fewer encoders tend to incorporate more recordings per individual (e.g., SPOS, FEEDTUM), allowing for the exploration of intra-individual variability in emotion expressions. This approach is particularly valuable for examining the coherence between the experience and expression of emotion within an individual, and how person-specific factors (e.g., personality traits) may affect this relationship. In contrast, datasets with a larger subject pool (e.g., BINED, EB+) can better capture inter-individual variability in facial behaviours. This is vital for developing affective computing systems that are robust to individual differences in facial features and generalize effectively to new faces. Despite the large variation in sample size, most databases show a preference for medium numbers of encoders ranging from 40 to 90, suggesting a trend towards databases that balance a larger, yet feasible, number of encoders.

Another notable trend across datasets is the overrepresentation of young adults (approximately 50%), possibly driven by the convenience of recruiting within academic settings. Merely a subset of databases contains a wider age span (e.g., BP4D-spontaneous, BP4D+, EB+). While this trend might seemingly represent a demographic snapshot, it risks obscuring the influence of age on emotional expression including factors such as cognitive maturation/decline, muscle atrophy, and wrinkles (Houstis & Kiliaridis, 2009; Ko et al., 2021). This skewed representation presents a potential gap in understanding the nature of facial expressivity in children and older adults. To compensate for this marked preference, some databases have expanded their recruitment criteria to include a wider age range, enriching their sample size and providing a more inclusive representation of the age spectrum (e.g., 4DFAB, UvA-Nemo). Certain databases have targeted specific age groups like children (ChildEFES, LIRIS-CSE); nonetheless, the underrepresentation of the elderly population persists.

Most databases have a fairly equal representation of male and female encoders, although some are slightly skewed towards one gender group (e.g., NVIE, UT-Dallas). Balanced gender ratios are paramount for stimulus development to reflect known differences in emotion processing between the sexes (Wiswesser et al., 2018). Despite efforts to incorporate diverse ethnic backgrounds, there remains a skewed focus on White/Caucasian and Asian encoders, likely reflecting the geographical location of data acquisition. While this approach provides valuable insights into culture-specific differences in expression, it poses challenges to the broader cross-cultural generalisability of a dataset. The representation of diverse ethnic backgrounds is of particular importance for spontaneous facial expressions, which may be subject to cultural dialects (Elfenbein et al., 2007).

3.3.3 Measurement and annotation

The effectiveness of elicitation techniques is profoundly tied to the accuracy and comprehensiveness of annotation (Lucey et al., 2010). Databases must accurately specify the emotional content of recordings, otherwise, they risk becoming a mere collection of stimuli without interpretational value (Sneddon et al., 2012). However, data annotation is a labour- and time-intensive task that demands considerable effort by the experimenter, such as determining the labelling criteria which remain to date an open question (Fanelli et al., 2010). This process might be further complicated by the subtlety and complexity of spontaneous expressions. Having well-annotated videos of facial behaviour considerably amplifies the value of a database, especially for affective computing research which relies on the training and testing of machine algorithms (Zhang et al., 2014). Additionally, annotation assists database users in selecting stimuli corresponding to specific features of interest.

As illustrated in Table 3.1, most databases provide annotation to some extent, serving as empirical ground truth for facial expressions. These annotations commonly adhere strictly to predefined emotion categories (e.g., basic six emotions) or dimensions and align with the emotion-inducing stimulus designed to elicit an affective reaction (Lucey et al., 2010). This adherence fosters uniformity and comparability across databases, providing a systematic and validated framework for interpreting FEs. While access to well-labelled data enables a systematic and validated framework for interpretation, annotation in terms of only one emotion category can be problematic, especially when there is an inconsistency between what is shown on the screen (e.g., an amusing scene) and what is experienced by the encoder (e.g., surprise, disgust). To avoid the potential risks of oversimplifying multifaceted emotional states, it is imperative to treat data labels with caution (Fanelli et al., 2010; Zhalehpour et al., 2017).

A subset of databases has integrated a more holistic approach that combines nominal categorisation of emotions and facial patterns to enrich the understanding of FEs (Gerardo &

Menezes, 2019). Here, the Facial Action Coding System (FACS; Ekman et al., 2002) might prove useful for analysing subtle and complex facial configurations (e.g., in BP4D+, EB+, MMI). Approximately 30% of FEDBs have adopted FACS coding to comprehensively quantify facial behaviours in terms of action units (AUs). Unlike other annotation approaches relying on a limited set of predefined emotion labels, FACS directly measures observable AUs without presumptions about the underlying affective state. This enables greater flexibility in interpreting facial behaviours from a broader perspective, especially in the context of spontaneous expressions, where complex combinations of AUs often involved. Also, researchers can refer to the extensive empirical evidence linking AUs to a wide range of affective states that might otherwise be overlooked by fixed emotion categories.

Despite the advantages of various annotation approaches in exploring and validating different theoretical spectrums, each method comes with its unique challenges. An ideal database would incorporate both emotions and AU annotations to maximise its potential utility (e.g., DSME-3D, SAMM). The implementation of these diverse methods requires careful consideration, and often human annotators play a critical role in the process.

For meticulous and reliable annotation, employing a diverse range of annotators - from naïve observers to individuals specifically equipped with FACS coding skills - is recommended. Naïve observers, despite their lack of formal training, can offer valuable insights into the general population's perception of FE. Their judgement can serve as a useful reference point for understanding how FEs are interpreted in everyday contexts. In contrast, FACS-trained coders bring a higher level of reliability and objectivity, adhering to a strict coding framework that reduces personal interpretation and potential biases, thus enhancing the overall quality of the FEDBs (Ekman et al., 2002). Their ability to decipher complex and mixed emotions (Sayette et al., 2001), which often poses challenges to untrained human annotators (Naab & Russell, 2007), is an additional significant advantage.

Irrespective of the chosen annotation method, it is crucial to secure a robust level of agreement among annotators to ensure consistency and validity in the assignment of emotion/AU labels. Various statistical measures, such as Fleiss' or Cohen's Kappa or other analyses like intra-class correlation (ICC), have been implemented to ascertain interrater agreement.

In addition to human annotators, approximately 10% of databases employed semi-automated approaches, where automated facial expression analysis (AFEA) tools aid in streamlining the annotation process, enabling the efficient management of annotating large volumes of recordings. AFEA tools bring remarkable computational efficiency, allowing the processing of vast FEDBs in a fraction of the time that manual annotation would require (Pantic & Rothkrantz, 2004). Their capability in the simultaneous analysis of emotions, in addition to AUs, is an appealing feature of their algorithmic sophistication (Bishay et al., 2023). Nevertheless, it is important to note that the effectiveness of AFEA tools is intertwined with the quality of training data and the robustness of the algorithms utilised (Dhall et al., 2012). This highlights the indispensable role of human researchers in ensuring good-quality automated annotations.

3.3.4 Evaluation methods

To validate the emotional content of recordings, most FEDBs provide self-reports of subjective states as a resource for stimulus assessment. This introspective approach allows encoders to articulate their feelings, offering a convenient, albeit not always reliable (Cowen & Keltner, 2017), window into a person's emotional experience. These methods frequently employ categorical ratings, enabling individuals to classify their feelings into distinct sets such as the basic six emotions. Complementing this, self-observation of an encoder's own FEs merges internal emotional states with external manifestations (e.g., CASME II), weaving together a comprehensive narrative of emotion display. Moreover, encoders were asked to provide their emotional recognition of elicitation stimuli (e.g., DEAP, OL-SFED), predicting the intended emotional induction and thus offering insights into the dynamics of how specific stimuli can elicit particular emotional responses. Some databases (e.g., BioVid, Emognition, PPB-Emo) have integrated continuous ratings in which affective states are mapped onto dimensional spaces including valence and arousal (and dominance) to capture subtle variations and gradations inherent in affective states (Russell, 1980). Some databases (e.g., PEDFE, DynEmo) further measure emotion genuineness and action readiness (e.g., approach, avoidance), offering valuable insights for distinguishing between posed and spontaneous expressions and evaluating behavioural intentions.

While these mentioned tasks contribute fine-grained emotional details from the encoders' perspective, inherent challenges rooted in subjectivity persist (Matsumoto & Ekman, 2010), thereby necessitating complementary external observation to obtain an objective quality of emotional experiences.

The emotional content of recordings can also be validated by external inferences from human observers or machine recognition. Consensus judgments by naïve observers, who are asked to classify expressions, are typically considered to be more objective because they are less prone to social desirability and memory biases. In line with annotation, the majority (52%) of databases used categorical emotion recognition, while only five databases incorporated dimensional ratings. A mere 6% of FEDBs include both discrete and dimensional ratings. This tendency again emphasises a preference for discrete emotion paradigms, which may be due to their simplicity and straightforward interpretation. Furthermore, nearly 30% of databases evaluated AU recognition by using AFEA.

Naïve human observers, although deployed in only a handful of databases (e.g., PDSTD, PEDFE), play a critical role in complementing the intrinsic and subjective nature of self-reports by encoders. Their judgements bridge the gap between subjective and objective, internal and external ratings, enabling a more comprehensive database evaluation. However, the forced-choice paradigm often used for categorical emotion recognition may result in lower ecological validity by suggesting the use of emotion labels that might not otherwise be chosen (Russell, 1993; Wagner, 1997). Also, the cost and effort associated with collecting data from human participants can be large. For this reason, AFEA stands out as the primary evaluation method for the majority of databases (e.g., EB+, Emognition, iSAFE). AFEA classifiers extend the scope of emotion recognition beyond human capabilities, offering a standardised and efficient approach for processing large amounts of data. AFEA can provide both discrete and continuous ratings of emotions and AUs.

These classifiers, whether in-house or commercial software, play a significant role in the foundational methodology of database evaluation. However, their effectiveness is contingent on the quality of training data, and the robustness of the algorithms utilised (Koelstra et al., 2010). While AFEA efficiently captures facial movements, they often overlook contextual information, a crucial dimension for emotion recognition (Calvo & Nummenmaa, 2016). Despite the advantage given by the precision of AFEA, human annotation remains valuable for validating algorithms and addressing challenges that automated approaches cannot yet handle well.

For all validation techniques, it is crucial to note that an over-reliance on any single approach for measuring or labelling facial behaviours can be problematic. Emotions represent multifaceted processes that cannot be adequately processed by a single system (Gross & John, 1997). For instance, inconsistencies may emerge in how emotions are experienced, expressed, or perceived due to social norms that down- or up-regulate different components of emotional responses (Bonanno & Keltner, 2004; Eisenkraft & Elfenbein, 2010). To avoid oversimplifying the intricate nature of emotional experiences, it is imperative to utilise and integrate multiple, complementary methods to thoroughly measure and validate the relations between emotions and facial behaviours. Some databases (e.g., PEDFE, iSAFE) have adopted such a comprehensive framework by incorporating self-report data, observer ratings, and FACS coding, which maximise the potential utility of FEDBs.

3.4 Technical Features

This section reviews the technical features of FEDBs, i.e., stimulus numbers, technical controls and duration, frame rate and resolution, and modality. These features are crucial for the utilisation of FEDBs in human studies as well as for human-computer interaction. Table 3.2 provides a summary of these technical features for each dataset.

3.4.1 Stimulus number and technical control

The strategic development of a diverse range of emotional portrayals within FEDBs relies heavily on the active and thoughtful participation of a sufficient number of encoders. However, preserving data quality while simultaneously expanding the breadth of emotional portrayals is challenging. Databases vary considerably in the number of captured expressions, with the total amount of recordings ranging from as few as 26 (i.e., Face of Pain) to as many as 3455 (i.e., Vinereactor). Overall, most databases depict high numbers of recordings (approximately 400 to 700), indicating a wide spectrum of facial expressions being captured. Large stimulus numbers are particularly important for the training and testing of computer models sufficiently robust to stimulus variations. They can also act as a benchmark for comparing different expression recognition algorithms (Turk & Pentland, 1991; Valstar et al., 2015, 2017).

The recording environment plays an important role in the size of the database. Traditionally, most FEDBs have been cultivated in laboratory settings with a feasible number of encoders (mostly 20 to 90; see Table 3.2). Alternatively, some databases have opted for a remote approach, where participants are recruited online through crowdsourcing platforms (e.g., AM-FED+, Vinereactor). While remote recording alleviates recruiting difficulties and potentially offers more representative portrayals, it may simultaneously introduce potential noise due to technical constraints, such as inconsistent recording quality or visibility of the face (McDuff et al., 2019). Unique challenges in different recording planforms underscore the importance of meticulous control over the environment to ensure data quality. The decision to adopt one approach over the other during the construction of the database is strategic and should align with the intended research applications of the database.

It is worth noting that despite the logistical constraints of engaging a large number of participants for lab-based recordings, databases such as BP4D+ and BINED have successfully maintained a larger encoder pool and high-quality recordings. Besides environmental settings, recording conditions, such as camera resolution and frame rate, are essential in capturing intricate details of FEs. High resolution is particularly important for recording spontaneous expressions, which are characterised by their heterogeneous nature and subtle variations (Namba et al., 2017; Pfister et al., 2011).

3.4.2 Stimulus features

With regard to the recording quality, most databases utilise medium (640 × 480) to high-resolution (1920 × 1080) cameras. The latter captures more detailed changes in the morphological features (e.g., shapes & textures) of facial behaviour, which is crucial for the accurate representation of emotional states, particularly in the case of micro-expressions (Li et al., 2022). While lower resolution may sacrifice detail (Han et al., 2020), it can still be beneficial for training algorithms due to their smaller data size and easier manageability. Nevertheless, nominal resolution is particularly important for fine-grained analysis of spontaneous expressions, which often manifest heterogeneous and subtle facial configurations (Namba et al., 2017; Pfister et al., 2011). Therefore, a thoughtful calibration of recording resolution, considering the trade-offs between the level of details and computational demands, can optimise the quality and broad utilisation of FEDBs.

Frame rate refers to the number of frames recorded per second (fps). The higher the frame rate, the smoother and more fluid the facial movement appears to be. Therefore, the frame rate of recordings substantially determines the amount of temporal information encapsulated within FEDBs. The majority of FEDBs typically employ frame rates within the range of 30 to 60 fps, effectively balancing visual clarity and methodological efficiency (see Table 3.2). Such a trend implies that these frame rates are adept at capturing the sophisticated temporal dynamics inherent in spontaneous FEs, particularly the subtle and swift changes indicative of complex emotions. Some databases (e.g., CASME II, SAMM) opt for notably higher frame rates (100 to 200 fps), specifically equipped to accurately capture fleeting micro-expressions that persist only for a fraction of a second (Davison et al., 2018; Yan et al., 2014). A higher frame rate is preferred for capturing and analysing the dynamic trajectory of spontaneous displays (Krumhuber et al., 2023), which provides a more vivid depiction of how facial expressions unfold and progress over time (Leonard et al., 1991). The enhanced temporal resolution also facilitates the identification of rapid and transient facial movements that might be easily missed at lower frame rates.

Lastly, there is variability across databases in the duration of portrayals, ranging from 0.5 seconds for micro-expressions (e.g., SMIC) to 15 minutes (DynEmo). Extended durations intuitively offer more information. However, long videos can also encompass periods devoid of emotional content, especially if the encoder is recorded throughout the entire elicitation tasks (e.g., Emognition). Such non-emotive periods can introduce noise to the data, potentially complicating the analysis and recognition of facial expressions. For this reason, most databases have segmented their recordings, thereby concentrating on key expressive phases from onset to apex and offset.

3.4.3 Recording features

The strict control of recording environments that induce emotional expressions effectively mitigates the influence of extraneous variables. Factors such as ambient noise, lighting conditions, or distractions can be minimised, creating an environment that is optimally equipped for emotion elicitation (Fanelli et al., 2010; Gallagher, 2016). The ideal database strikes a balance between the quantity and quality of data, ensuring technical accuracy while maintaining the ecological validity of elicitation environments.

The recording setting is a prerequisite for ensuring data quality, with particular emphasis on lighting conditions and background consistency. Many databases (approximately 65%) have implemented additional lighting sources to establish stable illumination (see Table 3.2). The nature of these sources varies from stand lamps to LED lights, often placed next to the camera (e.g., CASME, ISED, NVIE). Stable lighting has profound implications for FE recognition. The direction, intensity, and colour of lights can significantly impact face processing. For instance, well-balanced illumination can enhance the clarity of facial features, thereby augmenting the recognisability of FE (Shi et al., 2011). Conversely, extreme lighting conditions, such as dim or overly bright lighting, can obscure the face from the surrounding environment (Koringa et al., 2017). As corroborated by several studies, illumination stability positively influences recognition accuracy (Stratou et al., 2011), while flickering or varying lights can hinder recognition (Wang et al., 2013; Nguyen et al., 2014). This highlights the significance of stable lighting in the quality of FEDBs.

Across all databases, controlling the recording background is common practice (see Table 3.2). Converging evidence suggests that visual factors, including the colour, complexity, and composition of the background, can significantly modulate the FE processing. For example, a high-contrast colour used in the background can accentuate FEs, thereby facilitating their recognition (Minami et al., 2018). Conversely, complex backgrounds may detract attention from the face, impeding recognition (Righart & De Gelder, 2008; Sannikov et al., 2017). Social factors, such as the presence of other people in the background, have also been shown to influence the interpretation of FEs (Kret & Gelder, 2010; Wieser & Brosch, 2012). In consequence, most databases have used plain backgrounds to improve standardisation.

3.4.4 Modality

Facial expressions serve as a powerful means for the transmission and interpretation of emotional states. Consequently, databases have predominantly focused on visual cues of expressed emotions (i.e., images, videos, or image sequences). However, other modalities such as audio (or audio-visual) and physiological signals offer equally significant insights into affective states (Juslin & Laukka, 2003; Vanny et al., 2013).

Audio signals are a rich source of cues for emotion recognition, particularly when visual cues are ambiguous (He et al., 2020). As a consequence, some databases include the sound of stimulus material to provide additional indicators of the intended emotion (Miolla et al., 2022). Others have recorded encoders whilst speaking (e.g., BAUM-1, HUMAINE) or discussing the elicitation stimulus (e.g., CAM3D, RECOLA, RU-FACS). Vocal cues, such as tone of voice, pitch, and speech patterns, supplement the information that cannot be captured by FEs alone (Scherer et al., 1985). The moving mouth is also a crucial region for conveying various emotional expressions (Barrett et al., 2019; Mehu & Scherer, 2015). This multimodal combination of facial and vocal cues is integral to emotion recognition, as they are interconnected and mutually informative.

In addition to vocal cues, the assessment of physiological responses, such as heart rate (HR), skin conductance level (SCL), electroencephalogram (EEG) and electromyography (EMG), is becoming increasingly popular (26%). The integration of these modalities offers a comprehensive and nuanced depiction of emotional responses, providing invaluable information when facial or vocal cues alone are insufficient (Bulagang et al., 2020; Kim & André, 2008). These indicators may be particularly pertinent for distinguishing between emotions like fear and surprise, which share similar facial actions (Zhao et al., 2017). Furthermore, physiological signals can reveal emotional responses that may not be overtly expressed through facial or vocal cues (Kim & André, 2008; Kreibig, 2010). However, data acquisition and synchronisation are challenging, thereby limiting the practicality of this type of recording in large-scale studies. As a result, most databases with this modality tend to have smaller encoder sizes, mostly between 16 and 50 (Video-fNIRS, Emognition). Despite these challenges, the multimodality of FEDBs presents a promising avenue for advancing the field of emotion expression research beyond what can be learned from a single modality.

3.5 Practical features

Committing to open science principles promotes knowledge sharing and offers long-term benefits for future research. To facilitate the utility of publicly available stimulus sets, this section provides practical information about dataset accessibility while emphasising ethical compliance for data usage. Table 3.3 summarises information on how to access the datasets and potential ethical restrictions to be considered.

To fully leverage the usefulness of spontaneous FEDBs, a multifaceted approach needs to be adopted that considers various factors. The accessibility and transparency of the data are fundamental for maximising the utility of a database. To this end, most databases provide access through a dedicated website link (see Table 3.3). These platforms often detail the database's key features and offer additional information (e.g., experimental manipulation, annotation) beyond the published article. In addition, the author's email address serves as an initial point of contact for inquiries about the database. Transparency and accessibility are instrumental for replication and validation purposes of research findings, thereby enabling informed decisions about the appropriate use of databases.

Several databases adopted the Open Science Framework (OSF), a platform designed to promote transparency, accessibility, and reproducibility in scientific research (Open Science Collaboration, 2015). The principles of open science can be applied to the creation of new databases as well as the maintenance of existing FEDBs. Transparency in the creation procedure can increase the awareness of researchers in terms of database characteristics, enabling them to determine the utility of the stimulus set for different contexts, cultures and machine algorithms.

Researchers have adopted various practices to ensure the ongoing distribution of datasets while safeguarding the responsible and ethical usage of datasets. Many databases mandate a signed End User License Agreement (EULA) to protect participants' rights and prevent potential data misuse. As such, most datasets are restricted to academic research purposes only, with additional consents required for commercial use (e.g., BP4D). These EULA forms can be accessed through the website links or by directly contacting database authors through the email address listed in Table 3.3.

Moreover, some databases are directly distributed through the Open Science Framework (OSF) platform without requiring any EULA. This practice streamlines data acquisition by removing administrative barriers and bypassing lengthy processes of obtaining ethical approval. Adherence to open science practices also offers long-term benefits to future database users by ensuring sustained access to data well beyond the initial publication date. By embracing these principles, database authors can guarantee the continued distribution of their data, thereby promoting knowledge distribution and collective advancement of the field in facial expression research.

Finally, it is important to note that almost all FEDBs restrict their use to academic purposes, ensuring that the data is used in line with the intended purpose of advancing academic research. Some databases (e.g., BP4D-Spontaneous, BP4D+, EB+) impose handling fees for data maintenance and delivery for over 10TB data. The EULA and access fees may also help to maintain the quality and consistency of the data and protect the rights of the participants involved, by ensuring that the data is used responsibly and ethnically.

3.6 General Discussion

In the last two decades, there has been a major shift in facial expression research towards more ecologically valid facial stimuli (Krumhuber et al., 2023). Unlike posed displays depicting highly standardised portrayals to maximise their recognisability, spontaneous expressions do not involve fixed signals of emotion (Parkinson, 2013), making them more variable but representative of affective responses seen in real life. Such growing interest in stimulus validity has notably accelerated the development of FEDBs, with a number of papers surveying existing corpora (e.g., Diconne et al., 2022; Guerdelli et al., 2022; Haamer et al., 2018; Siddiqui et al. 2022; Weber et al., 2018). Yet, a comprehensive review focusing exclusively on spontaneous, dynamic expressions is currently missing.

The present paper aims to fill that gap with the ultimate purpose to assist readers in their decisions about stimulus selection. While previous works have primarily focused on the general characteristics of a limited number of FEDBs, this review provides an in-depth exploration of the unique properties and features of spontaneous and dynamic FEDBs. In addition to detailing well-known databases, this review also sheds light on lesser-known databases in the field. This approach not only deepens our understanding of the specific characteristics of FEDBs but also offers insights into individual strengths and limitations, as well as the current state of the field. Consequently, it provides a more holistic perspective on the use of these resources in emotion research. The review was organised around three distinct themes: conceptual, technical, and practical features. This thematic approach provides a comprehensive perspective that can serve as a valuable resource for researchers. It can assist them in selecting and adapting stimuli suitable for their specific research objectives, thereby facilitating more effective and targeted use of FEDBs.

This review revealed a categorical focus on basic emotions, which has been the dominating approach of most FEDBs. The trend seemingly mirrors the enduring influence of discrete emotion theories such as BET according to which expressions are strictly categorised based on specific emotions. However, emotions may not be experienced in isolation but rather as blends or mixes of different emotional states (Du & Martinez, 2015). Instead, they interact and even morph into one another to provide nuanced meanings. The extensive focus on basic emotions highlights a lack of diversity in inducing more intricate and complex emotional states.

Among all FEDBs there is a clear trend towards a broader range of emotion categories beyond the basic six, thereby acknowledging the complexity and diversity of human experiences. Approximately 25% of databases extend to incorporate non-basic and more nuanced emotions (Kossaifi et al., 2019; Tcherkassof et al., 2013). Such expansion not only enriches the theoretical understanding of emotions but also holds practical significance for AFEA in terms of its ability to generalise to a wider spectrum of everyday emotional phenomena (Bänzinger et al., 2011; Gunes & Pantic, 2010).

This trend may also align with contemporary perspectives on emotions, such as the constructivist or appraisal theory of emotion (e.g., Barrett, 2006; Ellsworth, 2013), which posit that emotions are not fixed entities but are constructed from a variety of psychological and physiological components and can vary widely across individuals and contexts. This advancement in the field offers a deeper understanding of the human emotion spectrum. Moreover, it enriches our theoretical understanding of emotions, with practical implications for the development of more sophisticated AFEA algorithms that frequently rely on databases for training.

Although some FEDBs incorporate a mixture of elicitation techniques (active, passive, interactive), the majority rely on (audio-) visual materials (i.e., images, films, video-clips) for emotion induction. Unlike naturalistic displays which are sourced directly from real-world contexts or the Internet, thereby introducing noise to the data, spontaneous expressions are evoked under experimental conditions. The induction approach serves as a useful method for eliciting the intended emotion and simplifies the annotation process by aligning the emotion labels with the elicitation materials (Gross & Levenson, 1995; Yan et al., 2014), thereby ensuring consistent and replicable responses across senders. However, a significant concern is that only a handful of FEDBs have validated these materials or pre-tested their stimuli (Li et al., 2022; Saganowski et al., 2022). Increasingly, concerns have been raised about the effectiveness of such materials. Given that individuals may appraise materials in diverse ways (Barrett et al., 2019; Smith & Ellsworth, 1985), materials that have not undergone thorough validation tests may not be effective in inducing the intended emotions. Also, it may limit the number of emotion-inducing situations typically experienced in real life. Future work may aim for greater variety in experimental methods for inducing various emotions, also piloting materials/tasks for their effectiveness in evoking the relevant emotional state.

Alternatively, other procedures such as the revival of autobiographical memories could further diversify elicitation techniques (Levenson, 2007), but may be challenging to elicit in the laboratory (Siedlecka & Denson, 2018). Furthermore, self-appraised emotional experience from encoders during the autobiographical recall may increase the heterogeneity of expression, providing a more accurate reflection of real-life expressions (Philippot et al., 2003). However, the efficiency of autobiographical recall may be influenced by the time elapsed from the original experience and may result in relatively low expression intensity (Fradera & Ward, 2006; Nandrino et al., 2019). Also, summoning specific imagery events that are emotionally connoted may be highly dependent on individuals' cognitive ability (Addis et al., 2007; Robert, 2007).

Additionally, it is important to note that very few databases used social interactions to elicit spontaneous expressions. Social functionalism argues that emotional expressions coordinate individuals' behaviours within social interactions, serving three key functions: providing information to others, acting as incentives for social behaviour, and evoking specific responses in observers (Keltner & Kring, 1998; van Kleef, 2016). In real-life contexts, facial expressions signal intentions, social motives, and responses to others' actions (Frith, 2009; Parkinson, 2005). Without capturing these communicative elements, spontaneous FEDBs may miss key mechanisms of how facial expressions function in communication.

While interactive expressions have higher ecological validity, reflecting communicative value of facial expressions, there is a trade-off between ecological and internal validity. The lack of experimental controls in interaction makes it challenging to establish the ground truth (i.e., underlying emotional affects and specific emotional stimuli/antecedent events that elicited the expressions). In contrast, although the ecological validity of induced expressions might be constrained by experimental control, their ground truth can be directly validated through various experimental settings (e.g., preselected stimuli, self-reports, psychophysiological measurements). The controlled induction methods ensure researchers know precisely how the emotional expressions were elicited and what stimuli were involved, which may not be directly observable in interactional conditions. Accordingly, spontaneous expressions strike a balance between ecological and experimental validity. Therefore, while acknowledging the importance of investigating expressions in communicative contexts, this review focuses on spontaneous databases due to their methodological strengths in establishing ground truth and maintaining experimental control, which are crucial for drawing reliable conclusions about the relationships between emotional experiences and facial expressions.

Many FEDBs feature a decent number of encoders, with a relatively equal gender balance, although a notable focus on young adults as well as White/Caucasian and Asian encoders persists. For the training and testing of computer models, it will be important to collect large amounts of data from diverse demographics. Also, stimulus sets with higher temporal resolution (> 30 fps) are needed for capturing rapid facial movements.

Besides sender-relevant characteristics, it is important to note that not all emotions are equally easy to elicit using experimental methods. For example, anger may be difficult to induce in a controlled setting (Siedlecka & Denson, 2019), particularly when participants are aware that they are being filmed. In addition, facial behaviour obtained in the laboratory may be restricted in motion due to fixed camera positions. In the future, recording conditions could be less constrained by filming in natural environments (with multiple and hidden cameras) that allow for greater privacy, without compromising the experimental control in data acquisition (i.e., noise level, illumination, occlusion).

While induction materials are effective in eliciting the target emotion, it is possible that more than one emotion is felt by the encoder. Moreover, there may be considerable variability in how individuals appraise and respond to the stimulus content. At the moment, database validation approaches rely mainly on categorical emotion labels, which fail to capture subtle differences in cognitive and affective dimensions of emotion. Future research is needed to provide more fine-grained labels, thereby utilising both categorical and dimensional approaches to capture variability in emotional experiences (Cowen et al., 2021). This may also include meta-data such as audio and physiological signals (e.g., heart rate, skin conductance) for gaining complementary insights into emotional states (Jerritta et al., 2011; Juslin & Laukka, 2003), particularly when those are mixed.

Among all validation efforts, AFEA stands out as the primary method. The trend is likely stems from the overarching goal of many FEDBs to refine existing computer algorithms, with the potential to improve human-machine interaction. This focus has solidified AFEA as a reliable tool for facial expression recognition (Lewinski et al., 2014), often comparable to or even surpassing human performance (Del Líbano et al., 2018; Krumhuber et al., 2021; Lewinski et al., 2014). It should be noted though that their accuracy hinges on the integrity of the training data and the robustness of the underlying algorithms (Koelstra et al., 2010). Specifically, classifiers often train on specific segments of a database, reserving the rest for testing (Gupta et al., 2022). While this strategy may yield high recognition accuracy, it may also result in artificially inflated accuracy scores. This phenomenon highlights the need for benchmarking using standardised stimulus sets for training and validation (Valstar et al., 2011), thereby ensuring a more realistic approximation of real-life recognition. Also, most past efforts hinge on proprietary in-house algorithmic models which may not be easily accessible for cross-laboratory research (Dupré et al., 2020). In the future, it will be important to conduct cross-classifier and cross-corpus validations to allow for greater transparency in database assessment. The present article provides a first step in comparatively evaluating multiple spontaneous and dynamic FEDBs for basic emotion research, thereby highlighting their commonalities and differences.

For technical quality, many FEDBs nowadays provide consistent illumination and recording backgrounds, which signifies an improvement in their recording protocols compared to earlier databases (Bänziger et al., 2012; Krumhuber et al., 2017). While a considerable number of databases now offer medium to high-resolution recordings, there remains potential for further enhancement, particularly in achieving frame rates that surpass 30 fps to capture smoother temporal characteristics of facial expressions.

The integration of physiological responses offers a more comprehensive framework for emotion assessment and enables the detection of subtle emotional experiences (Bornemann et al., 2012). The inclusion of multimodal data (beyond audio and physiology) remains an evolving aspect of database construction. Similarly, modalities such as 3D face and body gestures hold promise as valuable sources for enhancing emotion recognition that 2D facial expressions alone cannot provide. The inclusion of 3D face meshes is also quite rare, which can offer a more accurate depiction of the depth, volume and dynamics of facial movements (Fanelli et al., 2010; Zhang et al., 2016). Such information can enhance emotion recognition, particularly under challenging conditions such as varying lighting or head poses (Pei et al., 2021; Malawski et al., 2014). In the future, more work could be done to capture other types of emotional information, such as hand gestures (Mahmoud et al., 2011) and thermal images (Nguyen et al., 2014).

3.6.2 Outlook and Future Research

The demand for additional spontaneous FEDBs continues to be an open discourse in psychology and affective computing. Existing databases have undeniably advanced our understanding of human affective states and bolstered various applications relevant to clinical research, security, and education. Nevertheless, the necessity for a comprehensive and systematic review of these databases remains indispensable.

This paper aimed to present an overview of 62 publicly available DBs in facial expression research, thereby illustrating the extensive selection available to researchers. Besides a few widely known databases, we introduced a range of databases that may not have received equivalent attention in prior work. These FEDBs offer valuable insights into spontaneous emotional experiences that researchers may seek to study in more detail in the future. In this context, more interactive tasks such as interpersonal discussions (Douglas-Cowie et al., 2011), gameplay (Saha et al., 2019), or personal events triggering (Nandrino et al., 2019) could be used to elicit expressions that closely mimic those found in real-life scenarios. By utilising interactive tasks alongside passive induction, researchers are able to cultivate more realistic and diverse datasets. This methodological integration allows for an in-depth exploration of how recognition of spontaneous expression functions, examining the diverse ways in which emotion might be expressed across various contexts and individuals. It also sheds light on the complex mechanisms that govern our capacity to accurately perceive these expressions. Such insights stand to refine the role that spontaneous expressions play in natural human communication.

While the self-report has traditionally been used to validate facial expressions, future research could explore alternative validation methods, such as physiological measures or observer ratings, especially as the field moves towards more naturalistic settings. The use of wearable devices may be useful in capturing real-time emotional responses, offering additional details to understand emotional responses. Additionally, shifting from self-report to observer-based annotation could mitigate limitations tied to self-assessment, providing an unbiased view of facial expressions. Moreover, while the validation process often hinges on in-house techniques of AFEA, these techniques may lack broad generalisability. Open-science practices can help address this issue by promoting transparency and reproducibility of techniques. Facial expression recognition analysis (FERA) challenges represent an important milestone in testing new DBs and allow for a direct comparison of different machine algorithms, helping to benchmark effective techniques and build communality across various research laboratories.

As such, implementing robust and diverse elicitation techniques, coupled with reliable annotation methods and the adoption of standardised evaluation methods, is of paramount importance. In addition to visual stimuli, additional methods could include narrated scenarios

(Li et al., 2013), interactive tasks (Cavicchio & Poesio, 2012; Rehg et al., 2013), or controlled interviews (Zhang et al., 2014) that can generate both basic and complex emotions. Such implementation could capture meaningful variations in spontaneous facial expressions while minimising artificially generated responses.

Future research could benefit from a more extensive consideration of the encoder age. While recent studies have shown increased interest in children's facial expressions (Khan et al., 2019; Littlewort et al., 2011), there remains a notable lack of representation for elderly individuals in many spontaneous FEDBs. Incorporating a more diverse age range among encoders could enhance our understanding of age-related variability in facial expressions across different populations.

Additionally, the exploration of emotional complexity within videos warrants further attention. Real-life emotional experiences often involve a mix of emotions, and a single video clip may capture several distinct emotional expressions. However, many FEDBs primarily assign a single emotion label per video. Future work could aim to identify and analyse these complex emotional experiences by considering the co-occurrence of emotional expressions.

Lastly, while some databases have endeavoured to capture expressions in specific contexts (e.g., driving, education; Bian et al., 2019; Li et al., 2020), a greater contextual variance in databases would be desirable. Furthermore, most FEDBs contain stimuli with frontal views of the face, thereby limiting their applicability in real-world scenarios where profile views are often encountered (Matsumoto & Hwang, 2011).

In conclusion, this paper provides valuable information for the field of facial expression research by systematically reviewing existing spontaneous and dynamic facial expression databases. The meticulous examination of 62 databases illuminates the depth and breadth of available resources, thereby empowering researchers to make informed decisions. Such endeavour optimises the use of existing resources and advocates for a shift in focus from creating new databases to exploiting existing data. Furthermore, the review highlights gaps in the literature, thereby paving the way for more diverse and realistic portrayals of emotional behaviours. As such, it lays the foundation for future research by promoting further advancements in the field of facial expression research.

Table 3.1. Conceptual features of 61 spontaneous datasets of dynamic facial expressions

Database	Expression		Encoder demographics				Measurement and validation					
	Emotions	Elicitation	Ethnicity	Age	Gender	N	Annotation	FACS coding	Emotion recognition	Type of rating	Self- report	Physiology
4DFAB	6 Basic emotions	Videos	Caucasian, Asian, Hispanic	5 to 75	120 M, 60 F	180	-	-	AFEA	Discrete emotions	-	-
4DME	Positive, negative, surprise, repression, others	Movies	Caucasian, Asian, Hispanic	22 to 57	-	56	Emotions, AU, valence	Y	AFEA	-	-	-
AM-FED	Amusement	Super Bowl commercials	Worldwide	>13	-	-	AU. Smile	Y	AFEA	AUs, smile	Liking, familiarity, rewatchability	-
AM-FED+	Various emotions	Video ads	Worldwide	21 to 82 (*M* = 47.6)	-	416	AU, Smile	Y	AFEA	AUs, smile	Liking, familiarity	-
AMHUGE	Amusement, neutral	Documentary, movie trailer, comedy show, YouTube video	Italian	18 to 54	27 M, 9 F	36	Valence, Arousal	Y	AFEA	AUs	Pleasure, arousal, dominance	EDA, BVP, temperature
BAUM-1	6 basic emotions, boredom, contempt, confusion, thinking, concentration, bothered, neutral	Movies, video ads, TV shows, IAPS images	Turkish	19 to 65	18 M, 13 F	31	Emotions	-	AFEA	Discrete emotions	-	-
BINED	6 basic emotions, frustration	Films, wire tracking task, touching objects	Irish, Peruvian	>18	137M, 119F	256	Emotions, valence	-			Emotion	-
BioVid Heat Pain	Pain, happiness, sadness, anger, disgust, fear, valence, arousal	Thermode (pain), films, IAPS images	German	18 to 65	45M, 45F	90	Expressive-ness	Y	AFEA	Pain	Emotion, valence, arousal	SCL, ECG, EEG, EMG
BioVid Emo	Anger, disgust, fear, amusement, sadness	Films	German	18 to 65	44M, 50F	86 (ex)	-	-			Emotion, valence, arousal	SCL, ECG, tEMG
BNU-LSVED	Happiness, surprise, disgust, confidence, puzzlement, frustration, neutral	Teaching videos	Chinese	18 to 31	19M, 64F	83	Emotions	-	-	-	Emotion, valence, arousal	-

(continued)

Table 3.1. (continued)

Database	Expression		Encoder demographics				Measurement and validation					
	Emotions	Elicitation	Ethnicity	Age	Gender	N	Annotation	FACS coding	Emotion recognition	Type of rating	Self- report	Physiology
BP4D-Spon	Happiness/amusement, sadness, surprise/startle, anger/upset, fear/nervous, disgust, embarrassment, pain	Various tasks	Asian, Black, white, Hispanic	18 to 29	18M, 23F	41	AUs	Y	AFEA, human observers	Discrete emotions, AUs	Emotion	-
BP4D+	happiness/amusement, sadness, anger, disgust, startle/surprise, joyful surprise, skeptical, fear/nervous, embarrassment, pain	Various tasks	Asian, Black, White, Hispanic, others	18 to 66	58M, 82F	140	AUs	Y	AFEA	Discrete emotions, AUs	Emotion	ECG, EDA, RSP, BVP
BU-EEG	Pain, neutral	Cold pressor task, meditation	Asian, Mid-Eastern, White, Hispanic, other	18 to 38	22M, 7F	29	Pain, relax, AUs	Y	AFEA	Pain, neutral	Pain	EEG
CAL	Confusion, interest, surprise, boredom, happiness, annoyance, neutral, other	Card matching task	Various ethnicities	21 to 32	3M, 5F	8	Emotions	-	Human observers	Discrete emotions	Emotion	-
CAM3D	Anger, surprise, thinking, concentration, unsure, confusion, triumph, frustration, boredom, neutral	Computer maze, riddles, word repetition, flickering screen/lights	Caucasian, Asian, Middle Eastern	24 to 50 (M = 27)	3M, 4F	7 (ex)	Emotions	-	Human observers	Discrete emotions	-	-
CASME	Amusement, sadness, disgust, surprise, contempt, fear, repression, tense	Videos	Chinese	M = 22.03	22M, 13F	35	Emotions, AUs	Y	AFEA	Discrete emotions	Emotion	-
CASME-II	Happiness, disgust, surprise, repression, others	Video ads, films	Chinese	M = 22.03	22M, 13F	26(ex)	Emotions, AUs	Y	AFEA	Discrete emotions	Emotion	-

(continued)

Table 3.1. (continued)

Database	Expression		Encoder demographics				Measurement and validation					
	Emotions	Elicitation	Ethnicity	Age	Gender	N	Annotation	FACS coding	Emotion recognition	Type of rating	Self- report	Physiology
CAS(ME)2	Positive (happiness), negative (disgust, anger), surprise, others	Video ads, films	Chinese	19 to 26 (*M* = 22.6)	9M, 13F	22	Emotions, AUs	Y	AFEA	Discrete emotions	Emotion	-
CAS(ME)3	6 basic emotions, others	Video ads, films/movies, mock interview	Chinese	*M* = 22.7	112M, 135F	247	Emotions, AUs	Y	AFEA	Discrete emotions	Emotion	EDA, ECG, RSP, PPG
DEAP	Valence, arousal	Music videos	European	19 to 37 (*M* = 26.9)	16M, 16F	22 (ex)	-	-			Valence, arousal, dominance. Liking, familiarity	EEG, EMG, GSR, BVP, temperature
DECAF	Amusement/happiness /funniness, excitement, anger, disgust, fear, sadness, shock	Movies, music videos	-	*M* = 27.3	16M, 14F	30	-	-	AFEA	Discrete emotions	Valence, arousal	MEG, hEOG, ECG, tEMG
DEFE	Happiness, anger, neutral	Videos	Chinese	19 to 56 (*M* = 27.3)	47M, 13F	60	-	-	AFEA	Discrete emotions, valence, arousal, dominance, AUs	Emotion	-
DISFA	Joy, disgust, fear, sadness, surprise	YouTube videos	White, Black, Hispanic, Asian	18 to 50	15M, 21F	27	AUs	Y	AFEA	AUs	-	-
DSME-3D	Happiness, anger, sadness, surprise, disgust, others	Films	Chinese	*M* = 25.6	12M, 21F	33 (ex)	Emotions, AUs	Y	AFEA	Discrete emotions	Emotion	-
DynEmo	Cheerfulness, curiosity, pride, astonishment, disgust, moved, boredom, fright, annoyance, shame, neutral	Videos, video ads, pictures, defective software, positive feedback, false belief	Caucasian	25 to 65 (*M* = 48)	176M, 182F	358	-	-	Human observers	Discrete emotions	Emotion, action readiness, valence, arousal	-

(continued)

Table 3.1. (continued)

Database	Expression		Encoder demographics				Measurement and validation					
	Emotions	Elicitation	Ethnicity	Age	Gender	N	Annotation	FACS coding	Emotion recognition	Type of rating	Self- report	Physiology
EB+	Happiness/amusement, sadness, anger, disgust, startle/surprise, joyful surprise, skeptical, fear/nervous, embarrassment, pain	Various tasks	Asian, Black, White, Hispanic, others	18 to 66	82M, 118F	200	AUs	Y	AFEA	Discrete emotions, AUs	Emotion	EEG, BVP, HR, EDA, SKT, ACC, GYRO
Emognition	Amusement, anger, awe, disgust, enthusiasm, fear, liking, sadness, surprise, neutral	Films	Polish	19 to 29 (*M*=22.4)	22M, 21F	43	-	Y	AFEA	Discrete emotions, AUs	Emotion, valence, arousal, motivation	EMG
EmoPain	Pain	Physical exercises	Caucasian, Black, Asian	*M* = 43.8	21M, 29F	-50	Pain	-	AFEA	Pain	-	-
Faces of Pain	Pain	Cold pressor task		-	6M, 20F	26	AUs	Y	AFEA, human observers	Genuine-ness, AUs	-	-
FEEDTUM	6 basic emotions, neutral	Films	White/ Caucasian	-	-	19		-	AFEA, human observers	Discrete emotions	-	-
HUMAINE	Various emotions	Various tasks	Various ethnicities	-	-		Emotions	-	-	-	Various states	ECG, CSR, skin temperature, breathing
ISED	Happiness, surprise, sadness, disgust	Movies	Indian	18 to 22	29M, 21F	50	Emotions	-	AFEA	Discrete emotions	Emotion	-
KTFE	6 basic emotions, neutral	Videos, game	Thai, Japanese, Vietnamese	11 to 32	16M, 10F	26	Emotions	-	AFEA	Discrete emotions	Emotion	-
KTFEv2	6 basic emotions, neutral		Thai, Japanese, Vietnamese	11 to 32	-	30	Emotions	-	AFEA	Discrete emotions	Emotion	-
LIRIS-CSE	Disgust, fear, happiness, sadness, surprise	Cartoon videos, movies	Various ethnicities	6 to 12 (*M* = 7.3)	5M, 7F	12	Emotions	-	AFEA, Human observers	Discrete emotions	-	-

(continued)

Table 3.1. (continued)

Database	Expression		Encoder demographics				Measurement and validation					
	Emotions	Elicitation	Ethnicity	Age	Gender	N	Annotation	FACS coding	Emotion recognition	Type of rating	Self- report	Physiology
MAHNOB -HCI	Amusement, disgust, fear, joy, sadness, neutral	Movies	Various ethnicities	19 to 40 (*M*=26.1)	13M, 17F	27 (ex)	-	-	-	-	Emotion, valence, arousal, dominance, predictability	EEG, ECG, GSR, RSP, skin temperature
MAHNOB -Laughter	Amusement	Videos	12 countries	*M* = 27.5	12M, 10F	22 (ex)	Laughter	-	AFEA	Laughter	-	-
MMEW	6 basic emotions, others	Videos	Chinese	22 to 35 (*M*=22.4)	-	36	Emotions, AUs	Y	AFEA	Discrete emotions	-	-
MMI-Part IV & V	Disgust, happiness	Cartoon videos, comedy shows, pictures	European, South American, Asian	20 to 32	13M, 12F	25	Emotions, AUs, laughter	Y	-	-	-	-
MUG	6 basic emotions, neutral	Videos	White/ Caucasian	20 to 35	51M, 35F	82 (ex)	-	-	-	-	-	-
NVIE	6 basic emotions	Films	Chinese	17 to 31	157M, 58F	315	Emotions, valence, arousal	-	AFEA	Discrete emotions, valence, arousal, genuine-ness	Emotion, valence, arousal	-
OL-SFED	Enjoyment, confusion, fatigue, distraction, neutral	Teaching videos	Chinese	17 to 26 (*M*=20.1)	29M, 53F	82 (ex)	Emotions	-	AFEA	Discrete emotions	Emotion	-
OPEN-EmoRec_II	Valence, arousal, neutral	IAPS images, Ulm picture set, card game	-	*M* = 44.3	7M, 23F	30 (ex)	Valence, expressive-ness	-	-	-	Valence, arousal, dominance	EEG, EMG, SCL, RSP, BVP
PDSTD	Sadness, neutral	Films	White/ Caucasian	18 to 33 (*M*=21.5)	24F	24	-	-	Human observers	Discrete emotions, valence, arousal, genuine-ness	Emotion	-

(continued)

Table 3.1. (continued)

Database	Expression		Encoder demographics				Measurement and validation					
	Emotions	Elicitation	Ethnicity	Age	Gender	N	Annotation	FACS coding	Emotion recognition	Type of rating	Self- report	Physiology
PEDFE	6 basic emotions	Films, video ads, YouTube videos, unpleasant smell, computer games	White/ Caucasian	20 to 30	-	56 (ex)	Emotions, AUs	Y	Human observers	Discrete emotions, genuine-ness	Emotion, genuineness	
PPB-Emo	6 basic emotions, neutral	Videos	Chinese	19 to 58 (*M* = 28.1)	31M, 9F	40 (ex)	-	-	-	-	Emotion, valence, arousal, dominance	EEG
RECOLA	Positive, negative, neutral	Film, comedy sketch, image	French, Italian, German, Portuguese	18 to 32 (*M* = 22)	19M, 27F	23 (ex)	Valence, arousal	-	-	-	Emotion, valence, arousal	ECG, EDA
RU-FACS	Various emotions	False opinion paradigm		-	-	100	AUs	Y	AFEA	AUs	-	-
SAMM	6 basic emotions, contempt	Movie, self-image	Various ethnicities	19 to 57 (*M* = 33.2)	16M, 16F	32	Emotions, AUs	Y	AFEA	AUs	-	
SAMM long	6 basic emotions, contempt	Movie, self-image	Various ethnicities	19 to 57 (*M* = 33.2)	16M, 16F	32	Emotions, AUs	Y	AFEA	AUs	-	-
SASE-FE	Anger, disgust, happiness, sadness, surprise, contempt	YouTube videos	White/ Caucasian, Asian, African	19 to 36	32M, 22F	54	-	-	AFEA	Discrete emotions	-	
SEWA	Disgust/distress, pleasure/liking, confusion, disliking, interest/boredom	Video ads	British, German, Hungarian, Greek, Serbian, Chinese	18 to 65	201M, 197F	398 (ex)	-	-	-	-	Emotion	
SHSD	Sadness	Movies	White/ Caucasian	20 to 22	6M, 12F	18	Emotions	Y	AFEA	Discrete emotions, AUs	-	

(continued)

Table 3.1. (continued)

Database	Expression		Encoder demographics				Measurement and validation					
	Emotions	Elicitation	Ethnicity	Age	Gender	N	Annotation	FACS coding	Emotion recognition	Type of rating	Self- report	Physiology
SIE-Intensity	Anger, happiness, sadness	Videos	Japanese	~ 20	9M, 1F	10	Emotions	-	AFEA, human observers	Discrete emotions	Emotion	-
SMIC	Disgust, fear, happiness, sadness, surprise	Movies	White/ Caucasian, Asian	22 to 34 (*M* = 28.1)	10M, 6F	16 (ex)	Emotions	-	AFEA	Discrete emotions, valence	Emotion	-
SPOS	6 basic emotions	Films	White/ Caucasian, Asian	-	4M, 3F	7	Emotions	-	AFEA	Valence, genuine-ness	Emotion	-
UNBV-McMaster	Pain	Shoulder rotation tasks	White/ Caucasian	-	63M, 66F	129	Pain, AUs	Y	AFEA	Pain, AUs	Pain	-
UT-Dallas	6 basic emotions, boredom, disbelief, laughter, puzzlement, neutral	Movies, TV programs	Asian, Black, White, Hispanic, Other	18 to 25	76M, 208F	284	Emotions	-	-	-	-	-
UvA-NEMO smile	Enjoyment	Videos	White/ Caucasian	8 to 76	215M, 185F	357 (ex)	Smiles	-	AFEA	Genuine-ness	-	-
Video-fNIRS	Happiness, disgust, neutral	Facial expression videos, IAPS images	-	*M* = 25	10M, 6F	16	Emotions	-	AFEA	Discrete emotions	Valence, arousal	EEG, fNIRS, GSR, BVP, RSP
Vine-reactor	Amusement	Comedy videos	-	-	-	222 (ex)	AUs	Y	AFEA	AUs	Emotion	-

Note. Key descriptions: N: (ex) = exclusions. Emotion recognition: AFEA = Automated facial expression analysis. Unspecified emotion categories from the original articles are noted as “various emotions”.

Table 3.2. Technical features of 61 spontaneous datasets of dynamic facial expressions

Database	Stimulus features							Recording					Additional features
	N (Stimulus)	Duration	Frame rate	Resolution	Modality	Sound	Speaking	Viewpoint	Visible	Setting	Background	Lighting control	
4DFAB	1.8M frames	-	60 fps	1200 × 1600, 640 × 480	V	-	-	Frontal	HD	Lab	Plain	Y	3D+4D, Kinect RGB and depth videos, posed expressions
4DME	1560	<0.5 to 9.82s	60 fps	1200 × 1600, 640 × 480	V	-	-	Multiple	HD	Lab	Plain	Y	2D+4D, Kinect RGB and depth videos; instruction to hide true feelings and maintain neutral face
AM-FED	242	~60s	14 or 15 fps	320 × 240	V	-	-	Frontal	HD, UB	Online	Non-plain	-	Measures of head scales, pose, and illumination; annotation of gender; webcam recordings
AM-FED+	1044	~28s	14 fps	320 × 240	V	-	-	Frontal	HD, UB	Online	Non-plain	-	Annotation of gender, age, country; webcam recordings
AMHUGE	-	25 to 117s	30 fps	1024 × 768	V	-	-	Frontal	HD, UB	Lab	Plain	Y	Kinect RGB video and depth sequences
BAUM-1	1184	0.43 to 9.34s (*M* = 1.82)	30 fps	576 × 720	V, A, AV	Y	Y	Multiple	HD	Lab	Plain	Y	Also contains posed speech (BAUM-1a)
BINED	1400	5 to 180s	-	720 × 576, 1920 × 1080	V	-	-	Frontal	HD, UB	Lab	Non-plain	-	Some emotion self-reports are open-ended
BioVid Heat Pain	>8700	5.5s	25Hz	1388 × 1038	V	-	-	Multiple	HD	Lab	plain	Y	Measures of head pose; Kinect RGB and depth videos; also contains posed expressions
BioVid Emo	-	32 to 245s (*M* = 68)	25Hz	1388 × 1038	V	-	-	Multiple	HD	Lab	Plain	Y	Kinect RGB and depth videos
BNU-LSVED	1572	-	-	40 × 40	IS	-	-	Multiple	HD	Class-room	Non-plain	-	-
BP4D-Spontaneous	328	~1 min	25 fps	1040 × 1392	V, AV	Y	Y	Frontal	HD	Lab	Plain	Y	2D+3D videos; measures of head pose; observer confidence ratings
BP4D+	1.4M frames	-	25 fps	1040 × 1392	V, AV	Y	Y	Frontal	HD	Lab	Plain	Y	2D+3D+thermal videos; measures of head pose
BU-EEG	29	~90s	24 fps	-	V	-	-	Frontal	HD	Lab	Non-plain	-	Also contains posed expressions

(continued)

Table 3.2. (continued)

Database	Stimulus features							Recording					Additional features
	N (Stimulus)	Duration	Frame rate	Resolution	Modality	Sound	Speaking	Viewpoint	Visible	Setting	Background	Lighting control	
CAL	< 247	0.4 to 15.9s	30 fps	320 ✕ 240	V	-	-	Frontal	HD	Lab	Plain	-	Emotion self-reports are also open-ended; annotation of confusion, interest, boredom, annoyance
CAM3D	108	*M* = 6s	30 fps	720 ✕ 576	AV	Y	Y	Frontal	HD, UB	Lab	Non-plain	-	Kinect RGB and depth videos, analysis of hand-over-face gestures; annotation of mental states; emotion observer judgements are open-ended
CASME	195	<250ms to 4mins	60 fps	1280 ✕ 720, 640 ✕ 480	V	-	-	Frontal	HD	Lab	Plain	-	Instruction to hide true feelings and maintain neutral face
CASME-II	247	<250ms to 2.5mins	200 fps	640 ✕ 480	V	-	-	Frontal	HD	Lab	Plain	Y	Instruction to hide true feelings and maintain neutral face
CAS(ME)2	444	<500ms to 4s	30 fps	640 ✕ 480	V	-	-	Frontal	HD	Lab	Plain	Y	Instruction to hide true feelings and maintain neutral face
CAS(ME)3	6107	<500ms to 8s	30 fps	1280 ✕ 720	V, A	-	-	Frontal	HD	Lab	Plain	Y	Instruction to hide true feelings and maintain neutral face/suppress facial movements; voice analysis; Kinect RGB and depth videos
DEAP	-	1 min	50 fps	-	V, AV	-	-	Frontal	HD	Lab	Plain	Y	-
DECAF	-	51.1 to 128.2s	20 fps	-	V, AV	-	-	Frontal	HD	Lab	plain	Y	Near-infrared facial videos
DEFE	164	30s	30 fps	1920 ✕ 1080, 648 ✕ 480	V	-	-	Frontal	HD	Lab	Non-plain	Y	Emotion elicitation before task recording
DISFA	130k frames	4mins	20 fps	1024 ✕ 768	V	-	-	Frontal	HD	Lab	Plain	Y	-
DSME-3D	373	<260ms to 500ms	60 fps	848 ✕ 480	V	-	-	Frontal	HD	Lab	Plain	Y	2D+3D, Kinect RGB and depth videos; instruction to hide true feelings and maintain neutral face/suppress facial movements
DynEmo	358	1 to 15mins	25 fps	768 ✕ 576	V, AV	Y	Y	Multiple	HD, UB	Lab	Plain	Y	Hidden camera

(continued)

Table 3.2. (continued)

Database	Stimulus features							Recording					Additional features
	N (Stimulus)	Duration	Frame rate	Resolution	Modality	Sound	Speaking	Viewpoint	Visible	Setting	Background	Lighting control	
EB+	1261	44s	25 fps	1040 × 1392	V, AV	Y	Y	Frontal	HD	Lab	Plain	Y	-
Emognition	430	49s to 2 mins	60 fps	1920 × 1080	V	-	-	Frontal	HD, UB	Lab	Plain	Y	-
EmoPain	-	-	58 fps	1024 × 1024	AV	-	-	Multiple	HD, FB	Lab	Non-plain	Y	Chronic pain patients and healthy controls; annotation of body movements; motion capture
Faces of Pain	26	60s	-	-	V	-	-	Frontal	HD	Lab	Plain	Y	Also contains posed pain expressions
FEEDTUM	399	30 to 300s	25 fps	640 × 480	IS	-	-	Frontal	HD	Lab	Plain	Y	-
HUMAINE	50	3s to 2mins	-	-	AV	Y	Y	Multiple	HD, UB, FB	Mixed	Non-plain	-	Annotation of gender, modality, context; speech/language, gesture
ISED	428	1 to 10s	50 fps	1920 × 1080	V	-	-	Frontal	HD	Lab	Plain	Y	Hidden camera
KTFE	-	-	5 fps	-	V	-	-	Frontal	HD	Lab	plain	Y	Also contains static images and thermal videos
KTFEv2	-	-		-	V	-	-	Frontal	HD	Lab	Plain	Y	-
LIRIS-CSE	208	*M* = 5s	25 fps	720 × 480, 800 × 600, 1920 × 1080	V	-	-	Frontal	HD, UB	Mixed	Plain, non-plain	-	-
MAHNOB-HCI	-	34.9 to 117s	60 fps	780 × 580	V, A	-	-	Multiple	HD, UB	Lab	Plain	Y	Eye gage
MAHNOB-Laughter	563	*M* = 1.65min	25 fps	720 × 576	V, A, AV	-	-	Frontal	HD	Lab	Non-plain	Y	Also contains posed smiles, posed laughter, speech, thermal videos
MMEW	300	-	90 fps	1920 × 1080	V	-	-	Frontal	HD	Lab	Plain	-	Instruction to hide true feelings and maintain neutral face
MMi-Part IV & V	392	-	-	640 × 480	AV	Y	Y	Frontal	HD	Lab	Plain, non-plain	-	-

(continued)

Table 3.2. (continued)

Database	Stimulus features							Recording					Additional features
	N (Stimulus)	Duration	Frame rate	Resolution	Modality	Sound	Speaking	Viewpoint	Visible	Setting	Background	Lighting control	
MUG	66139 images	-	19 fps	896 × 896	IS	-	-	Frontal	HD	Lab	Plain	Y	Also contains posed expressions
NVIE	-	3 to 4 mins	30 fps	704 × 480	IS	-	-	Frontal	HD	Lab	Plain	Y	Also contains posed expressions, thermal videos
OL-SFED	1274	1 to 8s	30 fps	1280 × 720	V	-	-	Frontal	HD	Class-room	Plain	-	Webcam recordings
OPEN-EmoRec_II	-	6s	-	-	V, AV	-	-	Frontal	HD		-	Y	-
PDSTD	72	30s	30 fps	1920 × 1080	V	-	-	Frontal	HD	Lab	Non-plain	-	-
PEDFE	707	*M* = 3.0s	30 fps	-	AV	-	-	Frontal	HD, UB	Lab	Plain, Non-plain	-	Also contains posed expressions and static images; hidden camera
PPB-Emo	240	30s	30 fps	1920 × 1080	AV	-	-	Multiple	HD, UB	Lab	Non-plain	-	Body gesture, thermal videos; emotion elicitation before task recording
RECOLA	-	3.5mins	25Hz	1080 × 720	AV	Y	Y	Frontal	HD	Lab	Non-plain	Y	Annotation of social behaviors
RU-FACS	-	2.5mins	-	-	AV	Y	Y	Multiple	HD	Lab	Plain	-	-
SAMM	338	<0.7s	200 fps	2040 × 1088	V	-	-	Frontal	HD	Lab	Plain	Y	Instruction to hide true feelings and maintain neutral face/suppress emotion
SAMM long	147	*M* = 35.3s	200 fps	2040 × 1088	V	-	-	Frontal	HD	Lab	Plain	Y	Instruction to hide true feelings and maintain neutral face/suppress emotion
SASE-FE	~ 324	3 to 4s	100 fps	1280 × 960	V	-	-	Frontal	HD	Lab	Plain	-	Also contains posed expressions
SEWA	1592	~1min	20 to 30 fps	320 × 240, 640 × 360	V, A, AV	-	-	Frontal	HD	Online	Non-plain	-	Annotation of head gesture, hand gesture, mimicry, liking, agreement
SHSD	-	-	100 fps	640 × 480	V	-	-	Frontal	HD	Lab		Y	Hidden camera

(continued)

Table 3.2. (continued)

Database	Stimulus features							Recording					Additional features
	N (Stimulus)	Duration	Frame rate	Resolution	Modality	Sound	speaking	Viewpoint	Visible	Setting	Background	Lighting control	
SIE-Intensity	30	2 to 4mins	5 fps	-	V	-	-	Frontal	HD, UB	Lab	Plain	Y	Also contains thermal videos
SMIC	164	~0.5s	25, 100 fps	640 ✕ 480	V	-	-	Frontal	HD	Lab	-	Y	Also contains near-infrared facial videos; Instruction to hide true feelings and maintain neutral face/suppress emotion
SPOS	147	*M* = 13s	25 fps	640 ✕ 480	V	-	-	Frontal	HD	Lab	-	Y	Also contains posed expressions, near-infrared facial videos
UNBV-McMaster	200	-	-	320 ✕ 240	V	-	-	Multiple	HD	Lab	Plain	-	Shoulder pain patients
UT-Dallas	-	5s	30 fps	720 ✕ 480	V	-	-	Frontal	HD	Lab	Plain	Y	-
UvA-NEMO smile	597	*M* = 3.9s	50 fps	1920 ✕ 1080	V	-	-	Frontal	HD	Lab	Plain	Y	Also contains posed expressions
Video-fNIRS	-	-	-	-	V	-	-	Frontal	HD	Lab	plain	-	-
Vinereactor	3455	2 to 7s	-	320 ✕ 240	V	-	-	Frontal	HD	Online	Non-plain	-	Webcam recordings

Note. Key descriptions: Stimuli: V = video; A = audio; AV = audio-visual. Visible elements: HD = head; UB = upper body; FB = full body.

Table 3.3. Practical features of 61 spontaneous datasets of dynamic facial expressions

Database	Practical Information			
	Website	Email address	Access	Payment
4DFAB	https://ibug.doc.ic.ac.uk/resources/4dfab/	m.pantic@imperial.ac.uk	Email (not yet available)	-
4DME	-	xiaobai.li@oulu.fi	EULA form	-
AM-FED	https://www.affectiva.com/facial-expression-dataset-/	amfed@affectiva.com	EULA form	-
AM-FED+	https://www.affectiva.com/facial-expression-dataset-/	amfed@affectiva.com	EULA form	-
AMHUGE	http://amhuse.phuselab.di.unimi.it/	phuselab@di.unimi.it	EULA form	-
BAUM-1	http://archive.ics.uci.edu/ml/datasets/BAUM-1	s.zhalehpour@gmail.com; cigdem.erogluerdem@gmail.com	Email	-
BINED	-	g.mckeown@qub.ac.uk	Email	-
BioVid Heat Pain	https://www.nit.ovgu.de/BioVid.html	sascha.gruss@uni-ulm.de	EULA form	-
BioVid Emo	https://www.nit.ovgu.de/BioVid.html	sascha.gruss@uni-ulm.de	EULA form	-
BNU-LSVED	-	hejun@bnu.edu.cn	Email	-
BP4D-Spontaneous	https://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html	lijun@cs.binghamton.edu	Email	$400
BP4D+	https://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html	lijun@cs.binghamton.edu	Email	$1,250
BU-EEG	https://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html	lijun@cs.binghamton.edu	Email	$500
CAL	-	Peter.Robinson@cl.cam.ac.uk	Email	-
CAM3D	https://www.repository.cam.ac.uk/handle/1810/291335	marwamm@gmail.com	Email	-
CASME	http://casme.psych.ac.cn/casme/e1	fuxl@psych.ac.cn	EULA form	-
CASME-II	http://casme.psych.ac.cn/casme/e2	fuxl@psych.ac.cn	EULA form	-
CAS(ME)2	http://casme.psych.ac.cn/casme/e3	fuxl@psych.ac.cn	EULA form	-
CAS(ME)3	http://casme.psych.ac.cn/casme/e4	fuxl@psych.ac.cn	EULA form	-
DEAP	https://www.eecs.qmul.ac.uk/mmv/datasets/deap/	i.patras@qmul.ac.uk	EULA form	-

(continued)

Table 3.3. (continued)

Database	Practical Information			
	Website	Email address	Access	Payment
DECAF	https://decaf-dataset.github.io/	decaf.mhug@gmail.com	EULA form	-
DEFE	-	liwenbo@cqu.edu.cn	Email	-
DISFA	http://mohammadmahoor.com/disfa/	mmahoor@du.edu	EULA form	-
DSME-3D	-	fengpingwang@stu.xjtu.edu.cn	Email (not yet available)	-
DynEmo	https://dynemo.univ-grenoble-alpes.fr/?page=inscription	anna.tcherkassof@upmf-grenoble.fr	Email	-
EB+	https://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html	lijun@cs.binghamton.edu	Email	$500
Emognition	https://github.com/Emognition/Emognition-wearable-dataset-2020	emotions@pwr.edu.pl	EULA form	-
EmoPain	https://wangchongyang.ai/EmoPainChallenge2020/	n.berthouze@ucl.ac.uk	Email	-
Faces of Pain	-	marni@salk.edu	Email	-
FEEDTUM	https://www.jade-hs.de/team/frank-wallhoff/databases/	frank.wallhoff@jade-hs.de	EULA form	-
HUMAINE	-	g.mckeown@qub.ac.uk	EULA form	-
ISED	https://sites.google.com/site/iseddatabase/	iseddatabase@gmail.com	EULA form	-
KTFE	-	hungnv@hcmue.edu.vn	Email	-
KTFEv2	-	hungnv@hcmue.edu.vn	Email	-
LIRIS-CSE	https://childrenfacialexpression.projet.liris.cnrs.fr	rizwan.khan@bhu.edu.pk	EULA form	-
MAHNOB-HCI	https://mahnob-db.eu/hci-tagging/	hci_tagging_db@mahnob-db.eu	EULA form	-
MAHNOB-Laughter	https://mahnob-db.eu/laughter/	laughter_db@mahnob-db.eu	EULA form	-
MMEW	https://faculty.sdu.edu.cn/_resources/group1/M00/00/3C/CgECYWNKGs2AS7aAAALQs4B2n0w884.pdf	benxianyeye@163.com	Email	-
MMI-Part IV & V	https://mmifacedb.eu/	mmi_face_db@mahnob-db.eu	EULA form	-
MUG	http://mug.ee.auth.gr/fed/	adelo@eng.auth.gr	EULA form	-

(continued)

Table 3.3. (continued)

Database	Practical Information			
	Website	Email address	Access	Payment
NVIE	http://nvie.ustc.edu.cn	sfwang@ustc.edu.cn	EULA form	-
OL-SFED	-	luweigang@ouc.edu.cn	Email	-
OPEN-EmoRec_II	-	harald.traue@uni-ulm.de	Email	-
PDSTD	https://osf.io/uyjeg/?view_only=24474ec8d75949ccb9a8243651db0abf	dkuester@uni-bremen.de	OSF	-
PEDFE	https://osf.io/cynsx/	alessio.miolla@phd.unipd.it	OSF	-
PPB-Emo	-	guogang@cqu.edu.cn	Email	-
RECOLA	https://diuf.unifr.ch/main/diva/recola/	diufFrecola@unifr.ch	EULA form	-
RU-FACS	https://inc.ucsd.edu/mplab/grants/project1/research/rufacs1-dataset.html	mbartlet@ucsd.edu	Email	-
SAMM	http://www2.docm.mmu.ac.uk/STAFF/m.yap/dataset.php	M.Yap@mmu.ac.uk	EULA form	-
SAMM long	http://www2.docm.mmu.ac.uk/STAFF/m.yap/dataset.php	M.Yap@mmu.ac.uk	EULA form	-
SASE-FE	https://icv.tuit.ut.ee/databases/	shb@icv.tuit.ut.ee	EULA form	-
SEWA	https://db.sewaproject.eu/	sewa@mahnob-db.eu	EULA form	-
SHSD	-	shb@ut.ee	Email	-
SIE-Intensity	-	namnp.khmt171@pg.hcmue.edu.vn	Email	-
SMIC	https://www.oulu.fi/en/university/faculties-and-units/faculty-information-technology-and-electrical-engineering/center-machine-vision-and-signal-analysis	Xiaobai.Li@oulu.fi	EULA form	-
SPOS	https://www.oulu.fi/en/university/faculties-and-units/faculty-information-technology-and-electrical-engineering/center-machine-vision-and-signal-analysis	Xiaobai.Li@oulu.fi	EULA form	-

(continued)

Table 3.3. (continued)

Database	Practical Information			
	Website	Email address	Access	Payment
UNBC-McMaster shoulder pain	http://www.jeffcohn.net/resources/	mer160@pitt.edu	EULA form	-
UT Dallas	https://labs.utdallas.edu/facelab/database/	otoole@utdallas.edu	EULA form	-
UvA-NEMO smile	http://www.uva-nemo.org/index.html	th.gevers@uva.nl	EULA form	-
Video-fNIRS	-	arman.savran@yasar.edu.tr		-
Vinereactor	http://vinereactor.org/	ek826@drexel.edu	Email (not yet available)	-

CHAPTER 4

Empirical Evaluation of Spontaneous and Dynamic Facial Expression Databases for Basic Emotion Research

4.1 Introduction

Facial expressions (FEs) are inherently spontaneous and dynamic, posing unique challenges in the field of emotion recognition. Despite their importance, most previous research on FE recognition has relied on posed, static images captured at peak intensity (Dawel et al., 2021; Krumhuber et al., 2013). While this approach offers high experimental control, it has led to criticisms for its dependence on exaggerated, stereotypical expressions (Matsumoto & Hwang 2017; Nelson & Russell, 2013), often contributing to high recognition accuracy observed in earlier studies (Calvo & Nummenmaa, 2016). This distinctiveness of posed expressions lies in their development nature, designed to depict a singular emotion, devoid of any mixed or contradictory signals. Such deliberate emphasis can lead to a stylised combination of facial features, potentially atypical in spontaneous emotional expressions (Carroll & Russell, 1997; Scherer & Ellgring, 2007).

With the growing interest in spontaneous FEs, numerous facial expression databases (FEDBs) have been developed in recent years, encompassing a wide spectrum of expressions depicting more realistic and ecologically valid facial behaviours (Dawel et al., 2021). Despite their higher ecological validity, these databases often showcase facial actions that are complex and subtle, amplifying the ambiguity of their emotional content (Cohn et al., 2007), thereby posing challenges for accurate emotion recognition from faces. While some studies have shown recognition accuracy of spontaneous expressions surpassing chance levels (Grimm et al., 2006; Krumhuber et al., 2020), the accuracy rates in most studies have been generally low, failing to achieve the level of accuracy of posed expressions (Dupré et al., 2019; Calvo & Nummenmaa, 2016).

The heterogeneous composition of spontaneous expressions presents significant challenges for their classification and interpretation (Berenbaum & Rotter, 1992; Kim & Sutharson, 2023; Pfister et al., 2011). The subtle facial configurations can also go unnoticed or are erroneously interpreted (Komori & Onish, 2015; Le Ngo et al., 2016; Sato & Yoshikawa, 2007), particularly when devoid of contextual indicators (Hess & Hareli, 2015). Consequently, such tendency in turn results in substantially decreased recognition accuracy previously ranging from 15% to 65% (Wagner, 1990; Kayyal & Russell, 2013).

4.1.1 Spontaneous facial expression databases

Despite the increasing availability of spontaneous FEDBs, their practical application in research remains somewhat limited. Many studies opt to test their own databases (e.g., Zhang et al., 2014; Saganowski et al., 2022), often bypassing the opportunity for comparative analysis with databases from other sources. This practice could potentially confine the generalisability of findings. Additionally, there is a noticeable inclination in decoding research to focus on only a few spontaneous FEDBs, generally around two to four databases (Chanti & Caplier, 2018; Reddy et al., 2019). This selective usage raises questions about the extrapolation and validity of findings concerning emotion recognition, as conclusions might be tightly bound to specific databases being used.

Spontaneous FEDBs offer an extensive array of features, encompassing a wide spectrum of emotions and their corresponding facial actions (Wang et al., 2013; Zhang et al., 2014). The richness of these databases captures not just the nuances present in FEs, but also provides a more realistic portrayal of diverse emotional experiences (Tcherkassof et al., 2013). However, the inherent complexity of these recordings collectively poses significant challenges in their accurate recognition and interpretation (Calvo & Nummenmaa, 2016; Saumure et al., 2018), highlighting the urgent need to explore what makes spontaneous FEs recognisable.

The meticulous construction of spontaneous FEDBs hinges on the elicitation techniques employed. A vast majority of FEDBs (over 90% of databases; Gross & Levenson, 1995; Schaefer et al., 2010) rely on emotionally evocative videos or images to induce genuine emotional responses. Other favoured techniques include immersive emotional tasks (Littlewort et al., 2007; Sneddon et al., 2012) and in-depth interviews (Zhang et al., 2014). The selection of elicitation methods is carefully considered to ensure that the expressions captured are both genuine and align with the emotion the database aims to represent (Gross & Levenson, 1995).

Once expressions are elicited, most FEDBs validate the emotional content of recordings by incorporating self-reports, wherein encoders reflect on their emotional state, providing a direct measure of the experienced emotion. Although insightful, this method is complemented by external assessment to ensure objectivity. The way FEDBs are externally validated varies from human observer evaluation to machine recognition, with a majority employing automated facial expression analysis (AFEA) for emotion and action unit (AU) recognition. Human observers, often unfamiliar with the recordings, are deemed more objective. This method capitalises on the human ability to perceive and interpret emotional expressions, albeit subject to individual variances (Krumhuber et al., 2021; Yitzhak et al., 2017). On the other hand, AFEA classifiers present a cost-effective approach for analysing large amounts of data. Their main advantage lies in delivering standardised, objective measurements across different datasets, significantly reducing the subjectivity and variability associated with human interpretation (Dupré et al., 2019).

The various methods adopted to elicit and validate spontaneous FEs significantly impact the quantity and quality of data central to emotion recognition research (Küster et al., 2020; Krumhuber et al., 2017). This issue, although pivotal, often goes underemphasised in the development of databases, with some databases missing out on the evaluation process. A prevailing trend reveals that many studies validated their databases via in-house AFEA classification, often without inter-platform comparative evaluations. While these classifiers demonstrate proficiency within their native training datasets, their reliability wanes when confronted with novel expressions, especially spontaneous FEDBs where displays are often heterogeneous (Yitzhak et al., 2017). This discrepancy risks overly optimistic results of a database's versatility and adaptability. The field's current trajectory lacks a cohesive normative standard that encapsulates the diversity of spontaneous FEDBs documented in the literature. Such lapses underline the imperative for standardised cross-corpus evaluations, fostering an environment where databases can be compared to each other. Instituting such a coordinated benchmark may help accelerate the progress in the field by guiding researchers, to review, compare, and contrast existing study findings on spontaneous FEs.

4.1.2 Previous works on cross-corpus evaluation

Earlier efforts in cross-corpus evaluation have predominantly centred on algorithmic advancements rather than the stimulus set itself (Zavarez et al., 2017), aimed to test the applicability of algorithms developed on one specific FEDB to others that differ in terms of various characteristics (e.g., elicitation method, recording condition etc). Collective evidence (Chen et al., 2020; Ryumina et al., 2020) argues that algorithms perform well within the same dataset used for training (intra-corpus), but their performance often degrades significantly when tested across different datasets (cross-corpus). In this vein, cross-corpus evaluation is crucial in determining the generalisability of classifiers and databases (Mayer et al., 2014;

Ryumina et al., 2022). Initiative challenges like Facial Expression Recognition Analysis (FERA; Valstar et al., 2015, 2017) and Emotion Recognition in the Wild (EmotiW; Dhall et al., 2023) have extended cross-corpus evaluation by testing models against a curated set of FEDBs, each with unique characteristics. These challenges have not only facilitated a direct comparison of FEDBs but also spurred advancements in algorithms to enhance cross-dataset performance.

Yet, cross-corpus research in FE recognition ventures beyond algorithmic and technical precision. The variability in database content and construction presents unique challenges. Many FEDB publications have indeed focused on validating their own databases or comparing recognition rates across a limited selection. However, only a few studies delved into the inherent variability and complexity of human expressions, aiming to understand how these factors impact recognition performance across diverse datasets.

In past efforts, a significant emphasis was placed on understanding the variability and challenges posed by spontaneous or "in-the-wild" databases. Studies have incorporated a range of 2 (Zhang et al., 2022) to 10 (Ryumina et al., 2022) different stimulus sets, encompassing both posed and spontaneous expression databases in both static and dynamic formats. The recognition rates varied widely, with general spans from as low as 30% (Zhang et al., 2021) to over 99% (Ryumian & Karpov; 2020). This performance range was heavily contingent on the testing databases and conditions under which the expressions were collected (i.e., posed or spontaneous). A recurrent trend across these studies is that emotion recognition involving spontaneous expressions often shows a noticeable drop in recognition rates, especially when algorithmic models were trained on posed datasets.

For dynamic FEDBs, Krumhuber and colleagues (2020) empirically demonstrated the considerable variability in recognition accuracy across 14 databases, corroborating the notion that spontaneous expressions are inherently more difficult to recognise. These expressions consistently yield lower recognition rates compared to posed ones, primarily due to their unpredictable and naturalistic presentation. The study further revealed that facial AU configurations in posed expressions are more prototypical, leading to higher classification accuracy. Similarly, Benitez-Quiroz and colleagues (2016) and Chen (2020) systematically compared the performance of various AFEA algorithms across FEDBs, showing significant variability in accuracy rates that depended on the specific database used for training. This variability was often linked to differences in emotional intensity and the presence of non-prototypical cues within spontaneous databases. Additionally, Zhang and colleagues (2022) conducted a cross-corpus evaluation specifically for micro-expression databases, highlighting the unique challenges these subtle and fleeting expressions present in generalising performance across different datasets.

The construction and annotation methodologies of databases heavily influence the performance of cross-corpus evaluation. For example, methodological rigour and diversity in annotation approaches, play a substantial role in the accurate recognition of across different datasets (Ryumina et al., 2020). Research further emphasised the bias inherent in spontaneous FEDBs, illuminating how the emotion-induction methods, the diversity of subjects, and the number of annotators contribute to biases that may affect the performance of algorithmic tools (Wang et al., 2012). Moreover, Ryumina and colleagues (2022) showed that databases with a broad range of emotion categories, including subtle and complex expressions, often present a greater challenge for recognition. The expertise of the annotator was also found to influence the AFEA performance (i.e., naïve vs trained coder). Static versus dynamic databases further escalate the varied performance, as AFEA systems often struggle with dynamic sequences. These past efforts not only shed light on the algorithmic models used for cross-corpus testing but also highlight the pressing need for considering diverse characteristics across databases.

Beyond the recognisability of databases, cross-corpus research has extended to various aspects of FEs, such as the level of interest (Yeasin et al., 2013), or perceived intensity (Krumhuber et al., 2020), as well as degrees of depression symptoms (Pampouchidou et al., 2020). Additionally, a segment of past works has opted for a meta-analytic approach or integrated reviews to scrutinise the algorithmic methodologies coupled with recognition rates across databases (Jia et al., 2020; Chen et al., 2020). Despite these diverse approaches, the integration of spontaneous and dynamic FEDBs into cross-corpus evaluation has been still sparse, often lacking dynamic quality. This deficiency may be traced back to an overreliance on proprietary in-house tools for automated classification, which are not widely disseminated within the research community. This restricted access, along with an absence of cross-corpus evaluation, risks compromising the research findings dependent on selected stimuli. This gap highlights a critical need within the field: a focused examination of how well spontaneous expressions across different corpora can be recognised and interpreted.

4.1.3 Present research

Building on the systematic review presented in Chapter 3, this study aims to empirically evaluate spontaneous and dynamic FEDBs through a cross-corpus analysis. By utilising the commercial software AFFDEX, we aim to standardise the comparative approach in testing spontaneous FEDB. This study ventures beyond the conventional constraints imposed by proprietary algorithmic tools and posed static expressions, which may inadequately represent the subtlety and variability inherent in human expressions. The restricted access, along with an absence of cross-corpus evaluation, risks compromising the research findings dependent on selected stimuli, potentially leading to an overly optimistic view of a database's applicability. Additionally, we seek to revisit the featural characteristics identified in Chapter 2 as significantly influencing the recognisability of spontaneous expressions.

Specifically, the current study integrates emotion classification with detailed AU analysis within the context of spontaneous expressions. According to the extensive review in Chapter 3, database evaluations often concentrate on either emotion or AU recognition, which might unintentionally obscure a holistic understanding of facial behaviour. Notably, a substantial portion of databases (about 60%) prioritise emotion recognition, while a smaller portion (approximately 30%) delves into detailed FACS analysis. This integrated approach, which merges the nominal categorisation of emotions with specific facial action patterns, aims to address the issue of oversimplification (Zhalehpour et al., 2016). It is also important to note that some databases reviewed in Chapter 3 lack an evaluation process which can limit their reliability and versatility.

Previous database evaluations typically concentrated on recognition performance without adequately investigating how intrinsic expression characteristics – prototypicality, ambiguity, complexity - impact the performance. While prior studies have tapped into these attributes (Kinchella & Guo, 2021; Matsumoto & Hwang, 2014), they typically do so separately, without integrating them into a cohesive framework that accounts for their sole impact or in combinations. In contrast, our research in Chapter 2 provides a comprehensive analysis of these featural characteristics within a unified framework, offering insights into their individual and combined contributions to recognition accuracy. Specifically, higher prototypicality and complexity improve recognition, whereas ambiguity has the opposite effect (Kim et al., 2023). However, given that these findings are primarily based on posed expressions, further investigation is needed to understand their influence on the recognition of spontaneous emotional displays. Furthermore, there has been a pronounced emphasis on controlled, posed expressions (or with a few spontaneous expressions), which may not accurately reflect the spontaneous variability encountered in daily life. This methodological oversight hampers the understanding of the efficacy of these parameters in spontaneous expression recognition.

In alignment with the previous approach, we focused on the classification of FEs portraying the six basic emotions, also a choice set by the capabilities of the AFFDEX software. Considering the widespread accessibility of numerous commercially available software tools, utilising these for cross-corpus evaluations can offer more practical and accessible methodologies (Cohn & Sayette, 2010). Drawing from previous research, we hypothesised that while spontaneous displays might present challenges in classification (Calvo & Nummenmaa, 2016; Yitzhak et al., 2017), the recognition accuracy would likely surpass mere chance level (Grimm et al., 2006; Krumhuber et al., 2020; Pfister et al., 2011); a finding which may be explained by the interplay of featural characteristics between prototypicality, complexity and ambiguity in spontaneous expressions (Kim et al., 2023). We predict that higher expression prototypicality and complexity would enhance recognition accuracy, whereas greater ambiguity is expected to reduce recognition accuracy. To the best of our knowledge, this is the first study that not only compares a large number of publicly available spontaneous FEDBs for recognition but also critically examines the influence of parameters solely focused on spontaneous expression recognition.

4.2 Method

4.2.1 Stimulus material

Spontaneous FEs in the form of video clips or image sequences were sourced from a collection of databases. This selection was rigorously guided by the in-depth review conducted in Chapter 2. Deviating from the broad selection criteria typically employed, our approach was specifically tailored to identify databases that had not only been thoroughly reviewed but were also readily available for empirical testing. We prioritized databases featuring three to six basic emotions, aligning with the analytical capabilities of AFFDEX. This selection process resulted in a final selection of 21 databases featuring three to six basic emotions for emotion and AU classification via AFFDEX (see Table 4.1). While most databases encompassed five to six basic emotions, 4DME, CAS(ME)2, and ISED were exceptions, featuring only three to four emotions.

Delving into the specifics of the chosen databases, expressions within databases were primarily elicited by leveraging video-inducing techniques, which involve the presentation of emotionally evocative video clips to induce emotions from participants (for details, please see Chapter 2). Such techniques have been empirically validated for their efficacy in inducing targeted emotions (Gross & Levenson, 1995). Beyond video stimuli, a few databases (BINED, BP4D, EB+) incorporated multiple elicitation techniques, including touching objects or other active engagement methods, further enriching the emotional repository.

The video recordings predominantly showcase a frontal view of encoders, maintaining consistency in head orientation and facial visibility. The majority of these recordings adhere to a frame rate ranging from 25 to 60 frames per second (fps), to capture the fluid dynamics of FEs. The duration of videos varied significantly between databases, reflecting the diverse nature of emotion elicitation and recording protocols employed. Moreover, the resolution quality amongst databases also varies, ranging from medium (640 x 480) to high (2040 x 1088) definitions. Conclusively, the video recordings mostly feature a plain background, coupled with controlled lighting conditions, to accentuate the visibility of faces.

In pursuing a consistent representation across varied databases, a systematic selection protocol was imperative. To this end, we incorporated stratified random sampling (lliyasu & Etikan al., 2021). This approach demarcates the overarching population into distinct, homogeneous subsets, referred to as ‘strata’. In the context of the current study, we judiciously undertook selections within the two strata of emotions and gender.

Consequently, we selected five portrayals of each emotion category from each gender (male and female) through a process of random sampling (e.g., Figure 4.1). This process yielded a range of 30 to 60 portrayals per database (10 videos per emotion, maintaining a balanced gender representation). Noteworthy exceptions were SAMM and LIRIS databases, which contained fewer than 10 portrayals for certain emotions (e.g., disgust). The selection process culminated in a total of 1,060 spontaneous and dynamic expressions, derived from 537 female and 522 male encoders. The duration of the selected stimuli ranged from a minimum of 0.5 seconds (SMIC; Li et al., 2013) to a maximum of 2.5 minutes (BioVid; Zhang et al., 2016).

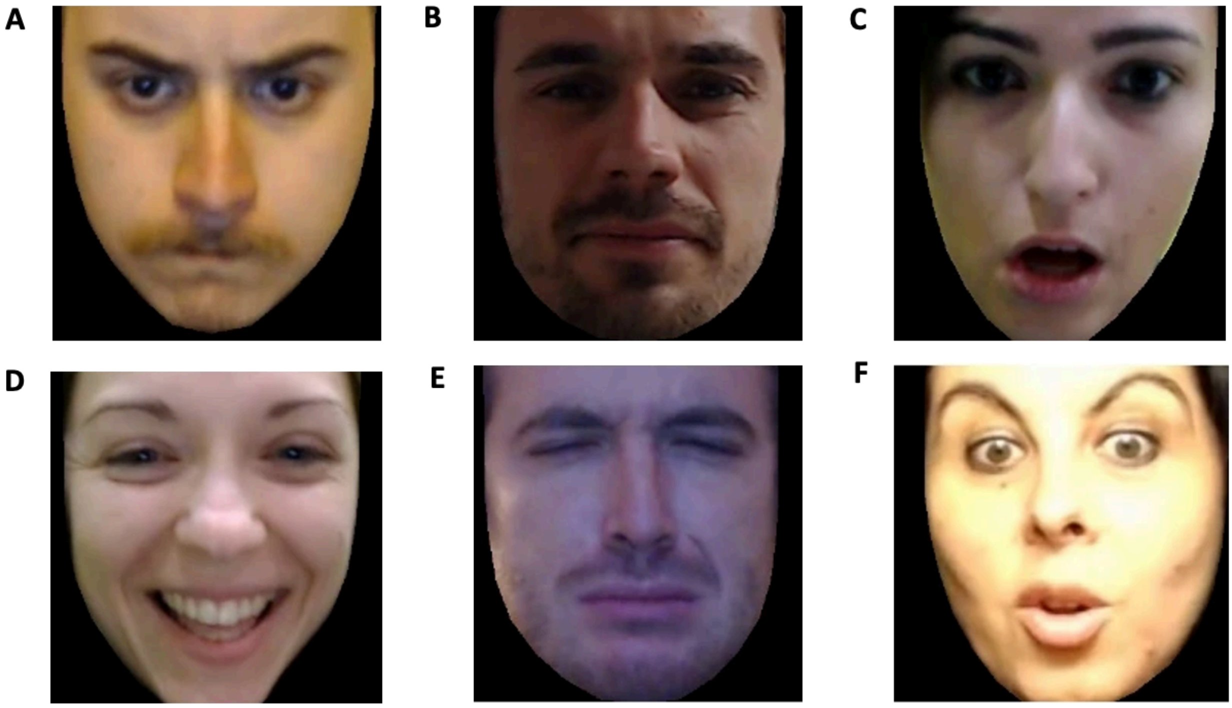


Figure 4.1. Example faces of basic emotions. A) anger, B) disgust, C) fear, D) happiness, E) sadness, F) surprise (examples from PEDFE database; Miolla et al., 2022)

Table 4.1. Characteristics of videos from 21 spontaneous databases

Database	**Videos**		**Encoders**		
	Emotion (n)	**Videos(n)**	**Female**	**Male**	**Total**
4DME	3	29	15	14	29
BAUM-1	6	58	29	29	58
BINED	6	59	29	30	59
BioVid	5	50	25	25	50
BP4D	5	50	25	25	50
CAS(ME)2	4	40	20	20	40
CAS(ME)3	6	60	32	28	60
CASME	5	40	15	25	40
CASME-II	5	40	22	18	40
DISFA	5	50	25	25	50
DynEmo	5	50	25	25	50
EB+	5	50	25	25	50
Emognition	6	60	30	30	60
FEEDTUM	6	60	30	30	60
ISED	4	37	19	18	37
LIRIS	5	50	28	22	50
MMEW	6	57	26	31	57
NVIE	6	59	30	29	59
PEDFE	6	59	30	29	59
SAMM	6	52	34	18	52
SMIC	5	50	24	26	50
Total	-	1060	537	522	1060

Note. Some databases consist of a limited number of emotions

4.2.2 Machine analysis

Due to the discontinuation of FACET software used in chapter 2, the selected 1,060 video stimuli underwent automated analysis using AFFDEX (v1.0; McDuff et al., 2016). AFFDEX, a commercial software developed by the company iMotion, employs advanced computer vision and machine learning algorithms to decode FEs, providing an objective measure of facial activity. The software operates by identifying and tracking specific facial landmarks within the video, which form the basis for the detection and classification of emotions and action units (Bishay et al., 2022).

AFFDEX is capable of recognising the six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) as delineated by Ekman (1992), thereby offering a comprehensive understanding of the emotional content conveyed by the face. In addition to recognizing basic emotions, AFFDEX quantifies the probability of 20 Action Units (AU1, 2, 4, 5, 6, 7, 9, 10, 12, 12L/R, 14, 15, 17, 18, 20, 24, 25, 26, 28, 43) presence, capturing subtle changes in FEs that may not be readily apparent to untrained human observers.

The software generates frame-by-frame probability scores, estimating the likelihood that a human observer would identify each frame as containing a specific emotion or AU. Although AFFDEX was primarily trained with posed datasets, it has demonstrated successful recognition rates for both posed and spontaneous expressions (Stöckli et al., 2016), underscoring its utility in diverse research contexts. Notably, in cross-classifier research, AFFDEX demonstrated robust performance for different types of expressions, as represented by its ROC and AUC curves, which reflect the software's sensitivity and specificity across different classifiers (Dupré et al., 2020). Given that AFFDEX is designed to output recognition scores for basic emotions, our analysis was confined to the six basic emotions as predefined by the database authors.

Building on the foundational work detailed in Chapter 2, this section revisits the application of prototypicality, ambiguity, and complexity within the context of video analysis, emphasising their roles in assessing spontaneous FEs. For a comprehensive description and measure of these parameters, readers are directed to Chapter 2. Our focus here is on evaluating how each parameter contributes to the accurate recognition of emotions in spontaneous expression databases, aligning with our overarching aim of providing cross-corpus evaluation. These parameters are integral for refining the systematic evaluation of the AFFDEX recognition performance, particularly discerning which parameters most significantly predict recognition accuracy.

4.3 Results

4.3.1 Emotion classification

In general, recognition accuracy was significantly higher than chance level of 16.7% (1/6), with an average correct classification of 30% (SD = 23%), $t(20) = 5.57$, $p < .001$, Cohen's $d = 1.22$. Accuracy rates of the majority of databases predominantly fell between 30% and 40%. It is noteworthy that some databases (4DME, CASME, CAS(ME)3, MMEW, SAMM and SMIC) lagged behind with recognition rates under 30%, spanning a range from a mere 6% to 23%.

For a more detailed cross-corpus comparison, a binary emotion recognition outcome by AFFDEX was predicted using a Bayesian logistic regression model, with the type of database (21 databases) and the type of emotion (6 basic emotions) as predictors. The model was fit using the brms package in R, with four chains of 4000 iterations each. The Rhat parameter for each predictor in the model was equal to 1, indicating successful model convergence.

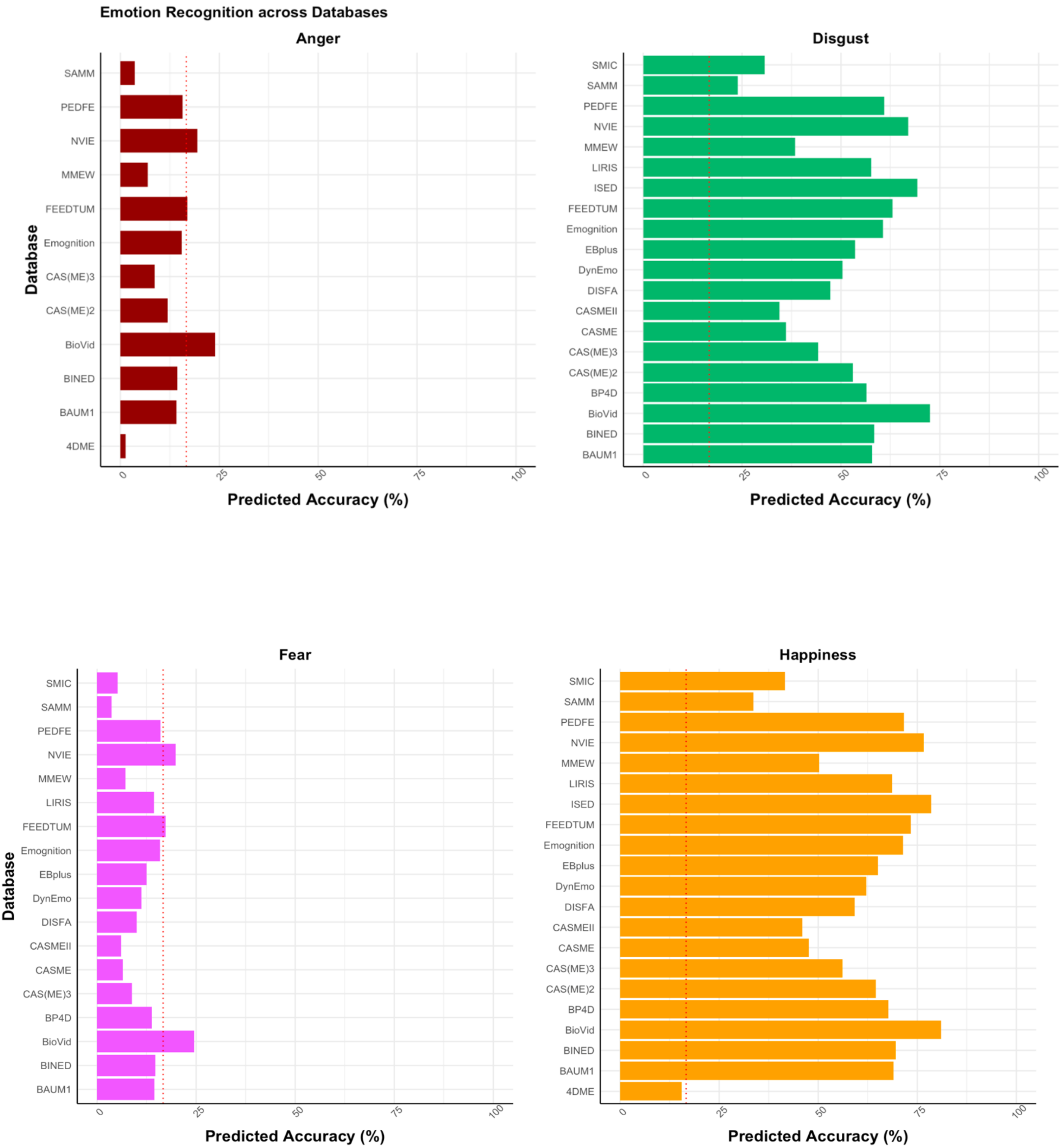
The model provided compelling evidence that the recognition accuracy profoundly varies across databases. Utilising the 4DME database as the reference level, the results revealed that the BioVid was the best-performing database, with a notably high odds ratio (exp(beta) = 28.88, exp(95%CI)[7.03, 135.79]). This superior performance was closely followed by ISED (exp(beta) = 24.76, exp(95%CI)[5.79, 124.09]), NVIE (exp(beta) = 21.81, exp(95%CI)[5.74, 102.34]), and FEEDTUM (exp(beta) = 18.38, exp(95%CI)[4.76, 86.57]), suggesting a higher efficacy for AFFDEX recognition. In support of this notion, predicted accuracy was generally higher for those databases reaching above 35%.

In stark contrast, databases such as SAMM (exp(beta) = 3.17, exp(95%CI)[0.68, 15.64]), SMIC (exp(beta) = 4.55, exp(95%CI)[1.06, 22.48]), CASME2 (exp(beta) = 5.46, exp(95%CI)[1.34, 26.38]), CASME (exp(beta) = 5.86, exp(95%CI)[1.37, 28.36]), and MMEW (exp(beta) = 6.54, exp(95%CI)[1.56, 31.41]) demonstrated poor recognition performance, with their predicted recognition rates below 20%. This disparity highlights the variability in classification and emphasises the necessity for a critical evaluation of database selection in spontaneous facial expression recognition. The remaining databases occupied a middle ground, with the odds ratios ranging from 8.36 (CAS(ME)3) to 16.73 (PEDFE). Overall, the results reaffirm the difficulty in emotion classification of spontaneous FEDBs, with a moderate variability in performance.

Further, results also indicated that recognition accuracy differed as a function of emotion. Compared to the reference level anger, happiness (exp(beta) = 14.80, exp(95%CI)[7.84, 28.94]) emerged as the most accurately recognised emotion, followed by disgust (exp(beta) = 9.01, exp(95%CI)[4.83, 17.38]). The findings align with the inherent expressiveness and distinctiveness of these emotions. However, emotions such as sadness (exp(beta) = 0.62, exp(95%CI)[0.28, 1.36]), surprise (exp(beta) = 1.87, exp(95%CI)[0.95, 3.77]) and fear (exp(beta) = 1.04, exp(95%CI)[0.51, 2.19]), showed comparatively lower accuracy, indicating potential challenges in AFFDEX for recognising these emotions (Figure 4.2).

Intriguingly, the analysis did not indicate a noticeable interaction effect between the choice of database and the type of emotion, suggesting that the influence of emotion on recognition accuracy is relatively uniform across different databases. This absence of interaction points to the intrinsic characteristics of spontaneous databases while suggesting potential challenges of AFFDEX in recognising certain emotions.

As shown in Figure 4.3, the confusion matrix reveals a distinct pattern of misclassification by the AFFDEX, with a notable propensity towards classifying various emotions as “Disgust”. Specifically, when the true emotion was anger, fear, sadness, or surprise, it was misinterpreted as disgust 46%, 38%, 54% and 42% of the time, respectively. In contrast, “Happiness” was correctly classified over 60% of the time, standing as a significant exception to this trend. These patterns underscore the challenges faced by automated classifiers like AFFDEX in accurately distinguishing between certain emotions, particularly in differentiating disgust from anger, fear, sadness and surprise.


Emotion Recognition across Databases
Anger
SAMM
PEDFE
NVIE
MMEW
FEEDTUM
Emognition
CAS(ME)3
CAS(ME)2
BioVid
BINED
BAUM1
4DME
Database
0
25
50
75
100
Predicted Accuracy (%)
Disgust
SMIC
SAMM
PEDFE
NVIE
MMEW
LIRIS
ISED
FEEDTUM
Emognition
EBplus
DynEmo
DISFA
CASMEII
CASME
CAS(ME)3
CAS(ME)2
BP4D
BioVid
BINED
BAUM1
Predicted Accuracy (%)
Fear
SMIC
SAMM
PEDFE
NVIE
MMEW
LIRIS
FEEDTUM
Emognition
EBplus
DynEmo
DISFA
CASMEII
CASME
CAS(ME)3
BP4D
BioVid
BINED
BAUM1
Database
Predicted Accuracy (%)
Happiness
SMIC
SAMM
PEDFE
NVIE
MMEW
LIRIS
ISED
FEEDTUM
Emognition
EBplus
DynEmo
DISFA
CASMEII
CASME
CAS(ME)3
CAS(ME)2
BP4D
BioVid
BINED
BAUM1
4DME
Predicted Accuracy (%)

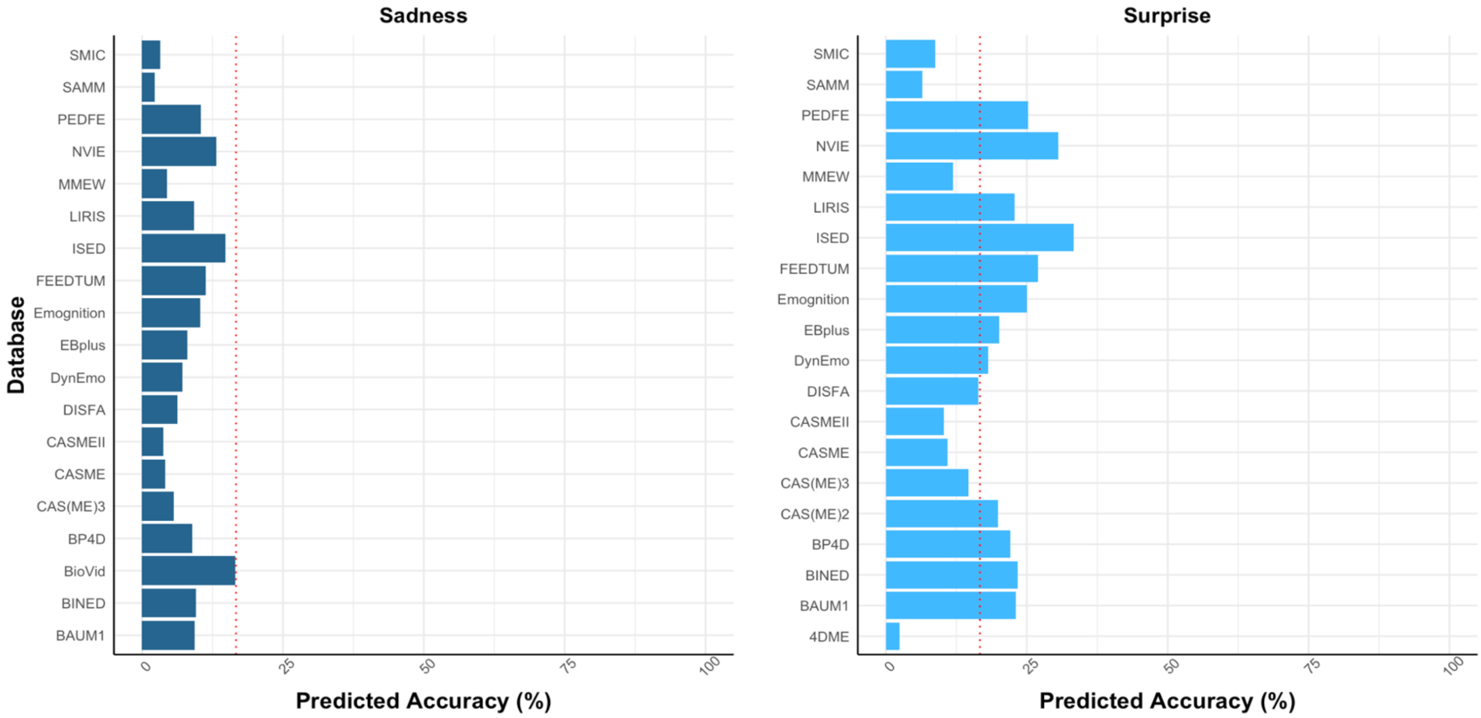


Figure 4.2. Mean recognition accuracy of databases for six basic emotions

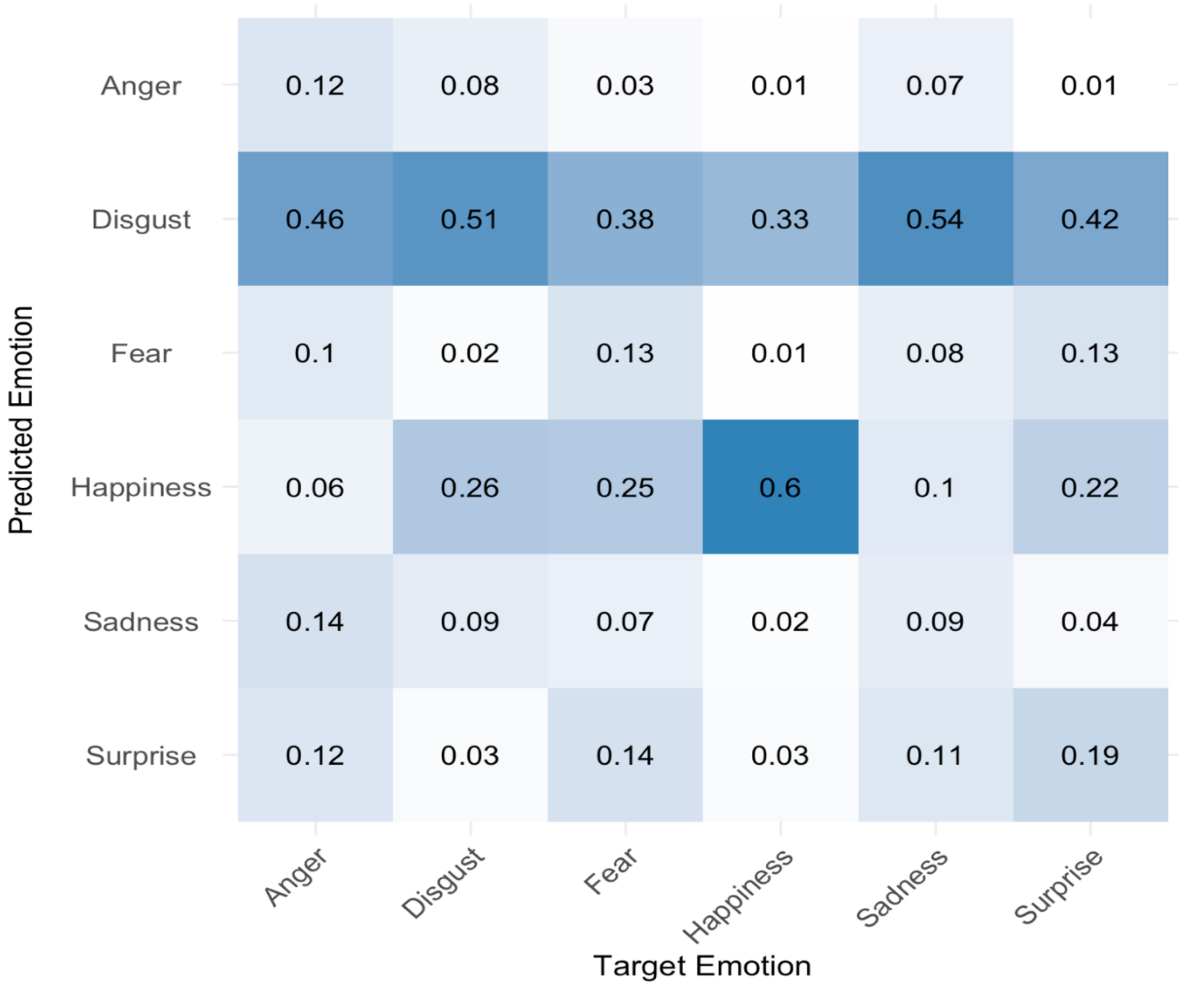


Figure 4.3. Confusion matrix across six basic emotions

4.3.2 Cluster analysis

To discern patterns of emotion recognition performance across various databases, a hierarchical cluster analysis was conducted. This methodological approach aimed to group databases based on their similarities in recognition. The Euclidean distances calculated from the emotion classification accuracies were used to determine the proximity between databases, with shorter distances indicating greater similarity in emotion recognition performance.

The analysis segmented the databases into three distinct clusters based on their recognition accuracy (Figure 4.4). Cluster 1, characterised by superior accuracy, encompassed BioVid, FEEDTUM, ISED, and NVIE, distinguishing them as top-tier databases. Within this cluster, most emotions (all or with mere one exception) consistently surpassed chance-level accuracy, with disgust and happiness reaching rates beyond 60% and 75%, respectively. Cluster 2, more expansive in its composition, comprised databases such as CAS(ME)^2, CAS(ME)^3, DynEmo, and others. Here, happiness and disgust maintained commendable accuracy around 70% and 50%, but a minimum of two emotions fell below the chance threshold. Finally, Cluster 3, featuring 4DME, MMEW CASME, CASME2 SAMM and SMIC, often matched (or even below) against chance level. Although happiness and disgust achieved rates around 40%, all other emotions fell short of the chance threshold. Across all clusters, the standout accuracy of disgust and happiness mirrored patterns observed in the confusion matrix.

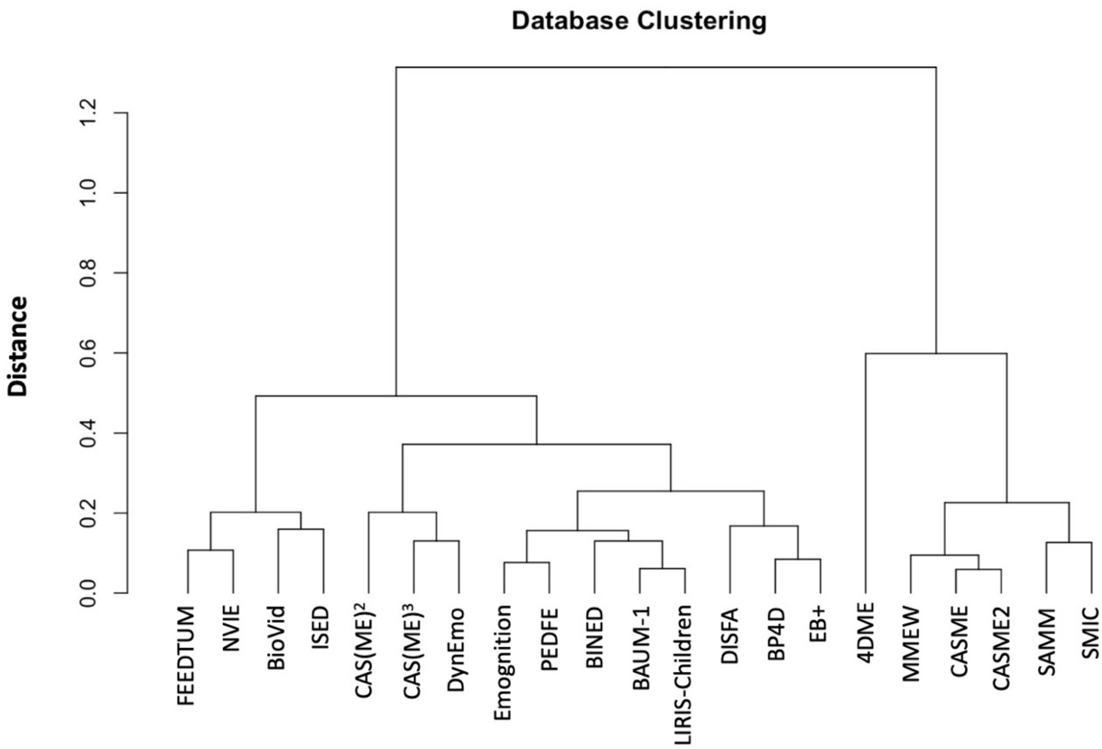


Figure 4.4. Dendrogram showing the hierarchical clustering of the 21 spontaneous databases

4.3.3 AU analysis

A FACS analysis was conducted by AFFDEX to investigate the degree to which the classification of the six basic emotions relies on individual facial actions. We estimated the relative contribution of the 20 AUs to emotion identification using a Bayesian penalised regression model with a regularised horseshoe prior (Piironen & Vehtari, 2017; Van Erp et al., 2019). The predicted number of non-zero coefficients was set to 1-5, mirroring the minimal number of AU prototypes for each emotion. Overall, happiness was the emotion best predicted by AUs (R^2 = 0.65), followed by sadness (R^2 = 0.48), fear (R^2 = 0.46), surprise (R^2 = 0.38), disgust (R^2 = 0.36) and anger (R^2 = 0.13).

When examining the classification by individual facial actions, results showed a varied contribution of individual AUs to emotion classification, as tabulated in 4.2. The findings corroborate, to an extent, the Basic Emotion Theory (Ekman et al., 2002), particularly noting that certain AUs crucially enhance the predictive accuracy of specific emotions. For instance, AU4 effectively predicted correct anger recognition while AUs 9 and 10 are more closely associated with disgust. Similarly, fear is predominantly represented by AU5 and AU20, happiness by AU6 and AU12, sadness by AU4 and AU15, and surprise by AU2 and AU26, respectively.

Notwithstanding these alignments, our findings revealed more flexible AU patterns than previously recognised within the scope of BET. Several prototypical AUs indicated by the FACS were observed to have a diminished or occasionally inverse effect on emotion recognition, such as AU5 and 10 in the context of anger. In addition, our data suggest the contribution of non-prototypical AUs (e.g., AU14 and 28 for anger, AU24 for disgust etc.) in spontaneous expression, albeit with minimal impact (see Table 4.2).

Table 4.2. AU relative contribution to emotion recognition performance

		Emotion exp(b)					
Database		**Anger**	**Disgust**	**Fear**	**Happiness**	**Sadness**	**Surprise**
AU1	Inner brow raise	0.26	-	**0.98**	0.85	**2.09**	**1.05**
AU2	Brow raise	0.99	0.53	**0.84**	1.02	0.50	**2.82**
AU4	Brow furrow	**1.83**	0.60	**0.87**	0.79	**5.37**	0.04
AU5	Eye widen	**0.43**	0.70	**3.41**	1.01	0.91	**0.78**
AU6	Cheek raise	0.80	0.59	0.88	**1.09**	**0.65**	1.26
AU7	Lid tighten	**1.30**	1.48	0.06	0.82	0.80	0.17
AU9	Nose wrinkle	0.74	**1.71**	0.61	0.96	0.91	0.56
AU10	Upper lip rise	**0.29**	**2.17**	0.85	0.89	0.41	0.15
AU12	Lip corner puller	0.06	0.31	0.03	**18.14**	0.76	0.21
AU12L/R	Smirk	0.95	1.04	0.90	-	0.95	1.02
AU14	Dimpler	1.07	1.04	-	0.99	1.03	0.97
AU15	Lip corner depressor	0.99	**0.92**	0.82	1.11	**1.24**	1.01
AU17	Chin raise	**-**	**1.14**	1.15	0.94	**0.98**	0.81
AU18	Lip pucker	0.98	1.40	1.18	0.74	0.69	0.77
AU20	Lip stretch	1.02	0.64	**1.16**	1.90	-	0.35
AU24	Lip press	**0.95**	1.17	0.93	0.99	0.99	-
AU25	Mouth open	0.99	**0.95**	**1.01**	1.33	1.18	1.26
AU26	Jaw drop	**1.01**	**1.04**	**0.59**	-	0.83	**1.32**
AU28	Lip suck	1.14	0.97	1.03	1.09	0.77	1.07
AU43	Eye closure	0.95	1.19	1.19	1.01	1.22	-

Note. Exponentiated regression coefficients (exp(beta)) of prototypes are printed in bold. The prior of p_0 were anger = 4, disgust = 1, fear = 5, happiness = 1, sadness = 2, surprise = 3.

4.3.4 Prototypicality, ambiguity and complexity

We subsequently examined the difference in parameter scores across databases and emotions, uncovering a notable trend in their distribution. Generally, these scores showed marginal variation, with most falling within the range of 20 and 30. On average, databases demonstrated low prototypicality, marked by a mean score of 24.66 (SD = 17.15). Notably, certain databases including, BAUM-1, BP4D, and ISED, stood out with higher prototypicality scores exceeding 35, aligning with those achieving moderate to high accuracy in emotion classification. In contrast, databases focusing on micro-expressions (particularly CAS(ME)2, CAS(ME)3, CASME-II, SMIC) exhibited notably lower prototypicality, with scores under 20. Together, the trend indicates that spontaneous expressions often deviate from the FACS prototypes, highlighting the diversity in their appearance.

In the assessment of ambiguity across databases, a consistent yet nuanced pattern emerged. The ambiguity scores predominantly centred around a mean of 28.79 (SD = 6.65), denoting a moderate level of ambiguity in the representation of emotions. This trend stands in contrast to the observed patterns in prototypicality. Specifically, micro-expression databases such as CAS(ME)2, CAS(ME)3, MMEW, SAMM and SMIC, exhibited slightly higher levels of ambiguity, with scores frequently surpassing 30. On the other hand, databases like BAUM1, BP4D and ISED showed lower ambiguity levels, generally around 25. Yet, generally moderate level of ambiguity across databases highlights the more complicated discernibility of spontaneous databases.

Finally, for the complexity, a significant variance across databases is observed. The overall mean complexity score stands at 18.01 (SD = 7.06), suggesting a relatively low to moderate complexity across databases. However, databases like BP4D and EB+ exhibit notably higher complexity scores around 38, indicating a wider range of variation across databases. Again, micro-expression databases such as CAS(ME)2, CAS(ME)3, CASME-II, SAMM and MMEW report much lower complexity scores, often below 10 (see Figure 4.5).

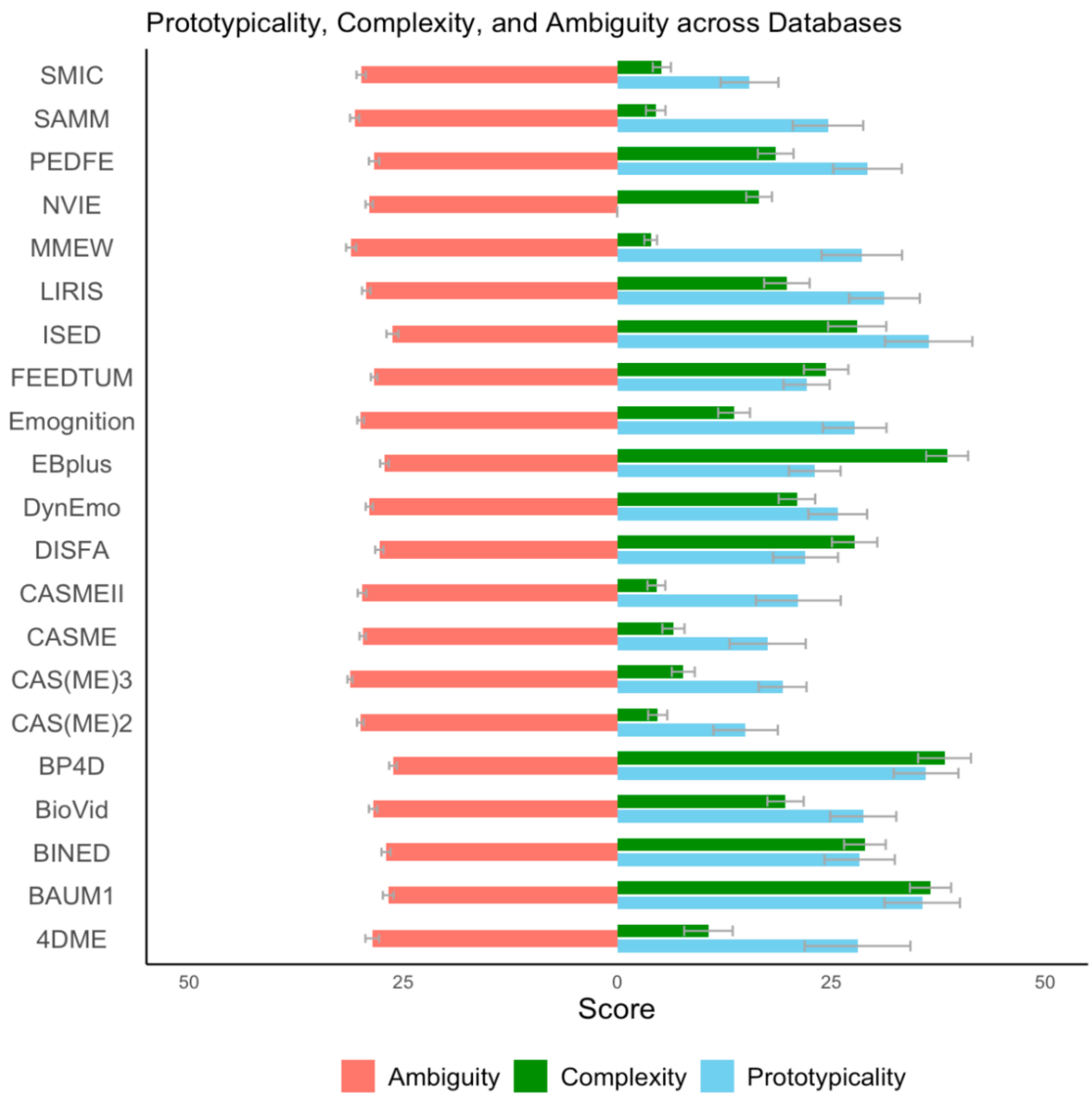
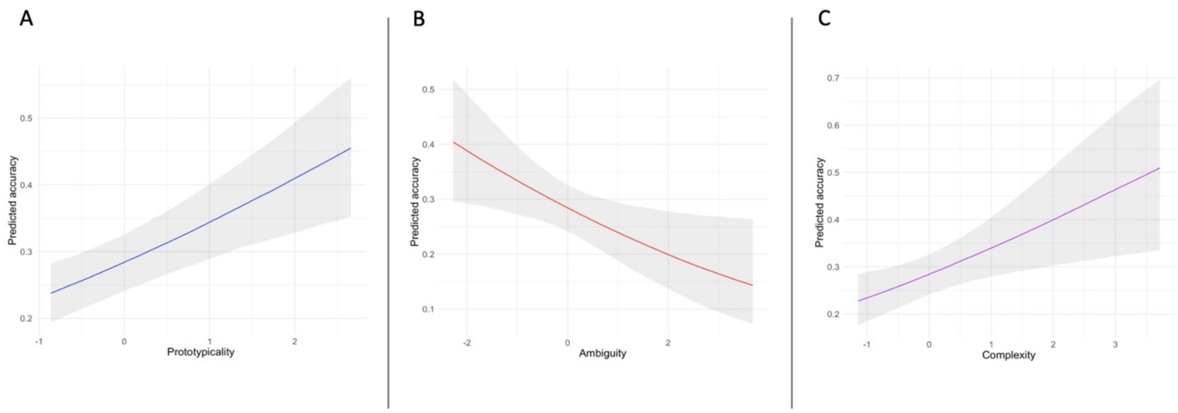


Figure 4.5. Prototypicality, ambiguity, and complexity across databases

In order to discern the relative contribution of each parameter**,** a Bayesian mixed-effects logistic regression model was fitted to predict classification accuracy based on fixed effects of prototypicality, ambiguity, and normalised complexity, with random intercepts for each database. The standard deviation of the random intercepts for each database was 0.33, indicating unexplained variability in accuracy across databases. The model diagnostics indicated successful convergence (Rhat = 1 for all parameters), underscoring the robustness of the model.

The model revealed that both prototypicality and complexity exerted a positive influence on accuracy. Specifically, a unit increase in prototypicality was associated with an increase in the odds of accuracy by a factor of 1.32 (exp(beta) = 1.32, exp(95%CI)[1.14, 1.53]). In a similar vein, a unit augmentation in complexity was associated with an increase in the odds of accuracy by a factor of 1.30 (exp(beta) = 1.30, exp(95%CI)[1.07, 1.58]). Conversely, ambiguity was found to negatively impact accuracy. A unit increase in ambiguity was associated with a decrease in the odds of accuracy by a factor of 0.79 (exp(beta) = 0.79, exp(95%CI)[0.65, 0.96]). Given the results, prototypicality appears to be the strongest positive predictor, but the difference between prototypicality and complexity is minimal. Ambiguity is a strong negative predictor, with its effect being in the opposite direction (Figure 4.6).

Figure 4.6. Predicted power of prototypicality, ambiguity, and complexity on recognition accuracy



Note. Regression line indicates the relationship between predicted accuracy and individual scores of (A) prototypicality, (B) ambiguity, and (C) complexity. The line shades represent upper and lower bounds 95% credible interval

4.4 General Discussion

The growing interest in ecologically valid facial expression stimuli has spurred the development of a multitude of spontaneous FEDBs over the past two decades. This proliferation has provided the scientific community with a diverse array of datasets, each distinct in its size and characteristics. While there have been isolated efforts to validate these resources (Wallhoff et al., 2006; Yan et al., 2014; Zhang et al., 2014), a comprehensive cross-corpus evaluation has still been lacking. Previous validation efforts on databases often focused on a narrow selection of spontaneous databases or primarily against posed ones (Krumhuber et al., 2021, encompassing both static and dynamic datasets (Cassidy et al., 2015). A comprehensive evaluation, particularly regarding expression characteristic metrics (prototypicality, ambiguity, complexity) is essential for meaningful database comparison, which has not been overtly explored. Moreover, given the variability in validation methodologies (Jia et al., 2021), a standardised approach becomes imperative. This study aims to evaluate various spontaneous FEDBs using the commercially available AFEA software, AFFDEX (iMotion). In doing so, we not only offer critical evaluation metrics but also shed light on the available sets in the field, striving to establish a foundational benchmark for further studies in spontaneous FE analysis.

The results reaffirm previous findings highlighting the inherent difficulties in recognising spontaneous FEs (Krumhuber et al., 2019; for review see Webster et al., 2021). Through an extensive evaluation of 21 databases, we observed a modest average classification accuracy of 30%. Nevertheless, while earlier conclusions about the challenges of spontaneous expression recognition often pivoted on limited spontaneous datasets, our expansive approach unveiled marked disparities in emotion classification accuracy across databases. Specifically, databases such as FEEDTUM, NVIE, BioVid and ISED emerged as the top-performing databases, achieving recognition rates surpassing 40% - a degree notably above the chance level. These databases predominantly feature expressions induced by video stimuli, specifically designed to trigger target emotions (Happy et al., 2015; Wallhoff et al., 2006). The successive performance of these databases may be partially explained by the high data quality and resolution they offer, suggesting a judicious balance between experimental control and distinctive representation of emotion, ultimately leading to improved algorithmic classification.

In contrast, databases such as 4DME, CASME, CASME 2, MMEW, SAMM, and SMIC, which primarily catalogue micro-expressions, demonstrated subsequently diminished recognition accuracy, often nearing or even falling below the chance level. Micro-expressions are fleeting, involuntary facial movements that manifest when individuals disclose an emotion they intend to conceal or regulate (Yan et al., 2014). These rapid and subdued expressions, typically lasting a mere 0.5 seconds (Davison et al., 2018), invariably complicate recognition, especially when compared to intense ones (Yitzhak et al., 2017). Despite their subtlety, micro-expressions offer insightful glimpses into genuine emotional states, warranting further exploration (Yan et al., 2013). Given these intrinsic challenges, the observed average accuracy of 16.9% in these databases may be unsurprising. However, when paired with specialised machine-learning techniques, these databases have showcased commendable recognition rates in their original publications (Li et al., 2022; Yan et al., 2013), highlighting their niche utility in the field of facial expression research.

The results collectively shed light on the intricate diverse representativeness inherent in spontaneous FEDBs. This disparity accentuates the notion that spontaneous FEDBs are not a monolithic category; instead, they span a spectrum, from easily recognisable databases to fleeting subtle micro-expressions. Such diversity in representation necessitates the pressing need for a granular comprehension of the distinct properties within spontaneous databases, essential for not only crafting more effective AFEA software but also advancing our understanding of spontaneous expression recognition.

Building on the systematic review in Chapter 3, the findings from Chapter 4 provide preliminary insights into the aspects of spontaneous FEDBs that may facilitate more accurate emotion recognition by AFEA systems. While our study did not directly examine how specific database characteristics influence machine performance, certain patterns emerge when exploring the top-performing databases. These databases often feature highly prototypical expressions captured from a frontal view with minimal head orientation, and they frequently relying on a single elicitation technique targeted for inducing basic emotions. These features likely contribute to their standardisation, which seem to align well with the technical capabilities of automated recognition tools, optimised for clear, unambiguous input.

By exploring the commonality within clusters of databases and the differences between clusters, the current research acknowledges the heterogeneity of spontaneous databases in terms of their constructive and representative aspects, thereby opening new avenues for both empirical and practical applications. The findings collectively aid in the informed-selection of databases that align with research objectives and suggest areas where further refinement of AFEA systems can take place, potentially leading to more robust and reliable models capable of handling the diverse nature of spontaneous FEs.

Findings also revealed marked variation in recognition accuracy across emotions within spontaneous databases. We focused on the basic six emotions for stimulus collection, aligning with the analytical capabilities of commercial AFEA classifiers. Rooted in Basic Emotion Theory (BET), these emotion categories stringently adhere to Ekman's criteria for universal expressions, representing core emotional repertoire (Ekman & Cordaro, 2011). Consistent with previous research (Krumhuber et al., 2019; Nummenmaa & Calvo, 2015), happiness, the sole positive emotion examined, emerged as the most accurately classified emotion, likely due to the distinctive facial prototypes such as raised cheeks and crow's feet, enhanced recognition (Ekman et al., 1990; Soussignan, 2002). Conversely, sadness was frequently misclassified as disgust, resulting in reduced accuracy. Notably, while both happiness and disgust achieved high recognition rates, the recurrent miscategorisation of emotions (anger, fear, sadness, surprise) as disgust yielded reduced recognition accuracies, with only surprise slightly exceeding the chance level. This trend raises questions about potential biases in the AFFDEX software, prompting concerns about its precision in classifying disgust. This challenge was more evident in micro-expression databases, suggesting their understated displays were erroneously classified, especially in the absence of a 'neutral' category in AFFDX classification. Overall, our findings spotlight the difficulty and potential pitfalls of relying merely on a single commercial software for spontaneous FEDBs, advocating for various validation methods in facial expression databases.

It is important to consider whether the misclassifications observed with AFFDEX are also present in human observers or other AFEA classifiers like FACET. Human observers often exhibit systematic confusion patterns, such as mistaking disgust for anger or fear for surprise, which remain relatively consistent across studies and databases (Calvo & Numenmaa, 2016). This consistent pattern is likely due to the shared facial muscle activations between these emotions (Ekman & Friesen, 1978). Given that machines are typically trained on human-annotated data, some classifiers tend to mirror these confusion patterns, as seen with FACET and FaceReader (Calvo et al., 2018; Lewinski et al., 2014). Notably, FACET tends to misclassify certain emotion as disgust to a certain extent; however, its misclassification errors are more evenly spread across different emotions (e.g., fear to happiness; Krumhuber et al., 2020). This variation across classifiers might be attributed to different dataset used in the training. Although our study does not directly compare the sensitivity and confusions patterns across various classifiers, understanding these differences is vital for ensuring the reliability of AFEA tools. Future research could benefit from examining how these biases manifest across different algorithmic tools and human ratings, potentially leading to more refined models that reduce misclassification errors.

Through a detailed FACS analysis, we examined the extent to which the classification of six emotions depends on individual facial actions. While FACS prototypes were shown to be the most salient AUs for accurate emotion classification, we identified other AUs indicative of emotions (e.g., AU7 and AU18 for disgust). Also, while FACS traditionally underscores the significance of AU combinations in characterising emotions (Ekman et al., 2002), our findings showed that frequently just a select group of these prototypes, or even an isolated AU, plays a pivotal role in the accurate classification of spontaneous expressions. This result resonates with earlier studies suggesting that a singular AU can be sufficiently representative of distinct emotions (Namba et al., 2017). This intricate involvement of AUs raises pivotal questions regarding the contribution of prototypical AUs in the context of spontaneous expression recognition. While our results affirm the significance of certain AUs as posited by Ekman's prototypes, they also highlight that not all prototypical AUs are equally influential. For example, in the case of anger, AU4 and 12L/R emerge as more critical than other AUs traditionally associated with this emotion. This disparity suggests a potential re-evaluation of what constitutes a 'full' versus a 'partial' prototype in the representation of real-life emotion. However, it is important to note that the most potent predictors for precise recognition invariably aligned with the AUs delineated in the FACS manual, emphasising their foundational role in emotion classification.

In support of this notion, prototypicality played a pivotal role, wherein expressions aligned closely with BET predictions facilitating enhanced recognition (Ekman et al., 2002). Notably, while prototypicality emerged as a significant predictor for static expressions, but not for dynamic expressions in earlier research using posed expressions (Kim et al., 2023), its influence becomes more pronounced in spontaneous facial displays. Previous investigations on prototypicality primarily centred on posed expressions, often amplifying their importance (Matsumoto & Hwang, 2014). Here, we showed the varying degrees of prototypicality within spontaneous FEDBs influencing the discernibility of key facial configurations, thereby impacting machine analysis. This dependency is likely rooted in the machine's historical training on posed or stylised expressions (Pantic & Bartlett, 2007).

The role of prototypical facial cues might be modulated by the expression intensity. Consistent with prior studies (Kim et al., 2023; Matsumoto & Hwang, 2014), the complexity of expressions markedly influenced machine performance. Elevated facial expressivity seems to bolster emotion decoding, suggesting a fundamental association between intensity and recognition (Hess et al., 1997; Wingenbach et al., 2016). Importantly, the complexity enumerates the number of active AUs, independent of their individual strength. This metric captures an abundance of facial actions, unfettered by the magnitude of singular AU contractions. This distinction is crucial as facial displays often manifest as a confluence of multiple AUs, not always aligning with prototypical representation of basic emotions.

The heterogeneity in prototypicality and complexity across FEDBs may be closely intertwined with the elicitation tasks and stimuli deployed for emotion induction. For instance, while most databases deployed video induction techniques without any restriction, micro-expressions databases, which frequently instruct participants to suppress their emotions, tend to showcase facial configurations in diminished intensity (Davison, 2018; Yan et al., 2013, 2014). Such attenuated intensity can engender increased confusion and uncertainty in emotion classification (Ichikawa et al., 2014; Matsumoto et al., 2002), presenting a formidable challenge for machine-based recognition.

Finally, ambiguous expressions displaying multiple basic emotions (i.e., when FEs are categorically ambiguous) consistently presented classification errors, particularly when measured by the degree to which contradictory emotional cues were present. Given that classification decisions typically rely on the most distinctive facial features (Calvo et al., 2012; Du et al., 2014; Fiorentini & Viviani, 2009), such ambiguity escalates the challenges faced by machine-driven systems in discerning discrete emotions (Calder et al., 2000; Neta & Whalen, 2010). In naturalistic settings, emotions rarely manifest in isolation. They often co-occur, creating various nuances of emotional states (Du et al., 2014; Du & Martinez, 2022). This intricate emotional tapestry not only highlights the diversity of human emotional experiences but also emphasises the complexity of recognising spontaneous expressions (Ito et al., 2017; Kinchella & Guo, 2021).

These findings collectively highlight the interdependent relationship between prototypicality, ambiguity and complexity within spontaneous emotional expressions. Expressions characterised by higher ambiguity often contain conflicting emotional cues, leading to less distinct prototypical FEs. Conversely, clearer expressions tend to have lower ambiguity and higher prototypicality. The generally low parameter scores across spontaneous databases likely stem from the low probability in emotion and AU ratings across databases used for measuring complexity scores. These findings underscore the intricate balance and the critical role these factors play in the accurate classification of FEs.

Until now, there is a prevailing trend where expressions are strictly labelled based on single emotions. While most of the existing databases contain single emotion labels that are indicative of targeted emotions, many videos present multiple emotional reactions, highlighting potential issues with annotation quality (e.g., smile after surprise elicitation). Such lapses in annotation potentially lead to erroneous conclusions in data interpretation and inadvertently introduce biases in subsequent analysis (Chen et al., 2021; Zeng et al., 2014). It is advisable to involve multiple annotators, ensuring their inputs are cross-validated to enhance annotation precision and move beyond mere single categorical labels for a more accurate representation of emotion.

A noteworthy limitation of the current study is its primary focus on induced spontaneous expressions, which might not fully capture the various types of expressions encountered in real-world scenarios. This emphasis likely stems from the prevalent use of emotion induction techniques in spontaneous databases. Autobiographical and other context-driven facial displays, which often provide a richer and more nuanced representation of emotions (Levenson, 2007), could be explored in future studies. Moreover, the study did not account for the impact of technical features, such as head orientation, duration, face box, resolution, and frame rate, all of which have been identified as pivotal determinants of machine recognition accuracy (Krumhuber et al., 2021). The lack of human observer ratings further constrains the applicability of our findings, primarily to machine-based recognition. Incorporating such human evaluation would establish a comparative benchmark against machine classifications, shedding light on the similarities and differences between machine and human recognition of spontaneous expressions.

To optimise our understanding of spontaneous expression, future research necessitates a paradigmatic shift from the prevailing elicitation techniques. While databases predominantly rely on video induction, the resulting expressions can occasionally produce a fixed appearance. For instance, the uniformity of gaze direction (e.g., eye fixation on the screen) might subtly influence the spontaneity of the expression (Ganel et al., 2005; Soussignan et al., 2018). Given that the contexts in which expressions manifest are unlikely to be steady in everyday life, databases should aim to encompass a wide array of emotional contexts that authentically mirror real-life expressions.

Notably, while a significant portion of examined databases presented a consistent frontal view, it might also be useful for FEDBs to encapsulate more flexibility in head orientation, ensuring that AFEA classifiers are equipped with training data that spans diverse facial viewing angles. Achieving this could involve leveraging multi-camera setups, offering a comprehensive view of facial movements (Tcherkassof et al., 2013; Zhalehpour et al., 2017; Zhang et al., 2016). These advanced setups are particularly beneficial for AFEA algorithms, which necessitate voluminous and varied training data, thereby enabling a fine-tuned analysis of a broad spectrum of spontaneous expressions.

Even though discrete emotions paradigms remain influential (Cordaro et al., 2018), emerging critiques warn against potential oversimplifications, urging a more detailed exploration of emotional diversities (Barrett et al., 2019; Kappas et al., 2013). In response, a few promising efforts have lately aimed to extend the range of elicitation techniques and include non-basic affective states. Some of the databases examined here (i.e., BP4D, DynEmo, Emognition etc.) reflect this shift, capturing a wider array of affective displays such as boredom, enthusiasm and being moved, which might otherwise be subsumed under broader discrete categories. Moreover, pioneering attempts to detect emotions in naturalistic settings using commercial classifiers have begun to capture the various subtle nuances of facial expressions (Bishay et al., 2022). It becomes incumbent upon future research to critically evaluate and validate spontaneous databases that transcend the basic emotion framework, aiming to fully embrace and accurately classify the diverse affective states. Such progress is not only pivotal for advancing emotion theory but also augments the versatility and applicability of FEDBs and AFEA classifiers. The present study constitutes an initial effort to offer cross-corpus validation of 21 spontaneous databases in a dynamic format. We hope that our contributions serve as a benchmark for assisting future progress in the field.

CHAPTER 5

General Discussion

Facial expressions play an integral role in daily social interactions, serving as primary indicators in the interpretation of others' mental states and intentions (Krumhuber et al., 2023). These expressions we encounter in everyday scenarios are inherently spontaneous and dynamic, revealing essential information about the underlying emotional states. Despite their significance, much of the past research has predominantly utilised posed static images, often captured at the peak of emotional displays. Although these controlled depictions have laid the groundwork for understanding basic emotions (Ekman et al., 1987), they do not adequately represent the progression and intricate morphology of genuine emotional expressions (Barrett et al., 2019), and thus are low in ecological validity. Converging evidence suggests that dynamic and spontaneous aspects are essential for an accurate and realistic representation of emotions and social signals (Hess et al., 1990; Namba et al., 2018).

A major concern within facial expression research lies in determining the specific circumstances under which dynamic information yields facilitative benefits and what attributes make expressions recognisable. Past research provided suggestive (Ambadar et al., 2005; Cunningham & Wallraven, 2009), albeit inconclusive, evidence for the dynamic advantage, with several studies showing no benefits of dynamic information (Gold et al., 2013; Kamachi et al., 2001). In Chapter 2, the current work showed that dynamic information plays a compensatory role in emotion recognition when static cues are insufficient for accurate representation in humans and machines (Kim et al., 2023). The recognisability of static and dynamic expressions appears to be significantly influenced by key featural parameters (prototypicality, ambiguity, and complexity). This research provided preliminary insights into how these parameters interact within the dynamic and static expressions. Nonetheless, since the majority of portrayals used in this work were from posed databases, several questions remain regarding the spontaneous aspects of facial expressions.

The persistent reliance on posed expressions in facial recognition research likely stems from the limited resources for and exploration into spontaneous FEDBs. Central to our investigation are the questions concerning which databases are available to study spontaneous expressions, and what unique conceptual and technical information underlies the spontaneous databases. Further inquiries pertain to how well these expressions are recognised in terms of both emotion and AUs. Herein, the systematic review and empirical evaluation of available resources is needed. This dissertation is dedicated to shedding light on the scope and intricacy of spontaneous expressions, aiming to foster a more ecologically valid comprehension of facial behaviour. By conducting a systematic and empirical analysis of spontaneous FEDBs, this research seeks to delineate how dynamic, genuine facial expressions contrast with their static, posed counterparts, focusing on aspects of recognizability and constructive aspects.

As this dissertation investigates the recognition of dynamic and genuine FEs, it becomes essential to assess how well these expressions are recognised by humans and machines. Previous research has demonstrated that machines typically outperform humans in recognising posed, prototypical expressions (Stöckli et al., 2018), yet they struggle for subtle and spontaneous expressions (Yitzhak et al., 2017). Although AFEA tools typically exhibit recognition and confusion patterns similar to human observers (Calvo & Nummenmaa, 2016), the inconsistent performance of these machines raises concern about their effectiveness in handling subtlety and ambiguity prevalent in everyday expressions. Findings from Chapter 2 provide evidence that, to some extent, machines are becoming capable of handling these, non-target, subtle, and ambiguous expressions. However, as highlighted in Chapter 4, the overall low performance of machines in recognising diverse spontaneous FEDBs suggests that their effectiveness may depend on the databases and tools for testing.

The current chapter is to provide a comprehensive overview and synthesis of research findings. Initially, I will summarise the key outcomes of the empirical studies and reviews conducted as part of this dissertation. Following this summary, I propose a conceptual framework that integrates the observed difference between posed, spontaneous and naturalistic expressions, offering a cohesive understanding of facial expressions. I will then delve into the significant contributions and implications of this research, highlighting its impact on the field of facial expression research. This will be followed by an evaluation of the strengths and limitations of our methodological approach, ensuring a balanced reflection on the conducted studies. Finally, the current chapter will outline recommendations for future research directions.

5.1 Summary of the Main Findings

The aim of the research reported in Chapter 2 was to investigate the dynamic advantage in emotion recognition, particularly contrasting static versus dynamic expressions and their effect on both human and machine recognition. Three featural parameters - prototypicality, ambiguity, and complexity – were further to identify the condition under which the movement confers its facilitative effects. This work utilised static images, each depicting peak expressions of targeted and non-targeted emotions, to compare them with corresponding dynamic representations. This methodology was specifically chosen to test conditions under which dynamic information substantially aids emotion recognition, particularly when static representations alone are more or less sufficient in depicting emotion.

The results showed notable distinctions between static and dynamic expression recognition. In alignment with the hypothesis, dynamic cues were found to aid recognition for non-target images but not for target images. Notably, results illuminated how the distinctiveness of expressions impacts recognition, as modulated by our featural parameters. High prototypicality and complexity were conducive to heightened recognition rates, underpinning the pivotal role of clear, expressive cues. Conversely, elevated levels of ambiguity typically hindered accurate emotion recognition. These patterns of recognition were consistently observed across both human observers and the AFEA tool.

Collectively, the results from Chapter 2 suggest the compensatory yet significant role of dynamic information, especially when static cues fail to represent target emotions. These results suggest that peak static faces could serve as a viable alternative for dynamic expressions, at least for emotion recognition. In both studies, the AFEA tool outperformed human observers across all three conditions – target, non-target, and dynamic -. This consistent performance, along with similar or higher recognition rates may suggest that automated tools are an effective alternative for emotion recognition tasks. These findings extend previous research by demonstrating the inconsistent effect of dynamic information on emotion recognition depending on the representativeness of static expressions.

In Chapter 3, I systematically reviewed spontaneous and dynamic facial expression databases, diverging from past reviews that predominantly cover posed, static datasets (Diconne et al., 2022; Guerdelli et al., 2022; Haamer et al., 2018). Despite the increasing awareness of ecological validity in facial expression research, a comprehensive review of available spontaneous resources and their impact on facial expression recognition was lacking. This chapter therefore aimed to bridge this knowledge gap by not only introducing a variety of datasets but also detailing the conceptual, technical, and practical facets of these resources. Our comprehensive assessment aspires to serve as a benchmark, equipping future researchers with a guide to making well-informed decisions, thereby enhancing the utility and versatility of spontaneous FEDBs in facial expression research.

In conceptual aspects, the review highlights the importance of elicitation techniques in developing spontaneous FEDBs. These methods, which span from passive viewing of emotion-evoking videos to active engagement in elicitation tasks, are pivotal in shaping the genuineness and quality of the recorded expressions. Such approaches also streamline the annotation process by labelling recordings directly from targeted emotions. The meticulous annotation of emotion or AU significantly amplifies the practicality and utility of databases in exploring both encoding and decoding aspects of facial expressions. Moreover, our review identified a persistent focus on basic emotions within these databases, likely reflecting the enduring impact of the discrete emotion paradigm. Although there is a noticeable shift towards incorporating a wide range of emotional nuances within some FEDBs, extending beyond basic emotion categories remains an area for continued effort.

Technically, there have been significant advancements in recording protocols within FEDBs, with many databases now providing standardised illumination and background, alongside the transition to higher-resolution recordings. However, the current work also identified the need for improvement in frame rates to capture detailed dynamics of facial behaviour more precisely. Additionally, the prominence of AFEA tools for assessing the emotional quality of recordings has reshaped database design, increasingly tailoring these resources towards improving human-machine interaction. However, the integration of physiological measures and multimodal data into these databases is still evolving, indicating room for further improvement to provide a more holistic representation of emotional expressions.

In synthesising the conceptual and technical aspects of the databases, our review highlights significant disparities in the development and application of spontaneous and dynamic FEDBs. While the increased number of spontaneous databases represents a significant leap toward ecological validity in facial expression research, this work pinpointed several persistent challenges that warrant attention. Specifically, there is a need for diversifying elicitation methods, broadening emotional categories and adopting rigorous annotation and evaluation processes. Addressing these gaps will not only advance the field but also ensure that these databases more accurately reflect real-life human expressions, thereby facilitating more robust and generalisable research outcomes.

Expanding upon the systematic review in Chapter 3, our study in Chapter 4 sought to empirically evaluate spontaneous and dynamic FEDBs through a cross-corpus analysis. Employing the commercial software AFFDEX, our objective was to refine and standardise comparative analysis for spontaneous databases. Past evaluations of these resources have been narrowly focused, often confined to single-corpus studies that fail to assess the generalisability of findings across varied stimulus sets. These evaluations predominantly concentrated on singular aspects – either emotion or AU recognition – without integrating a comprehensive framework. Additionally, this chapter delved into the specific roles of individual AUs in the context of spontaneous expression recognition and assessed their correspondence with the FACS prototypes. The investigation was further explored by revisiting and validating the featural characteristics of prototypicality, ambiguity, and complexity, outlined in Chapter 2, to ascertain their influence on the recognition of spontaneous expressions.

Our findings showed considerable difficulties inherent in recognising spontaneous facial expressions, surpassing mere chance levels. However, expanding previous findings (Dupré et al., 2019; Krumhuber et al., 2021a), our results revealed substantial disparities in classification accuracy across spontaneous FEDBs, highlighting the varied representativeness within spontaneous FEDBs for basic emotions. Interestingly, while certain FACS prototypes proved significant for accurate emotion classification, other AUs also appear to be indicators of specific emotions. Nonetheless, the most reliable predictors for precise recognition consistently aligned with prototypes delineated from the FACS manual. Consistent with findings from Chapter 2, heightened prototypicality and complexity were observed to enhance recognition performance. Conversely, increased ambiguity exacerbated the recognition challenges, complicating the recognition of discrete emotions.

Taken together, the findings from Chapter 4 highlight the diverse representativeness across spontaneous FEDBs, spanning from easily identifiable to subtle and fleeting. This disparity emphasises that spontaneous FEDBs are not homogenous but rather form a spectrum. Synthesising the insights from Chapter 3, our research extends previous understanding by highlighting factors and characteristics influencing spontaneous expression recognition. This heterogeneity across databases necessitates the need to investigate distinct properties within spontaneous databases and their impacts. Furthermore, our findings align with prior research (Namba et al., 2017; Girard et al., 2015), indicating that spontaneous expressions exhibit more flexible AU patterns contributing to accurate classification. Nevertheless, alignment between robust predictor AUs with FACS prototypes reinforces their consistent role in expression recognition.

5.2 Advancing the Ecological Validity in Facial Expression Research

The methodological preference for posed static displays ensures high levels of replicability and control, facilitating the detailed examination of specific facial cues and their cognitive interpretations (Dawel et al., 2021). Partly because posed expressions at their peak are easy to recognise, it has been argued that specific morphological cues identified from these expressions are the most reliable indicators of emotions across cultures (Frank et al., 1993). However, this approach may inadvertently overlook the complex morphological and temporal patterns in facial expressions as they manifest in the natural environment.

Posed expressions and spontaneous expressions are fundamentally different, not just in their outward presentation but also in their underlying neural mechanisms (Ekman et al., 1980;

Morecraft et al., 2001). This divergence raises critical questions about the generalisability and reliability of posed static expression-based findings to the spontaneous or naturalistic dynamic expressions encountered in daily life (Motley & Camden, 1988). Despite an emerging consensus acknowledging more complex properties involving spontaneous expressions (Happy et al., 2015), the comprehensive insights they afford have been markedly underexplored prior to this dissertation.

Given that expressions are spontaneous and dynamic, our research initiated a focused examination of the role of dynamic properties within facial expression recognition. I manipulated the temporal presentation of expressions to explore whether dynamic movements consistently improve recognition efficacy to both human perception and automated analysis. The recognition tasks spanned across both posed and spontaneous expressions, employing three distinct formats: target static, non-target static, and dynamic. This structured approach allowed us to directly assess the potential benefits that dynamic information provides, challenging existing paradigms that have predominantly relied on static images. The comparative analysis between human and machine recognition aimed to investigate the importance of dynamic information in both entities, thereby contributing to the refinements of computational models for more lifelike emotion detection.

Consistent with expectations, findings from Chapter 2 demonstrated that the dynamic qualities of expressions significantly influenced recognition accuracy, particularly for non-target static expressions that were less prototypical and complex but more ambiguous in their appearance. Interestingly, this recognitional advantage of dynamics was not evident when dealing with target-static images that were clear-cut in their prototypical features and complexity while being low in ambiguity. In this vein, the findings suggest that the role of dynamic information in aiding recognition is contingent upon the initial recognisability of static expressions; when static recognition rates are already near perfect, dynamic cues may offer limited additional clarity. However, most stimuli used in Chapter 2 were selected from posed databases, which inherently differ from the more subtle, less spontaneous expressions commonly observed in natural settings. It is therefore likely to be the case that dynamic expression may offer a more significant impact for effective social communication in daily interaction.

Transitioning to the research on spontaneous aspects, Chapters 3 and 4 collectively highlighted the significant variability inherent in spontaneous expression databases. Unlike posed expressions, which are typically generated through direct instructions to activate specific facial muscle movements or convey certain emotions (Cosker et al., 2011; Van Der Schalk et al., 2011), spontaneous expressions emerge from a myriad of elicitation techniques, adding layers of complexity to their annotation and subsequent evaluation. Regarding spontaneous databases, I observed a broad spectrum of emotional content, extending from the basic six emotions to a wider range of context-specific nuanced emotional states. This multifaceted nature of spontaneous expressions introduces substantial variability, not only in how these expressions are encoded and aligned with the intended emotions but also in their emotional intensity and clarity.

Consequently, the inherent variability across expressions has profound implications for the decoding of spontaneous expressions. Corroborating this notion, our empirical evaluation presented in Chapter 4 demonstrates that spontaneous databases exhibit significant differences in emotion recognition accuracy. Additionally, the analysis revealed that distinct AUs play varying roles in the successful recognition of these expressions, reinforcing the earlier research findings pointing towards variable morphological patterns in spontaneous expressions (Namba et al., 2017; Girard et al., 2015). These findings compel a re-evaluation of current methodologies in facial expression research, suggesting approaches that embrace the nuanced and dynamic nature of real-world emotional expressions. The studies presented in current research call for a more granular understanding of spontaneous expressions, highlighting the limitations of conventional methods that fail to address the complexity and diversity of human emotional experience.

Delving deeper into the distinctions among types of facial expressions, it is imperative to consider naturalistic expressions beyond the traditional dichotomy between posed and spontaneous categories, a topic that has not been extensively covered in this dissertation and in previous literature. Here, I define naturalistic expressions as expressions that occur in uncontrolled settings (Bian et al., 2024). These expressions starkly diverge from posed and spontaneous expressions, as they unfold naturally in real-life situations without any predetermined experimental conditions or elicitation methods. Emerging from authentic human interactions, naturalistic expressions capture the subtleties and complexities of emotions as they occur in daily experiences.

Naturalistic expressions may stand as the most ecologically valid expression in facial expression research, capturing the full spectrum of emotional experiences in their authentic context. These expressions are not limited to experimental control, thereby encompassing both visible emotional reactions and the contextual factors and interpersonal dynamics that influence these expressions. This unfiltered glimpse into genuine emotional experiences is crucial for understanding the intricate dynamics of human interaction and emotional communication.

However, the study of naturalistic expressions introduces distinct challenges, especially regarding data collection, ethical considerations, and the recognition process. Until now, most naturalistic databases collect portrayals from online platforms (e.g., YouTube, films; Erdem et al., 2014; Rosas et al., 2013), for which it is challenging to obtain consent from encoders. Given that they manifest in real-world social interactions, these expressions demand further attention to ethical standards, surpassing the requirements (e.g., controlled environment) typically associated when collecting posed or spontaneous expressions in a laboratory. Additionally, the unstructured nature of natural expressions adds layers of complexity to the accurate segmentation and interpretation of these expressions, often resulting in low recognition accuracy (Han et al., 2020).

Despite these hurdles, investigating naturalistic expressions offers unparalleled opportunities to enhance our grasp of authentic human emotional communication. This pursuit necessitates the adoption of novel methodologies to collect these expressions (e.g., one-to-one interaction, social interview) and technological advancements while upholding rigorous ethical guidelines. As this field expands, there is a growing imperative to forge new analytical frameworks. These frameworks should be designed to effectively integrate advanced computational techniques and robust data handling strategies, capable of capturing and recognising the unique, unscripted nuances presented by naturalistic expressions, thereby pushing the boundaries of our understanding of real-world emotional dynamics.

5.3 Implications and Contributions

The present dissertation offers empirical insights into the effects of dynamic and spontaneous elements in facial expression recognition, addressing seven key questions raised throughout the literature review. Firstly, this dissertation addresses the discourse surrounding the dynamic advantage in facial expression recognition, where prior research has yield mixed findings. Many studies have utilised degraded or suboptimal stimuli to demonstrate the facilitative effects of movement. However, the present research demonstrated that dynamic information facilitates recognition primarily when static cues are less informative regardless of visual conditions, such as in non-target expressions. While dynamic information did not offer a consistent advantage over static representations of peak moments (i.e., target expressions), it played an important role when the static images lacked clarity with high ambiguity. This aligns with the notion that the dynamic advantage is context-dependent, emerging most clearly in situations where static representations fail to provide sufficient emotional information.

Secondly, the dissertation explored whether the variability in static frames extracted from dynamic sequences influences recognition. Previous research has often compared static and dynamic expressions based on a singular static frame (Bassili et al., 1988, Bould & Morris, 2008; Wallraven et al., 2008), typically representing a peak moment. Despite the present research intended to extract ambiguous frames (though most emotive within sequence), findings revealed that the representativeness of static frames plays a crucial role in recognition. When a static frame captured the peak expression, it provided enough emotional clarity, making dynamic information unnecessary. However, in non-target images, dynamic sequences offered an advantage by providing additional temporal cues that facilitated recognition. This suggests that the timing of static frame extraction within a dynamic sequence significantly impacts its recognisability. The research again emphasised that dynamic sequences are more consistent in aiding recognition when static frames are less representative of the target emotion.

Thirdly, the present dissertation examined the featural parameters – prototypicality, intensity, and ambiguity – that influence recognition across static and dynamic, as well as posed and spontaneous expressions. While prior studies often did not delve into these factors in depth, the present work revealed that all three parameters collectively impact recognition. Generally, prototypicality and complexity confer a positive effect on recognition, while ambiguity reduces recognisability. This exploration also elucidates specific conditions under which motion enhances or fails to impact recognition. Moreover, this research provides empirical evidence on how featural characteristics inherent in everyday expressions shape recognition. Until now, previous studies primarily examine the effects of prototypicality (Ekman, 1992, 2003; Matsumoto & Hwang, 2014), ambiguity (Du et al., 2014; Fiorentini & Viviani, 2009; Kinchella & Guo, 2021) and intensity (Calder et al., 2000; Recio et al., 2014) have been examined in isolation. Our research reveals the synergetic effect among these factors, fundamentally altering our understanding of the facial expression recognition system.

Fourthly, the research extended beyond human perceptual analysis to explore how machines, like humans, respond to dynamic expressions. Previous research has shown that machines often struggle with dynamic expressions, particularly when compared to peak static snapshots. However, the present findings demonstrated that machines exhibit a dynamic advantage similar to humans, especially in recognising non-target expressions where ambiguity is high. While machines outperformed humans in recognising target images, they also showed comparable (or even better) performance to humans in dynamic context, particularly when dealing with subtle or ambiguous expressions. This suggests that machines can benefit from dynamic information, much like humans, when static cues are insufficient.

Fifthly, the research addressed the inconsistency in findings regarding human versus machine recognition capabilities. While some studies have shown humans outperforming machines for subtle expressions, others have suggested the opposite (Krumhuber et al., 2021b; Yitzhak et al., 2017). Building on prior research showing comparable performance of the machine to human observers, this work showed a machine advantage across different temporal phases, including peak and other time points of facial displays. Moreover, this research emphasises the practicality of automated recognition in reducing the costs and resources associated with video rating studies, allowing for more efficient data processing. Such findings suggest that algorithmic models are capable of handling not only in peak intensity of expressions but also dynamic and non-target expressions that are less standardised, subtle and ambiguous in their appearance. This shift towards a more cost-effective approach could significantly impact fields requiring the decoding of facial expressions, from psychological assessments to interactive media and user interface design.

Sixthly, the present dissertation first unveils an extensive array of database resources dedicated to the study of spontaneous and dynamic facial expressions, significantly contributing to the advancement of more ecologically valid research in this domain. While previous FEDB reviews primarily centred on posed and static databases, with only a handful of spontaneous datasets being explored (Diconne et al., 2022; Haamer et al., 2018), this dissertation broadens the scope. It systematically reviewed the diverse properties of FEDBs, highlighting how these features impact spontaneous facial expression recognition. This detailed examination serves as a crucial guide for researchers, aiding them in the careful selection of datasets that align with their specific needs. Additionally, this current work identifies remaining gaps within existing FEDB and expands the methodological approaches for studying spontaneous expressions, thereby illustrating the gaps further advancement is needed. This paves the way for the development and integration of more diverse and realistic portrayals of emotional behaviours.

Finally, the present research conducted a cross-corpus evaluation of spontaneous FEDBs, which addressed the question of whether significant variability exists in the recognition rates of spontaneous expressions across different databases. The findings aligned with and extended beyond previous research (Motley & Camden, 1988; Ngo et al., 2015; Tong et al., 2010) by demonstrating that spontaneous expressions, while difficult to recognise, exhibit varied accuracy rates across different databases. This variation seems not merely incidental but reflects the inherent diversity of human emotions as they manifest spontaneously. Such a revelation implies that research, both in human expressions and machine recognition, must adopt a more careful selection of stimuli that fit their research objectives. Additionally, the research found that recognition of spontaneous expression is frequently determined by a select few prototypes or even isolated AUs, including those not traditionally outlined in the FACS manual (e.g., AU7 for disgust). This result aligns with earlier studies suggesting that spontaneous expressions have more flexible AU patterns (Namba et al., 2017). The significant roles of prototypicality, ambiguity, and complexity in spontaneous expression recognition can be considered as an extension of our earlier findings (Kim et al., 2023), emphasising their essential impact on posed as well as spontaneous expression recognition. Notably, although our prior results demonstrated that automated classifiers possess the capability to classify subtle expressions, their average performance in classifying spontaneous expressions just exceeded the chance level. This performance rate may imply that the recognisability of spontaneous expressions may be influenced by other factors (e.g., conceptual factors, such as elicitation methods, encoder demographics) besides their subtlety. The current dissertation offers initial evidence of variability in spontaneous expression recognition throughout cross-corpus evaluation, aiming to clarify the determinants that influence spontaneous expression recognition.

5.4 Limitations and Future Research

Despite the significant contributions of this dissertation to addressing key questions in facial expression research, it is important to acknowledge specific limitations intrinsic to its methodological framework. One such limitation is the predominant utilisation of posed databases to assess the dynamic advantage in recognition. Spontaneous expressions, as opposed to posed ones, often feature more prolonged periods of subtle emotional displays (Hess et al., 1990), suggesting that the impact of movement on recognition could significantly differ between these expression types. Moreover, previous research has demonstrated that posed and spontaneous expressions exhibit distinct temporal dynamics in terms of timing, speed, and duration of onset, offset and apex (Cohn & Schmidt, 2004; Hess & Kleck, 1990;

Schmidt et al., 2009). This distinction necessitates future research to investigate the role of movement in spontaneous expression recognition.

Expanding the scope to spontaneous expressions could substantially refine our comprehension of the dynamic advantage. This adjustment would align with the evolving research paradigm that prioritises ecological validity, thereby facilitating a more realistic approximation of authentic human emotion recognition. Moreover, this approach could provide a clearer delineation of the conditions under which dynamic cues assert their facilitative benefits or detract from the recognition of genuine emotional states.

Given the predominant use of posed expressions in Chapter 2, such reliance may inadvertently bias recognition rates towards superior machine performance. In the present research, I demonstrated that machines outperformed human observers regardless of expression format (although mostly posed). Yet, this finding necessitates careful consideration of the context in which these algorithms excel and highlights the importance of the types of expressions used in training and testing phases. As automated classifiers are commonly trained using posed databases (Pantic & Barrett, 2007), they benefit highly from the uniformity of standardised expressions, favouring machine-based featural analysis. To counteract this bias and achieve a more equitable comparison, future research should incorporate a more balanced selection of posed and spontaneous expressions. This approach would ensure a realistic assessment of machine versus human performance in expression recognition, thereby refining the development and evaluation of automated systems and enhancing their applicability in real-world scenarios.

Such limitation in automated classifiers naturally lead to another important limitation issue: the inconsistent performance in the AFFDEX classifier, particularly its struggle with the accurate classification of certain emotional expressions, as demonstrated in Chapter 4. AFFDEX showed comparatively lower recognition accuracy for spontaneous and non-prototypical expressions, where misclassification rates were considerably high. This pattern in AFFDEX may stem from its reliance on proprietary mechanisms and the lack of transparency regarding its training datasets. Combined with earlier findings showing successful performance of AFFDEX for certain databases (Dupré et al., 2020; McDuff, 2017), the misclassification issues suggests that while such tools are effective in controlled conditions, they may be less reliable when applied to unfamiliar emotional displays that largely deviate from the training datasets. In this regard, single classifier-based evaluations pose significant risk, particularly when it comes to commercial software like AFFDEX, where necessary details, such as the underlying training datasets, are often not disclosed. The opacity of these proprietary systems makes it difficult to fully understand the biases and gaps inherent in their performance. For example, without access to the exact training data or the weightings assigned to different facial features, it becomes challenging to gauge why specific misclassification occur more frequently.

Future research must address this limitation by adopting a multi-system comparison approach. Rather than relying solely on single AFEA tool, including a broader array of classifiers with transparent architectures can facilitate more comprehensive evaluations. Doing so would provide a clearer understanding of not only the strengths and weakness of various systems but also the differences between testing datasets that influence classifiers' performance. Additionally, collaborations between academia and industry should strive to improve transparency in commercial tools, making it easier for researchers to identify where improvements in recognition algorithms are needed.

Another limitation of the present research stems from its focus on basic emotions in the recognition of spontaneous and dynamic expression, aligning with traditional trends in emotion research (Keltner et al., 2019). However, this framework may not adequately capture the complexity and subtlety of human emotional expressions, particularly in spontaneous forms that often embody a blend of multiple emotions (Du et al., 2014; Du & Martinez, 2015) or subtler affective nuances not covered by the BET (Russell, 1980). This restriction could limit the generalisability of our findings, confining them primarily to basic emotions while overlooking numerous emotional states.

As highlighted in the review presented in Chapter 3, spontaneous databases frequently capture a range of expressions that go beyond the conventional basic categories. Despite our focus on basic emotions being primarily determined by the capabilities and cost-effectiveness of commercial software tools (AFFDEX, FACET), this narrow focus may limit a thorough understanding of the varied emotional experiences that humans exhibit, especially in naturalistic settings. This restriction also hinders the complete assessment of several databases that feature expressions beyond basic emotions. Consequently, future research could benefit from utilising more sophisticated automated recognition algorithms trained with diverse stimulus sets that are equipped to classify a broader spectrum of emotional states.

One of the primary objectives of this dissertation was to delve into the role of dynamic information in facial expression recognition, a topic that remains inconclusive with varied conclusions in prior research. Despite converging evidence indicating that dynamic cues enhance recognition (Ambadar et al., 2005; Cunningham & Wallraven, 2009; Wehrle et al., 2000), and our own findings suggesting a compensatory role of dynamic information (Kim et al., 2023), gaps remain in our understanding, particularly regarding how different dynamic elements interplay at emotion recognition.

Future research could also investigate how the distinct phases of facial expressions contribute to accurate emotion identification. Dynamic expressions are typically categorised into onset, apex, and offset phases (Krumhuber et al., 2013), each bearing unique characteristics that could differentially influence recognition. A pertinent question arises: Are the facilitative effects more pronounced during the onset to apex phase due to its gradual unfolding, or do they become more apparent during the apex to offset phase, which accentuates the peak of the expressions (Fiorentini & Viviani, 2011; Recio et al., 2011; Yoshikawa & Sato, 2008)? Considering that onset and offset phases typically consist of subtler displays, their comparative impact on expression recognizability warrants thorough investigation. Such future studies could isolate specific movements within these phases to ascertain their individual contribution to the recognition of various emotional states.

Further extending the investigation of dynamic movements in facial expressions, future study should consider the role of speed and rhythm in facial muscle movements as potential predictors of emotion recognition. While the present work primarily focused on AU analysis to categorise emotional expressions, converging evidence suggests that the temporal dynamics – specifically, the speed and rhythm of facial movements - may provide additional, if not more precise, information for emotion classification (Krumhuber et al., 2013; Wehrle et al., 2000). These dynamic features may provide vital information that AUs alone cannot capture, particularly when distinguishing between subtle variations in emotional intensity.

The speed of facial movement is a crucial temporal dimension that could refine emotion recognition by indicating the intensity and immediacy of an emotional display. Fast, jerky movements often associated with more intense and urgent emotional states, such as anger, or excitement, as they reflect heightened muscle activation (Ambadar et al., 2005). For instance, the rapid eyebrows raise or swift mouth opening commonly signal surprise or fear. In contrast, slower movements tend to be linked to subtler, more restrained emotions, such as sadness or contemplation, where the muscle activation is less pronounced (Cohn & Schmidt, 2004). These temporal cues are particularly helpful in providing context-specific information that may clarify emotional states that could otherwise appear similar in a static frame. For example, a subtle smile and a smirk might be difficult to differentiate in a single static snapshot but becomes clear when observing the speed of the facial movements over time.

Beyond speed, the rhythm or flow of facial movements is another critical factor that could enhance emotion recognition. The smoothness and regularity of facial movement rhythms often contribute to perceptions of authenticity or genuineness in emotional displays. Genuine expressions, such as spontaneous smiles, typically exhibit a natural, rhythmic flow from onset to apex, while posed or exaggerated expressions tend to have irregular or disjointed movements, signalling inauthenticity (Ekman & Rosenberg, 2005; Cohn & Schmidt, 2004). In ambiguous expressions, where static features may not offer enough or contrasting information, rhythm can provide an additional layer of differentiation. By analysing the temporal structure of these expressions, future studies may be able to detect subtle emotional nuances that go unnoticed in static frames or morphology-based models.

Considering the compensatory role of dynamic information highlighted in the current research, the inclusion of speed and rhythm analysis could represent a crucial next step for advancing the field. Current automated recognition systems often focus on isolated moments of expression, neglecting the rich temporal dynamics that these features offer. Understanding how quickly expressions unfold and the fluidity with which they do so could improve recognition algorithms, enabling them to classify ambiguous or subtle expressions more accurately. Moreover, incorporating temporal dynamics into affective computing systems could make machine-based emotion recognition more human-like, allowing for real-time responses to the emotional states of others.

Expanding upon the role of movement in facial expression recognition, it is imperative in the future to delve deeper into how dynamic cues impact perceptions beyond mere emotion recognition, such as expression genuineness. Prior studies, such as those by Krumhuber and colleagues (2007), have illustrated the pivotal role of facial dynamics in cultivating the perception of trustworthiness and cooperative behaviour. In this context, the specific patterns or phases within an expression may hold key insights into perceptual outcomes (Dzhelyova et al., 2012). Additionally, while recognition rates for subtle and ambiguous expressions are typically lower, prior studies have indicated that such expressions are frequently perceived as more genuine (Ngo et al., 2015; Yitzhak et al., 2017). Such findings raise intriguing questions about whether movement impacts the perception of genuineness differently across various types of expressions, such as peak versus non-peak (or target versus non-target) expressions, suggesting a potential divergence from patterns identified in our research.

Future research could also examine how various factors, such as elicitation methods, influence the encoding and decoding of facial expressions. Previous works by Gross and Levenson (1995) and Mehu and Scherer (2015) have highlighted how emotional expressivity and interpretability vary substantially with different elicitation techniques, from passive reaction to autobiographical recall of emotional experiences. Despite this evidence, spontaneous expressions in existing research have often been approached as a monolithic phenomenon, lacking the dimensional depth that these varied elicitation techniques embody (Motley & Camden, 1988). Given the inherent variability in affective quality produced by different elicitation techniques, the findings from Chapter 4 revealing divergent recognition rates, may be unsurprising and provide evidence of how these methods impact the recognition of expressions. This variability seemingly mirrors the complexity seen in real-world expressions and challenges the prevailing notions of spontaneous expressions as uniform. Research by Aviezer, Ensenberg and Hassin (2017), which demonstrated that environmental context dramatically influences the perception of emotions, further supports the necessity for a multidimensional approach to understanding spontaneous expressions.

To this end, subsequent studies should extend beyond the traditional scope, exploring the multidimensional nature of spontaneous expressions as shaped by different elicitation methods, this could involve dissecting how specific methods influence the perceived intensity, authenticity, and ultimately the recognisability of expressions. Furthermore, by incorporating neurological findings in expression processing expressions and judgement (Jabbi et al., 2008; Sato et al., 2004), particularly for ambiguous expressions (Ito et al., 2017), it would be interesting to see how these brain regions respond to spontaneous expressions elicited through different methods. Given that posed and spontaneous expressions engage distinct neural pathways (Frank et al., 1993; Rinn, 1984), investigating the neural responses associated with different elicitation techniques may help elucidate the underlying mechanisms of emotion processing and reveal how authenticity and context influence neural activation patterns.

Acknowledging the impact of environmental context on emotional expressions, an essential avenue for further exploration emerges: the differentiation between spontaneous and naturalistic expressions in emotional research. Our review showed that spontaneous expressions are often elicited in laboratory settings, aiming to simulate real-life emotional reactions to emotion-evoking stimuli. Although this approach provides a good trade-off between experimental control and naturalness, it can still carry elements of artificiality, particularly concerning the recording environment. In contrast, naturalistic expressions emerge unprompted in everyday scenarios, presenting an authentic glimpse into human emotional experiences. Moreover, while laboratory-induced spontaneous expressions assume a direct relationship between the stimulus and its resultant expressions (Lucey et al., 2010), the recent findings assert that naturalistic expressions reflect a broader spectrum of individual responses shaped by personal appraisal of the situation (Lazarus, 1991; Schmidt et al., 2010). Consequently, expressions captured in natural settings might encompass a range of reactions from genuinely spontaneous to deliberately posed, contingent upon the individuals' interpretation and response to their surroundings.

Despite their distinct origins, spontaneous and naturalistic expressions have frequently been conflated in research, and mistakenly assumed to be interchangeable. This common oversight neglects the intricate complexity of emotional experiences as they naturally unfold

(Barrett et al., 2011). Additional factors, such as varying backgrounds, further complicate the recognition and interpretation of expression (Righart & De Gelder, 2008; Sannikov et al., 2017). Illustrative of this, recent studies by Kroczek and colleagues (2022) and Hsu, Sato and Yoshikawa (2020) have demonstrated that naturalistic expressions, captured in social interactions, can elicit stronger observer responses than their spontaneous counterparts generated in artificial settings. These empirical findings challenge the presumption of their interchangeability and call for a refined approach to categorise and analyse emotional expressions. In light of these insights, future investigations should attempt not only to establish clearer distinctions between spontaneous and naturalistic expressions but also to explore how these varying types of expressions are perceived and interpreted by observers.

In naturalistic environments, facial expressions are likely to co-occur with speech, merging verbal and non-verbal signals (Cai et al., 2020). This interplay between modalities is pivotal for effective social communication, as the congruence or discordance between verbal content and facial expressions can significantly influence how messages are perceived and interpreted. However, the intricacies of such spontaneous/naturalistic multimodal communication remain underexplored, particularly in how individuals integrate and prioritise these signals when they conflict. For example, a negative verbal message coupled with positive facial expressions presents unique challenges to the observers, complicating traditional communication that relies on single-channel models (Russell et al., 2017; Walther & D'Addario, 2001). Furthermore, mouth movements, even in the absence of clear verbal content, can significantly influence the interpretation of facial expressions, contributing to the complexity (Eisenbarth & Alpers, 2011; Meeten et al., 2015). Theories of multimodal communication posit that a holistic interpretation of emotional messages requires the integration of disparate signals, necessitating observers to reconcile multifaceted cues into a coherent narrative (Jokinen & Wilcock, 2012).

Given this backdrop, future research should extend beyond the traditional analysis of isolated facial expressions to consider the interplay between facial expressions and speech within spontaneous and naturalistic contexts. Besides verbal cues, physiological signals have increasingly become central to the study of multimodal emotional communication (Saganowski et al., 2022). These signals, such as heart rate and skin conductance, offer additional insight beyond what facial expressions or verbal messages alone can convey. For instance, fear and surprise facial expressions are often confusedly recognized due to their similar morphological patterns (Zhao et al., 2017). Physiological responses can aid in differentiating between these emotions by revealing the underlying autonomic arousal associated with each (Shu et al., 2018). This multimodal approach, integrating facial, verbal, and physiological data, is essential for understanding how emotions are expressed and perceived in everyday life.

In conclusion, this dissertation opens up several new avenues for future research by highlighting the spontaneous and dynamic aspects of facial expressions. It lays the foundation for future research to adopt a more ecologically valid approach to investigating facial expression. I hope this work acts as a catalyst for further exploration into the complex dynamics of facial expressions in spontaneous and naturalistic settings

References

Adams, A., Mahmoud, M., Baltrusaitis, T., & Robinson, P. (2015). Decoupling facial expressions and head motions in complex emotions. *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, 274–280. https://doi.org/10.1109/ACII.2015.7344583

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, *45*(7), 1363–1377. https://doi.org/10.1016/j.neuropsychologia.2006.10.016

Adolphs, R., & Tranel, D. (2004). Impaired judgments of sadness but not happiness following bilateral amygdala damage. *Journal of Cognitive Neuroscience*, *16*(3), 453–462. https://doi.org/10.1162/089892904322926782

Aifanti, N., Papachristou, C., & Delopoulos, A. (2010). *The MUG facial expression database*. *11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10*, Desenzano del Garda, Italy, pp1–4.

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, *16*(5), 403–410. https://doi.org/10.1111/j.0956-7976.2005.01548.x

Arsalidou, M., Morris, D., & Taylor, M. J. (2011). Converging evidence for the advantage of dynamic facial expressions. *Brain Topography*, *24*(2), 149–163. https://doi.org/10.1007/s10548-011-0171-4

Hyunwoo Kim

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion Perception from Dynamic and Static Body Expressions in Point-Light and Full-Light Displays. *Perception* 33, 717–746. https://doi.org/10.1068/p5096.

Atkinson, A. P., Vuong, Q. C., & Smithson, H. E. (2012). Modulation of the face- and body-selective visual regions by the motion and emotion of point-light face and body stimuli. *NeuroImage*, *59*(2), 1700–1712. https://doi.org/10.1016/j.neuroimage.2011.08.073

Aung, M. S. H., Kaltwang, S., Romera-Paredes, B., Martinez, B., Singh, A., Cella, M., Valstar, M., Meng, H., Kemp, A., Shafizadeh, M., Elkins, A. C., Kanakam, N., De Rothschild, A., Tyler, N., Watson, P. J., Williams, A. C. D. C., Pantic, M., & Bianchi-Berthouze, N. (2016). The Automatic Detection of Chronic Pain-Related Expression: Requirements, Challenges and the Multimodal EmoPain Dataset. *IEEE Transactions on Affective Computing*, *7*(4), 435–451. https://doi.org/10.1109/TAFFC.2015.2462830

Aviezer, H., Ensenberg, N., & Hassin, R. R. (2017). The inherently contextualized nature of facial emotion perception. *Current Opinion in Psychology*, *17*, 47–54. https://doi.org/10.1016/j.copsyc.2017.06.006

Baltrusaitis, T., Mahmoud, M., & Robinson, P. (2015). Cross-dataset learning and person-specific normalisation for automatic Action Unit detection. *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 1–6. https://doi.org/10.1109/FG.2015.7284869

Bänziger, T., & Scherer, K. R. (2007). Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The GEMEP Corpus. In A. C. R. Paiva, R. Prada, & R. W. Picard (Eds.), *Affective Computing and Intelligent Interaction* (Vol. 4738, pp. 476–487). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74889-2_42

Hyunwoo Kim

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion*, *12*(5), 1161–1179. https://doi.org/10.1037/a0025827

Barrett, L. F. (2006). Are Emotions Natural Kinds? *Perspectives on Psychological Science*, *1*(1), 28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in Emotion Perception. *Current Directions in Psychological Science*, *20*(5), 286–290. https://doi.org/10.1177/0963721411422522

Bartlett, M. S., Littlewort, G. C., Frank, M. G., Lainscsek, C., Fasel, I. R., & Movellan, J. R. (2006). Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, *1*(6), 22–35. https://doi.org/10.4304/jmm.1.6.22-35

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(3), 373–379. https://doi.org/10.1037/0096-1523.4.3.373

Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*, *37*(11), 2049–2058. https://doi.org/10.1037/0022-3514.37.11.2049

Ben, X., Ren, Y., Zhang, J., Wang, S.-J., Kpalma, K., Meng, W., & Liu, Y.-J. (2021). Video-based Facial Micro-Expression Analysis: A Survey of Datasets, Features and Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. https://doi.org/10.1109/TPAMI.2021.3067464

Hyunwoo Kim

Benitez-Quiroz, C. F., Srinivasan, R., & Martinez, A. M. (2016). EmotioNet: An Accurate, Real-Time Algorithm for the Automatic Annotation of a Million Facial Expressions in the Wild. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV, USA: IEEE), 5562–5570. https://doi.org/10.1109/CVPR.2016.600.

Benitez-Quiroz, C. F., Wilbur, R. B., & Martinez, A. M. (2016). The not face: A grammaticalization of facial expressions of emotion. *Cognition*, *150*, 77–84. https://doi.org/10.1016/j.cognition.2016.02.004

Berenbaum, H., & Rotter, A. (1992). The relationship between spontaneous facial expressions of emotion and voluntary control of facial muscles. *Journal of Nonverbal Behavior*, *16*(3), 179–190. https://doi.org/10.1007/BF00988033

Beringer, M., Spohn, F., Hildebrandt, A., Wacker, J., & Recio, G. (2019). Reliability and validity of machine vision for the assessment of facial expressions. *Cognitive Systems Research*, *56*, 119–132. https://doi.org/10.1016/j.cogsys.2019.03.009

Bian, C., Zhang, Y., Yang, F., Bi, W., & Lu, W. (2019). Spontaneous facial expression database for academic emotion inference in online learning. *IET Computer Vision*, *13*(3), 329–337. https://doi.org/10.1049/iet-cvi.2018.5281

Bian, Y., Küster, D., Liu, H., & Krumhuber, E. G. (2023). Understanding Naturalistic Facial Expressions with Deep Learning and Multimodal Large Language Models. *Sensors*, *24*(1), 126. https://doi.org/10.3390/s24010126

Bidet-Ildei, C., Decatoire, A., & Gil, S. (2020). Recognition of Emotions From Facial Point-Light Displays. *Frontiers in Psychology*, *11*, 1062. https://doi.org/10.3389/fpsyg.2020.01062

Hyunwoo Kim

Biele, C., & Grabowska, A. (2006). Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Exp Brain Res* 171, 1–6. https://doi.org/10.1007/s00221-005-0254-0.

Bishay, M., Preston, K., Strafuss, M., Page, G., Turcot, J., & Mavadati, M. (2023). AFFDEX 2.0: A Real-Time Facial Expression Analysis Toolkit. *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, 1–8. https://doi.org/10.1109/FG57933.2023.10042673

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

Bonanno, G., & Keltner, D. (2004). Brief Report The coherence of emotion systems: Comparing "on-line" measures of appraisal and facial expressions, and self-report. *Cognition & Emotion*, *18*(3), 431–444. https://doi.org/10.1080/02699930341000149

Bornemann, B., Winkielman, P., & Der Meer, E. V. (2012). Can you feel what you do not see? Using internal feedback to detect briefly presented emotional stimuli. *International Journal of Psychophysiology*, *85*(1), 116–124. https://doi.org/10.1016/j.ijpsycho.2011.04.007

Boucher, J. D. (1969). Facial Displays of Fear, Sadness and Pain. *Perceptual and Motor Skills*, *28*(1), 239–242. https://doi.org/10.2466/pms.1969.28.1.239

Bould, E., & Morris, N. (2008). Role of motion signals in recognizing subtle facial expressions of emotion. *British Journal of Psychology*, *99*(2), 167–189. https://doi.org/10.1348/000712607X206702

Brave, S., & Nass, C. (2009). Emotion in Human–Computer Interaction. In A. Sears & J. Jacko (Eds.), *Human-Computer Interaction Fundamentals* (Vol. 20094635, pp. 53–68). CRC Press. https://doi.org/10.1201/b10368-6

Hyunwoo Kim

Bulagang, A. F., Weng, N. G., Mountstephens, J., & Teo, J. (2020). A review of recent approaches for emotion classification using electrocardiography and electrodermography signals. *Informatics in Medicine Unlocked*, *20*, 100363. https://doi.org/10.1016/j.imu.2020.100363

Cabeza, R., Bruce, V., Kato, T., & Oda, M. (1999). The prototype effect in face recognition: Extension and limits. *Memory & Cognition*, *27*(1), 139–151. https://doi.org/10.3758/BF03201220

Cai, L., Dong, J., & Wei, M. (2020). Multi-Modal Emotion Recognition From Speech and Facial Expression Based on Deep Learning. *2020 Chinese Automation Congress (CAC)*, 5726–5729. https://doi.org/10.1109/CAC51589.2020.9327178

Calder, A. J. (1996). Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology*, *13*(5), 699–745. https://doi.org/10.1080/026432996381890

Calder, A. J., Rowland, D., Young, A. W., Nimmo-Smith, I., Keane, J., & Perrett, D. I. (2000a). Caricaturing facial expressions. *Cognition* 76, 105–146. https://doi.org/10.1016/S0010-0277(00)00074-3.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000b). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance* 26, 527–551. https://doi.org/10.1037/0096-1523.26.2.527.

Calvo, M. G., & Fernández-Martín, A. (2013). Can the eyes reveal a person's emotions? Biasing role of the mouth expression. *Motiv Emot* 37, 202–211. https://doi.org/10.1007/s11031-012-9298-1.

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods*, *40*(1), 109–115. https://doi.org/10.3758/BRM.40.1.109

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion*, *30*(6), 1081–1106. https://doi.org/10.1080/02699931.2015.1049124

Calvo, M. G., Avero, P., Fernández-Martín, A., & Recio, G. (2016). Recognition thresholds for static and dynamic emotional faces. *Emotion*, *16*(8), 1186–1200. https://doi.org/10.1037/emo0000192

Calvo, M. G., Fernández-Martín, A., & Nummenmaa, L. (2012). Perceptual, categorical, and affective processing of ambiguous smiling facial expressions. *Cognition*, *125*(3), 373–393. https://doi.org/10.1016/j.cognition.2012.07.021

Calvo, M. G., Fernández-Martín, A., Recio, G., & Lundqvist, D. (2018). Human observers and automated assessment of dynamic emotional facial expressions: KDEF-dyn database validation. *Frontiers in Psychology*, *9*, 2052. https://doi.org/10.3389/fpsyg.2018.02052

Calvo, M. G., Gutiérrez-García, A., Fernández-Martín, A., & Nummenmaa, L. (2014). Recognition of Facial Expressions of Emotion is Related to their Frequency in Everyday Life. *J Nonverbal Behav* 38, 549–567. https://doi.org/10.1007/s10919-014-0191-3.

Calvo, R. A., & D'Mello, S. (2010). Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *IEEE Transactions on Affective Computing*, *1*(1), 18–37. https://doi.org/10.1109/T-AFFC.2010.1

Carroll, J. M., & Russell, J. A. (1997). Facial expressions in Hollywood's portrayal of emotion. *Journal of Personality and Social Psychology*, *72*(1), 164–176. https://doi.org/10.1037/0022-3514.72.1.164

Cassidy, S., Mitchell, P., Chapman, P., & Ropar, D. (2015). Processing of Spontaneous Emotional Responses in Adolescents and Adults with Autism Spectrum Disorders:

Effect of Stimulus Type. *Autism Research*, *8*(5), 534–544. https://doi.org/10.1002/aur.1468

Cavicchio, F., & Poesio, M. (2012). The Rovereto Emotion and Cooperation Corpus: A new resource to investigate cooperation and emotions. *Language Resources and Evaluation*, *46*(1), 117–130. https://doi.org/10.1007/s10579-011-9163-y

Chanti, D. A., & Caplier, A. (2018). Spontaneous Facial Expression Recognition using Sparse Representation. *ArXiv*, https://doi.org/10.48550/ARXIV.1810.00362

Chen, D., Song, P., & Zheng, W. (2023). Learning Transferable Sparse Representations for Cross-Corpus Facial Expression Recognition. *IEEE Transactions on Affective Computing*, *14*(2), 1322–1333. https://doi.org/10.1109/TAFFC.2021.3077489

Chen, Y., & Joo, J. (2021). Understanding and Mitigating Annotation Bias in Facial Expression Recognition. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 14960–14971. https://doi.org/10.1109/ICCV48922.2021.01471

Cheng, S., Kotsia, I., Pantic, M., & Zafeiriou, S. (2017). 4DFAB: A Large Scale 4D Facial Expression Database for Biometric Applications. *ArXiv,* https://doi.org/10.48550/ARXIV.1712.01443

Cheng, X., Xue, L., Chen, Y., & Liu, L. (2014). Chinese Facial Expression Database with spontaneous expressions across multiple poses. *Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 1–6. https://doi.org/10.1109/ICCCNT.2014.6963009

Cheung, E. O., Slotter, E. B., & Gardner, W. L. (2015). Are you feeling what I'm feeling? The role of facial mimicry in facilitating reconnection following social exclusion. *Motivation and Emotion*, *39*(4), 613–630. https://doi.org/10.1007/s11031-015-9479-9

Coan, J. A., & Allen, J. J. B. (Eds.). (2007). *Handbook of emotion elicitation and assessment*. Oxford University Press.

Hyunwoo Kim

Cohn, J. F., & Sayette, M. A. (2010). Spontaneous facial expression in a small group can be automatically measured: An initial demonstration. *Behavior Research Methods*, *42*(4), 1079–1086. https://doi.org/10.3758/BRM.42.4.1079

Cohn, J. F., & Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, *02*(02), 121–132. https://doi.org/10.1142/S021969130400041X

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. *The handbook of emotion elicitation and assessment*, *1*(3), 203-221.

Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion*, *18*(1), 75–93. https://doi.org/10.1037/emo0000302

Cosker, D., Krumhuber, E., & Hilton, A. (2011). A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling. *2011 International Conference on Computer Vision*, 2296–2303. https://doi.org/10.1109/ICCV.2011.6126510

Costantini, E., Pianesi, F., & Prete, M. (2005). Recognising emotions in human and synthetic faces: The role of the upper and lower parts of the face. *Proceedings of the 10th International Conference on Intelligent User Interfaces*, 20–27. https://doi.org/10.1145/1040830.1040846

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, *114*(38). https://doi.org/10.1073/pnas.1702247114

Hyunwoo Kim

Cowen, A. S., Keltner, D., Schroff, F., Jou, B., Adam, H., & Prasad, G. (2021). Sixteen facial expressions occur in similar contexts worldwide. *Nature*, *589*(7841), 251–257. https://doi.org/10.1038/s41586-020-3037-7

Cowie, R., Douglas-Cowie, E., & Cox, C. (2005). Beyond emotion archetypes: Databases for emotion modelling using neural networks. *Neural Networks* 18, 371–388. https://doi.org/10.1016/j.neunet.2005.03.002.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, *18*(1), 32–80. https://doi.org/10.1109/79.911197

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

Darwin, C. (1872). *The expression of the emotions in man and animals.* John Murray. https://doi.org/10.1037/10001-000

Davison, A. K., Lansley, C., Costen, N., Tan, K., & Yap, M. H. (2018). SAMM: A Spontaneous Micro-Facial Movement Dataset. *IEEE Transactions on Affective Computing*, *9*(1), 116–129. https://doi.org/10.1109/TAFFC.2016.2573832

Dawel, A., Miller, E. J., Horsburgh, A., & Ford, P. (2021). A systematic survey of face stimuli used in psychological research 2000–2020. *Behavior Research Methods*, *54*(4), 1889–1901. https://doi.org/10.3758/s13428-021-01705-3

Dawel, A., Wright, L., Irons, J., Dumbleton, R., Palermo, R., O'Kearney, R., & McKone, E. (2017). Perceived emotion genuineness: Normative ratings for popular facial expression stimuli and the development of perceived-as-genuine and perceived-as-fake sets. *Behavior Research Methods*, *49*(4), 1539–1562. https://doi.org/10.3758/s13428-016-0813-2

De la Torre, F., & Cohn, J. F. (2011). "Facial Expression Analysis," in *Visual Analysis of Humans*, eds. T. B. Moeslund, A. Hilton, V. Krüger, and L. Sigal (London: Springer London), 377–409. https://doi.org/10.1007/978-0-85729-997-0_19.

Del Líbano, M., Calvo, M. G., Fernández-Martín, A., & Recio, G. (2018). Discrimination between smiling faces: Human observers vs. automated face analysis. *Acta Psychologica*, *187*, 19–29. https://doi.org/10.1016/j.actpsy.2018.04.019

Dente, P., Küster, D., Skora, L., & Krumhuber, E. (2017). Measures and metrics for automatic emotion classification via FACET. in *Proceedings of the Conference on the Study of Artificial Intelligence and Simulation of Behaviour (AISB)*, 160–163.

Devilly, G. J., & O'Donohue, R. P. (2021). A video is worth a thousand thoughts: Comparing a video mood induction procedure to an autobiographical recall technique. *Australian Journal of Psychology*, *73*(4), 438–451. https://doi.org/10.1080/00049530.2021.1997553

Dhall, A., Goecke, R., Lucey, S., & Gedeon, T. (2012). Collecting Large, Richly Annotated Facial-Expression Databases from Movies. *IEEE MultiMedia*, *19*(3), 34–41. https://doi.org/10.1109/MMUL.2012.26

Dhall, A., Singh, M., Goecke, R., Gedeon, T., Zeng, D., Wang, Y., & Ikeda, K. (2023). EmotiW 2023: Emotion Recognition in the Wild Challenge. *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION*, 746–749. https://doi.org/10.1145/3577190.3616545

Diconne, K., Kountouriotis, G. K., Paltoglou, A. E., Parker, A., & Hostler, T. J. (2022). Presenting KAPODI – The Searchable Database of Emotional Stimuli Sets. *Emotion Review*, *14*(1), 84–95. https://doi.org/10.1177/17540739211072803

Dobs, K., Bülthoff, 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.

Douglas-Cowie, E., Cox, C., Martin, J.-C., Devillers, L., Cowie, R., Sneddon, I., McRorie, M., Pelachaud, C., Peters, C., Lowry, O., Batliner, A., & Hönig, F. (2011). The HUMAINE Database. In R. Cowie, C. Pelachaud, & P. Petta (Eds.), *Emotion-Oriented Systems* (pp. 243–284). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-15184-2_14

Du, S., & Martinez, A. M. (2015). Compound facial expressions of emotion: From basic research to clinical applications. *Dialogues in Clinical Neuroscience*, *17*(4), 443–455. https://doi.org/10.31887/DCNS.2015.17.4/sdu

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

Dupré, D., Krumhuber, E. G., Küster, D., & McKeown, G. J. (2020). A performance comparison of eight commercially available automatic classifiers for facial affect recognition. *PLOS ONE*, *15*(4), e0231968. https://doi.org/10.1371/journal.pone.0231968

Dupré, D., Krumhuber, E., Küster, D., & McKeown, G. J. (2019). Emotion recognition in humans and machine using posed and spontaneous facial expression. PsyArXiv. https://doi.org/10.31234/osf.io/kzhds.

Durán, J. I., & Fernández-Dols, J.-M. (2021). Do emotions result in their predicted facial expressions? A meta-analysis of studies on the co-occurrence of expression and emotion. *Emotion*, *21*(7), 1550–1569. https://doi.org/10.1037/emo0001015

Dzhelyova, M., Perrett, D. I., & Jentzsch, I. (2012). Temporal dynamics of trustworthiness perception. *Brain Research*, *1435*, 81–90. https://doi.org/10.1016/j.brainres.2011.11.043

Edwards, K. (1998). The face of time: Temporal cues in facial expressions of emotion. *Psychological Science*, *9*(4), 270–276. https://doi.org/10.1111/1467-9280.00054

Ehrlich, S. M., Schiano, D. J., & Sheridan, K. (2000). Communicating facial affect: it's not the realism, it's the motion. in *CHI '00 Extended Abstracts on Human Factors in Computing Systems* (The Hague The Netherlands: ACM), 251–252. https://doi.org/10.1145/633292.633439.

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

Eisenkraft, N., & Elfenbein, H. A. (2010). The Way You Make Me Feel: Evidence for Individual Differences in Affective Presence. *Psychological Science*, *21*(4), 505–510. https://doi.org/10.1177/0956797610364117

Ekman, P. (1982). "Methods for measuring facial action," in *Handbook of methods in nonverbal behavior research.* (New York: Cambridge University Press), 45–135.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P. (2005). Basic Emotions. In T. Dalgleish & M. J. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 45–60). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470013494.ch3

Ekman, P. (2006). Darwin, Deception, and Facial Expression. *Annals of the New York Academy of Sciences*, *1000*(1), 205–221. https://doi.org/10.1196/annals.1280.010

Ekman, P., & Cordaro, D. (2011). What is Meant by Calling Emotions Basic. *Emotion Review*, *3*(4), 364–370. https://doi.org/10.1177/1754073911410740

Ekman, P., & Friesen, W. V. (1978). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129. https://doi.org/10.1037/h0030377

Hyunwoo Kim

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, *6*(4), 238–252. https://doi.org/10.1007/BF00987191

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face*. Malor Books.

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist*, *46*(9), 913–920. https://doi.org/10.1037/0003-066X.46.9.913

Ekman, P., & Rosenberg, E. L. (2005). *What the Face Reveals Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195179644.001.0001

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), 342–353. https://doi.org/10.1037/0022-3514.58.2.342

Ekman, P., Friesen, W. V. V., & Hager, J. C. (2002). *The Facial Action Coding System: A Technique for the Measurement of Facial Movement.* (San Francisco, CA: Consulting Psychologists Press)

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, *53*(4), 712–717. https://doi.org/10.1037/0022-3514.53.4.712

Ekman, P., Roper, G., & Hager, J. C. (1980). Deliberate Facial Movement. *Child Development*, *51*(3), 886. https://doi.org/10.2307/1129478

El Haj, M., Daoudi, M., Gallouj, K., Moustafa, A. A., & Nandrino, J.-L. (2018). When your face describes your memories: Facial expressions during retrieval of autobiographical memories. *Reviews in the Neurosciences*, *29*(8), 861–872. https://doi.org/10.1515/revneuro-2018-0001

Hyunwoo Kim

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Elfenbein, H. A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion*, *7*(1), 131–146. https://doi.org/10.1037/1528-3542.7.1.131

Ellsworth, P. C. (2013). Appraisal Theory: Old and New Questions. *Emotion Review*, *5*(2), 125–131. https://doi.org/10.1177/1754073912463617

Erdem, C., Turan, C., & Aydin, Z. (2015). BAUM-2: A multilingual audio-visual affective face database. *Multimedia Tools and Applications*, *74*(18), 7429–7459. https://doi.org/10.1007/s11042-014-1986-2

Erickson, K., & Schulkin, J. (2003). Facial expressions of emotion: A cognitive neuroscience perspective. *Brain and Cognition*, *52*(1), 52–60. https://doi.org/10.1016/S0278-2626(03)00008-3

Fabrício, D. D. M., Ferreira, B. L. C., Maximiano-Barreto, M. A., Muniz, M., & Chagas, M. H. N. (2022). Construction of face databases for tasks to recognize facial expressions of basic emotions: A systematic review. *Dementia & Neuropsychologia*, *16*(4), 388–410. https://doi.org/10.1590/1980-5764-dn-2022-0039

Facial Action Coding System. (2017). In M. Allen, *The SAGE Encyclopedia of Communication Research Methods*. SAGE Publications, Inc. https://doi.org/10.4135/9781483381411.n178

Fanelli, G., Gall, J., Romsdorfer, H., Weise, T., & Van Gool, L. (2010). A 3-D Audio-Visual Corpus of Affective Communication. *IEEE Transactions on Multimedia*, *12*(6), 591–598. https://doi.org/10.1109/TMM.2010.2052239

Hyunwoo Kim

Fiorentini, C., & Viviani, P. (2009). Perceiving facial expressions. *Visual Cognition*, *17*(3), 373–411. https://doi.org/10.1080/13506280701821019

Fiorentini, C., & Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *Journal of Vision*, *11*(3), 17–17. https://doi.org/10.1167/11.3.17

Fradera, A., & Ward, J. (2006). Placing events in time: The role of autobiographical recollection. *Memory*, *14*(7), 834–845. https://doi.org/10.1080/09658210600747241

Frank, M. G., Ekman, P., & Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology*, *64*(1), 83–93. https://doi.org/10.1037/0022-3514.64.1.83

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view.* Academic Press.

Frith, C. (2009). Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3453–3458. https://doi.org/10.1098/rstb.2009.0142

Gallagher, M. (2016). Sound as affect: Difference, power and spatiality. *Emotion, Space and Society*, *20*, 42–48. https://doi.org/10.1016/j.emospa.2016.02.004

Ganel, T., Goshen-Gottstein, Y., & Goodale, M. A. (2005). Interactions between the processing of gaze direction and facial expression. *Vision Research*, *45*(9), 1191–1200. https://doi.org/10.1016/j.visres.2004.06.025

Gaspar, A., Esteves, F., & Arriaga, P. (2014). On prototypical facial expressions versus variation in facial behavior: What have we learned on the "visibility" of emotions from measuring facial actions in humans and apes. In M. Pina & N. Gontier (Eds.), *The Evolution of Social Communication in Primates* (Vol. 1, pp. 101–126). Springer International Publishing. https://doi.org/10.1007/978-3-319-02669-5_6

Gepner, B., Deruelle, C., & Grynfeltt, S. (2001). Motion and emotion: A novel approach to the study of face processing by young autistic children. *Journal of Autism and Developmental Disorders* 31, 37–45. https://doi.org/10.1023/A:1005609629218.

Gerardo, P. C., & Menezes, P. (2019). Classification of FACS-Action Units with CNN Trained from Emotion Labelled Data Sets. *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 3766–3770. https://doi.org/10.1109/SMC.2019.8914238

Girard, J. M., Cohn, J. F., & De la Torre, F. (2015). Estimating smile intensity: A better way. *Pattern Recognition Letters,* 66, 13–21. https://doi.org/10.1016/j.patrec.2014.10.004

Girard, J. M., Cohn, J. F., Jeni, L. A., Sayette, M. A., & De La Torre, F. (2015). Spontaneous facial expression in unscripted social interactions can be measured automatically. *Behavior Research Methods*, *47*(4), 1136–1147. https://doi.org/10.3758/s13428-014-0536-1

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cognition & Emotion*, *22*(6), 1094–1118. https://doi.org/10.1080/02699930701626582

Gold, J. M., Barker, J. D., Barr, S., Bittner, J. L., Bromfield, W. D., Chu, N., Goode, R. A., Lee, D., Simmons, M., & Srinath, A. (2013). The efficiency of dynamic and static facial expression recognition. *Journal of Vision*, *13*(5), 23–23. https://doi.org/10.1167/13.5.23

Gosselin, P., Kirouac, G., & Doré, F. Y. (1995). Components and recognition of facial expression in the communication of emotion by actors. *Journal of Personality and Social Psychology*, *68*(1), 83–96. https://doi.org/10.1037/0022-3514.68.1.83

Grimm, M., Dastidar, D., & Kroschel, K. (2006). Recognizing emotions in spontaneous facial expressions. In *Proceedings: International Conference on Intelligent Systems and Computing (ISYC)*.

Gross, J. J. (2015). Emotion Regulation: Current Status and Future Prospects. *Psychological Inquiry*, *26*(1), 1–26. https://doi.org/10.1080/1047840X.2014.940781

Gross, J. J., & John, O. P. (1997). Revealing feelings: Facets of emotional expressivity in self-reports, peer ratings, and behavior. *Journal of Personality and Social Psychology*, *72*(2), 435–448. https://doi.org/10.1037/0022-3514.72.2.435

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition & Emotion*, *9*(1), 87–108. https://doi.org/10.1080/02699939508408966

Guerdelli, H., Ferrari, C., Barhoumi, W., Ghazouani, H., & Berretti, S. (2022). Macro- and Micro-Expressions Facial Datasets: A Survey. *Sensors*, *22*(4), 1524. https://doi.org/10.3390/s22041524

Gunes, H., & Pantic, M. (2010). Automatic, Dimensional and Continuous Emotion Recognition: *International Journal of Synthetic Emotions*, *1*(1), 68–99. https://doi.org/10.4018/jse.2010101605

Gunnery, S. D., & Ruben, M. A. (2016). Perceptions of Duchenne and non-Duchenne smiles: A meta-analysis. *Cognition and Emotion*, *30*(3), 501–515. https://doi.org/10.1080/02699931.2015.1018817

Gunnery, S. D., Hall, J. A., & Ruben, M. A. (2013). The Deliberate Duchenne Smile: Individual Differences in Expressive Control. *Journal of Nonverbal Behavior*, *37*(1), 29–41. https://doi.org/10.1007/s10919-012-0139-4

Guo, H., Zhang, X.-H., Liang, J., & Yan, W.-J. (2018). The dynamic features of lip corners in genuine and posed smiles. *Frontiers in Psychology*, *9*, 202. https://doi.org/10.3389/fpsyg.2018.00202

Gupta, A., D'Cunha, A., Awasthi, K., & Balasubramanian, V. (2016). *DAiSEE: Towards User Engagement Recognition in the Wild*. https://doi.org/10.48550/ARXIV.1609.01885

Haamer, R. E., Kulkarni, K., Imanpour, N., Haque, M. A., Avots, E., Breisch, M., Nasrollahi, K., Escalera, S., Ozcinar, C., Baro, X., Naghsh-Nilchi, A. R., Moeslund, T. B., & Anbarjafari, G. (2018). Changes in Facial Expression as Biometric: A Database and Benchmarks of Identification. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 621–628. https://doi.org/10.1109/FG.2018.00098

Haidt, J., & Keltner, D. (1999). Culture and Facial Expression: Open-ended Methods Find More Expressions and a Gradient of Recognition. *Cognition & Emotion*, *13*(3), 225–266. https://doi.org/10.1080/026999399379267

Halberstadt, J., Winkielman, P., Niedenthal, P. M., & Dalle, N. (2009). Emotional Conception: How Embodied Emotion Concepts Guide Perception and Facial Action. *Psychol Sci* 20, 1254–1261. https://doi.org/10.1111/j.1467-9280.2009.02432.x.

Hamann, S., & Canli, T. (2004). Individual differences in emotion processing. *Current Opinion in Neurobiology*, *14*(2), 233–238. https://doi.org/10.1016/j.conb.2004.03.010

Han, B., Yun, W.-H., Yoo, J.-H., & Kim, W. H. (2020). Toward Unbiased Facial Expression Recognition in the Wild via Cross-Dataset Adaptation. *IEEE Access*, *8*, 159172–159181. https://doi.org/10.1109/ACCESS.2020.3018738

Han, X., Ji, Z., & Wang, W. (2020). Low Resolution Facial Manipulation Detection. *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 431–434. https://doi.org/10.1109/VCIP49819.2020.9301796

Happy, S. L., Patnaik, P., Routray, A., & Guha, R. (2015). *The Indian Spontaneous Expression Database for Emotion Recognition*. https://doi.org/10.48550/ARXIV.1512.00932

Harwood, N. K., Hall, L. J., & Shinkfield, A. J. (1999). Recognition of facial emotional expressions from moving and static displays by individuals with mental retardation.

American Journal on Mental Retardation, *104*(3), 270. https://doi.org/10.1352/0895-8017(1999)104<0270:ROFEEF>2.0.CO;2

Hassin, R. R., Aviezer, H., & Bentin, S. (2013). Inherently Ambiguous: Facial Expressions of Emotions, in Context. *Emotion Review*, *5*(1), 60–65. https://doi.org/10.1177/1754073912451331

He, G., Liu, X., Fan, F., & You, J. (2020). Image2Audio: Facilitating Semi-supervised Audio Emotion Recognition with Facial Expression Image. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 3978–3983. https://doi.org/10.1109/CVPRW50498.2020.00464

Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology*, *40*(2), 129–141. https://doi.org/10.1016/S0167-8760(00)00161-6

Hess, U., & Hareli, S. (2015). The role of social context for the interpretation of emotional facial expressions. In M. K. Mandal & A. Awasthi (Eds.), *Understanding Facial Expressions in Communication* (pp. 119–141). Springer India. https://doi.org/10.1007/978-81-322-1934-7_7

Hess, U., & Kleck, R. E. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology*, *20*(5), 369–385. https://doi.org/10.1002/ejsp.2420200502

Hess, U., Blairy, S., & Kleck, R. E. (1997). The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior*, *21*(4), 241–257. https://doi.org/10.1023/A:1024952730333

Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta Psychologica*, *135*(3), 278–283. https://doi.org/10.1016/j.actpsy.2010.07.012

Höfling, T. T. A., Alpers, G. W., Gerdes, A. B. M., & Föhl, U. (2021). Automatic facial coding versus electromyography of mimicked, passive, and inhibited facial response to emotional faces. *Cognition and Emotion*, *35*(5), 874–889. https://doi.org/10.1080/02699931.2021.1902786

Horstmann, G. (2003). What do facial expressions convey: Feeling states, behavioral intentions, or actions requests? *Emotion*, *3*(2), 150–166. https://doi.org/10.1037/1528-3542.3.2.150

Horvat, M., Kukolja, D., & Ivanec, D. (2015). *Comparing affective responses to standardized pictures and videos: A study report*. https://doi.org/10.48550/ARXIV.1505.07398

Houstis, O., & Kiliaridis, S. (2009). Gender and age differences in facial expressions. *The European Journal of Orthodontics*, *31*(5), 459–466. https://doi.org/10.1093/ejo/cjp019

Hsu, C.-T., Sato, W., & Yoshikawa, S. (2020). Enhanced emotional and motor responses to live versus videotaped dynamic facial expressions. *Scientific Reports*, *10*(1), 16825. https://doi.org/10.1038/s41598-020-73826-2

Ichikawa, H., Kanazawa, S., & Yamaguchi, M. K. (2014). Infants Recognize the Subtle Happiness Expression. *Perception*, *43*(4), 235–248. https://doi.org/10.1068/p7595

Iliyasu, R., & Etikan, I. (2021). Comparison of quota sampling and stratified random sampling. *Biometrics & Biostatistics International Journal*, *10*(1), 24–27. https://doi.org/10.15406/bbij.2021.10.00326

Ito, T., Yokokawa, K., Yahata, N., Isato, A., Suhara, T., & Yamada, M. (2017). Neural basis of negativity bias in the perception of ambiguous facial expression. *Scientific Reports*, *7*(1), 420. https://doi.org/10.1038/s41598-017-00502-3

Izard, C. E. (1971). *The face of emotion*. Appleton-Century-Crofts.

Jabbi, M., & Keysers, C. (2008). Inferior frontal gyrus activity triggers anterior insula response to emotional facial expressions. *Emotion*, *8*(6), 775–780. https://doi.org/10.1037/a0014194

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology*, *24*(2), 187–192. https://doi.org/10.1016/j.cub.2013.11.064

Jerritta, S., Murugappan, M., Nagarajan, R., & Wan, K. (2011). Physiological signals based human emotion Recognition: A review. *2011 IEEE 7th International Colloquium on Signal Processing and Its Applications*, 410–415. https://doi.org/10.1109/CSPA.2011.5759912

Jia, S., Wang, S., Hu, C., Webster, P. J., & Li, X. (2021). Detection of Genuine and Posed Facial Expressions of Emotion: Databases and Methods. *Frontiers in Psychology*, *11*, 580287. https://doi.org/10.3389/fpsyg.2020.580287

Jian-zheng, L., Zheng, Z., Man-tian, L., & Chang, L. (2011). Action unit recognition based on motion templates and GentleBoost. *Proceedings - 7th International Conference on Networked Computing and Advanced Information Management, NCM 2011*.

Jokinen, K., & Wilcock, G. (2012). Multimodal Signals and Holistic Interaction Structuring. *Proceedings of 24th International Conference on Computational Linguistics (COLING 2012)*.

Jones, A. C., Gutierrez, R., & Ludlow, A. K. (2018). The role of motion and intensity in deaf children's recognition of real human facial expressions of emotion. *Cognition and Emotion* 32, 102–115. https://doi.org/10.1080/02699931.2017.1289894.

Jürgens, R., Grass, A., Drolet, M., & Fischer, J. (2015). Effect of Acting Experience on Emotion Expression and Recognition in Voice: Non-Actors Provide Better Stimuli than

Expected. *Journal of Nonverbal Behavior*, *39*(3), 195–214. https://doi.org/10.1007/s10919-015-0209-5

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

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

Kanade, T., Cohn, J. F., & Yingli Tian. (2000). Comprehensive database for facial expression analysis. *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, 46–53. https://doi.org/10.1109/AFGR.2000.840611

Kappas, A. (2013). Social regulation of emotion: Messy layers. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00051

Kätsyri, J., & Sams, M. (2008). The effect of dynamics on identifying basic emotions from synthetic and natural faces. *International Journal of Human-Computer Studies*, *66*(4), 233–242. https://doi.org/10.1016/j.ijhcs.2007.10.001

Kayyal, M. H., & Russell, J. A. (2013). Americans and Palestinians judge spontaneous facial expressions of emotion. *Emotion*, *13*(5), 891–904. https://doi.org/10.1037/a0033244

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology*, *68*(3), 441–454. https://doi.org/10.1037/0022-3514.68.3.441

Keltner, D., & Kring, A. M. (1998). Emotion, Social Function, and Psychopathology. *Review of General Psychology*, *2*(3), 320–342. https://doi.org/10.1037/1089-2680.2.3.320

Hyunwoo Kim

Keltner, D., Tracy, J. L., Sauter, D., & Cowen, A. (2019). What Basic Emotion Theory Really Says for the Twenty-First Century Study of Emotion. *Journal of Nonverbal Behavior*, *43*(2), 195–201. https://doi.org/10.1007/s10919-019-00298-y

Khan, R. A., Crenn, A., Meyer, A., & Bouakaz, S. (2019). A novel database of children's spontaneous facial expressions (LIRIS-CSE). *Image and Vision Computing*, *83–84*, 61–69. https://doi.org/10.1016/j.imavis.2019.02.004

Khan, S. A. (2017). Factors affecting the recognition accuracy of facial expressions. *MOJ Applied Bionics and Biomechanics*, *1*(4). https://doi.org/10.15406/mojabb.2017.01.00021

Kim, H. N., & Sutharson, S. J. (2023). Individual differences in spontaneous facial expressions in people with visual impairment and blindness. *British Journal of Visual Impairment*, *41*(3), 475–488. https://doi.org/10.1177/02646196211070927

Kim, H., Küster, D., Girard, J. M., & Krumhuber, E. G. (2023). Human and machine recognition of dynamic and static facial expressions: Prototypicality, ambiguity, and complexity. *Frontiers in Psychology*, *14*, 1221081. https://doi.org/10.3389/fpsyg.2023.1221081

Kim, J., & Andre, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(12), 2067–2083. https://doi.org/10.1109/TPAMI.2008.26

Kinchella, J., & Guo, K. (2021). Facial Expression Ambiguity and Face Image Quality Affect Differently on Expression Interpretation Bias. *Perception*, *50*(4), 328–342. https://doi.org/10.1177/03010066211000270

Knight, B., & Johnston, A. (1997). The Role of Movement in Face Recognition. *Visual Cognition* 4, 265–273. https://doi.org/10.1080/713756764.

Hyunwoo Kim

Ko, H., Kim, K., Bae, M., Seo, M.-G., Nam, G., Park, S., Park, S., Ihm, J., & Lee, J.-Y. (2021). Changes in Facial Recognition and Facial Expressions with Age. *Sensors*, *21*, 4858. https://doi.org/10.20944/preprints202104.0542.v1

Kocsor, F., Kozma, L., Neria, A. L., Jones, D. N., & Bereczkei, T. (2019). Arbitrary signals of trustworthiness – social judgments may rely on facial expressions even with experimentally manipulated valence. *Heliyon*, *5*(5), e01736. https://doi.org/10.1016/j.heliyon.2019.e01736

Koelstra, S., Pantic, M., & Patras, I. (2010). A Dynamic Texture-Based Approach to Recognition of Facial Actions and Their Temporal Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(11), 1940–1954. https://doi.org/10.1109/TPAMI.2010.50

Komori, M., & Onishi, Y. (2015). Analysis of Dynamic Characteristics of Spontaneous Facial Expressions. In *EAPCogSci*.

Koringa, P. A., Mitra, S. K., & Asari, V. K. (2017). Handling Illumination Variation: A Challenge for Face Recognition. In B. Raman, S. Kumar, P. P. Roy, & D. Sen (Eds.), *Proceedings of International Conference on Computer Vision and Image Processing* (Vol. 460, pp. 273–283). Springer Singapore. https://doi.org/10.1007/978-981-10-2107-7_25

Kossaifi, J., Walecki, R., Panagakis, Y., Shen, J., Schmitt, M., Ringeval, F., Han, J., Pandit, V., Toisoul, A., Schuller, B., Star, K., Hajiyev, E., & Pantic, M. (2021). SEWA DB: A Rich Database for Audio-Visual Emotion and Sentiment Research in the Wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(3), 1022–1040. https://doi.org/10.1109/TPAMI.2019.2944808

Koval, P., & Kuppens, P. (2012). Changing emotion dynamics: Individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion*, *12*(2), 256–267. https://doi.org/10.1037/a0024756

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Kret, M. E., & de Gelder, B. (2010). Social context influences recognition of bodily expressions. *Experimental Brain Research*, *203*(1), 169–180. https://doi.org/10.1007/s00221-010-2220-8

Kroczek, L. O. H., Lingnau, A., Schwind, V., Wolff, C., & Mühlberger, A. (2022). Observers predict actions from facial emotional expressions during real-time social interactions. *PsyArXiv*. https://doi.org/10.31234/osf.io/fzcau

Krumhuber, E. G., & Kappas, A. (2022). More What Duchenne Smiles Do, Less What They Express. *Perspectives on Psychological Science*, *17*(6), 1566–1575. https://doi.org/10.1177/17456916211071083

Krumhuber, E. G., & Skora, L. (2016). "Perceptual Study on Facial Expressions," in *Handbook of Human Motion*, eds. B. Müller, S. I. Wolf, G.-P. Brueggemann, Z. Deng, A. McIntosh, F. Miller, et al. (Cham: Springer International Publishing), 1–15. https://doi.org/10.1007/978-3-319-30808-1_18-1.

Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of Dynamic Aspects of Facial Expressions: A Review. *Emotion Review*, *5*(1), 41–46. https://doi.org/10.1177/1754073912451349

Krumhuber, E. G., Küster, D., Namba, S., & Skora, L. (2021a). Human and machine validation of 14 databases of dynamic facial expressions. *Behavior Research Methods*, *53*(2), 686–701. https://doi.org/10.3758/s13428-020-01443-y

Krumhuber, E. G., Küster, D., Namba, S., Shah, D., & Calvo, M. G. (2021b). Emotion recognition from posed and spontaneous dynamic expressions: Human observers versus machine analysis. *Emotion*, *21*(2), 447–451. https://doi.org/10.1037/emo0000712

Krumhuber, E. G., Skora, L. I., Hill, H. C. H., & Lander, K. (2023). The role of facial movements in emotion recognition. *Nature Reviews Psychology*, *2*(5), 283–296. https://doi.org/10.1038/s44159-023-00172-1

Krumhuber, E. G., Skora, L., Küster, D., & Fou, L. (2017). A Review of Dynamic Datasets for Facial Expression Research. *Emotion Review*, *9*(3), 280–292. https://doi.org/10.1177/1754073916670022

Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, *7*(4), 730–735. https://doi.org/10.1037/1528-3542.7.4.730

Kulke, L., Feyerabend, D., & Schacht, A. (2020). A Comparison of the Affectiva iMotions facial expression analysis software with EMG for identifying facial expressions of emotion. *Frontiers in Psychology*, *11*, 329. https://doi.org/10.3389/fpsyg.2020.00329

Küntzler, T., Höfling, T. T. A., & Alpers, G. W. (2021). Automatic Facial Expression Recognition in Standardized and Non-standardized Emotional Expressions. *Frontiers in Psychology*, *12*, 627561. https://doi.org/10.3389/fpsyg.2021.627561

Kunz, M., Mylius, V., Schepelmann, K., & Lautenbacher, S. (2008). Impact of age on the facial expression of pain. *Journal of Psychosomatic Research*, *64*(3), 311–318. https://doi.org/10.1016/j.jpsychores.2007.09.010

Küster, D., Baker, M., & Krumhuber, E. G. (2021). PDSTD - The Portsmouth Dynamic Spontaneous Tears Database. *Behav Res* 54, 2678–2692. https://doi.org/10.3758/s13428-021-01752-w.

Küster, D., Krumhuber, E. G., Steinert, L., Ahuja, A., Baker, M., & Schultz, T. (2020). Opportunities and Challenges for Using Automatic Human Affect Analysis in Consumer Research. *Frontiers in Neuroscience*, *14*, 400. https://doi.org/10.3389/fnins.2020.00400

Lampropoulos, A. S., Stathopoulou, I.-O., & Tsihrintzis, G. A. (2009). Comparative performance evaluation of classifiers for Facial Expression Recognition. In E. Damiani, J. Jeong, R. J. Howlett, & L. C. Jain (Eds.), *New Directions in Intelligent Interactive Multimedia Systems and Services—2* (Vol. 226, pp. 253–263). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-02937-0_23

Lander, K., & Butcher, N. L. (2020). Recognizing genuine from posed facial expressions: exploring the role of dynamic information and face familiarity. *Frontiers in Psychology*, *11*, 1378. https://doi.org/10.3389/fpsyg.2020.01378

Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Mem Cogn* 27, 974–985. https://doi.org/10.3758/BF03201228.

Lang, P. J., Bradley, M. M., Cuthbert, B. N., & Center for the Study of Emotion and Attention. (2020). *International Affective Picture System* [Data set]. https://doi.org/10.1037/t66667-000

Lazarus, R. S. (1991). Cognition and motivation in emotion. *American Psychologist*, *46*(4), 352–367. https://doi.org/10.1037/0003-066X.46.4.352

Le Ngo, A. C., Oh, Y.-H., Phan, R. C.-W., & See, J. (2016). Eulerian emotion magnification for subtle expression recognition. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1243–1247. https://doi.org/10.1109/ICASSP.2016.7471875

Lederman, S. J., Klatzky, R. L., Abramowicz, A., Salsman, K., Kitada, R., & Hamilton, C. (2007). Haptic Recognition of Static and Dynamic Expressions of Emotion in the Live

Face. *Psychological Science*, *18*(2), 158–164. https://doi.org/10.1111/j.1467-9280.2007.01866.x

Leonard, C. M., Voeller, K. K. S., & Kuldau, J. M. (1991). When's a Smile a Smile? Or how to Detect a Message by Digitizing the Signal. *Psychological Science*, *2*(3), 166–172. https://doi.org/10.1111/j.1467-9280.1991.tb00125.x

Leppänen, J. M., & Hietanen, J. K. (2004). Positive facial expressions are recognized faster than negative facial expressions, but why? *Psychological Research Psychologische Forschung*, *69*(1–2), 22–29. https://doi.org/10.1007/s00426-003-0157-2

Levenson, R. W. (2007). Emotion elicitation with neurological patients. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 158–168). Oxford University Press.

Lewinski, P., Den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, *7*(4), 227–236. https://doi.org/10.1037/npe0000028

Li, J., Dong, Z., Lu, S., Wang, S.-J., Yan, W.-J., Ma, Y., Liu, Y., Huang, C., & Fu, X. (2022). CAS(ME)^3: A Third Generation Facial Spontaneous Micro-Expression Database with Depth Information and High Ecological Validity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. https://doi.org/10.1109/TPAMI.2022.3174895

Li, W., Cui, Y., Ma, Y., Chen, X., Li, G., Zeng, G., Guo, G., & Cao, D. (2023). A Spontaneous Driver Emotion Facial Expression (DEFE) Dataset for Intelligent Vehicles: Emotions Triggered by Video-Audio Clips in Driving Scenarios. *IEEE Transactions on Affective Computing*, *14*(1), 747–760. https://doi.org/10.1109/TAFFC.2021.3063387

Li, W., Tan, R., Xing, Y., Li, G., Li, S., Zeng, G., Wang, P., Zhang, B., Su, X., Pi, D., Guo, G., & Cao, D. (2022). A multimodal psychological, physiological and behavioural dataset for human emotions in driving tasks. *Scientific Data*, *9*(1), 481. https://doi.org/10.1038/s41597-022-01557-2

Li, X., Cheng, S., Li, Y., Behzad, M., Shen, J., Zafeiriou, S., Pantic, M., & Zhao, G. (2022). 4DME: A Spontaneous 4D Micro-Expression Dataset With Multimodalities. *IEEE Transactions on Affective Computing*, 1–18. https://doi.org/10.1109/TAFFC.2022.3182342

Li, X., Pfister, T., Huang, X., Zhao, G., & Pietikainen, M. (2013). A Spontaneous Micro-expression Database: Inducement, collection and baseline. *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 1–6. https://doi.org/10.1109/FG.2013.6553717

Li, X., Zhang, X., Yang, H., Duan, W., Dai, W., & Yin, L. (2020). An EEG-Based Multi-Modal Emotion Database with Both Posed and Authentic Facial Actions for Emotion Analysis. *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, 336–343. https://doi.org/10.1109/FG47880.2020.00050

Limbrecht-Ecklundt, K., Scheck, A., Jerg-Bretzke, L., Walter, S., Hoffmann, H., & Traue, H. C. (2013). The effect of forced choice on facial emotion recognition: A comparison to open verbal classification of emotion labels. *GMS Psycho-Social-Medicine*. https://doi.org/10.3205/PSM000094

Littlewort, G. C., Bartlett, M. S., & Lee, K. (2007). Faces of pain: Automated measurement of spontaneous all facial expressions of genuine and posed pain. *Proceedings of the 9th International Conference on Multimodal Interfaces*, 15–21. https://doi.org/10.1145/1322192.1322198

Littlewort, G. C., Bartlett, M. S., Salamanca, L. P., & Reilly, J. (2011). Automated measurement of children's facial expressions during problem solving tasks. *Face and Gesture 2011*, 30–35. https://doi.org/10.1109/FG.2011.5771418

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011). The computer expression recognition toolbox (CERT). *Face and Gesture 2011*, 298–305. https://doi.org/10.1109/FG.2011.5771414

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*, 94–101. https://doi.org/10.1109/CVPRW.2010.5543262

Lundqvist, D., Flykt, A., & Öhman, A. (2015). *Karolinska Directed Emotional Faces* [Dataset]. https://doi.org/10.1037/t27732-000

Mahmoud, M., Baltrušaitis, T., Robinson, P., & Riek, L. D. (2011). 3D Corpus of Spontaneous Complex Mental States. In S. D'Mello, A. Graesser, B. Schuller, & J.-C. Martin (Eds.), *Affective Computing and Intelligent Interaction* (Vol. 6974, pp. 205–214). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24600-5_24

Malawski, F., Kwolek, B., & Sako, S. (2014). Using Kinect for Facial Expression Recognition under Varying Poses and Illumination. In D. Ślęzak, G. Schaefer, S. T. Vuong, & Y.-S. Kim (Eds.), *Active Media Technology* (Vol. 8610, pp. 395–406). Springer International Publishing. https://doi.org/10.1007/978-3-319-09912-5_33

Mandal, M., Poddar, S., & Das, A. (2015). Comparison of human and machine based facial expression classification. in *International Conference on Computing, Communication*

& Automation (Greater Noida, India: IEEE), 1198–1203. https://doi.org/10.1109/CCAA.2015.7148558.

Marian, D. E., & Shimamura, A. P. (2013). Contextual influences on dynamic facial expressions. *The American Journal of Psychology*, *126*(1), 53–66. https://doi.org/10.5406/amerjpsyc.126.1.0053

Martinez, B., Valstar, M. F., Jiang, B., & Pantic, M. (2019). Automatic Analysis of Facial Actions: A Survey. *IEEE Transactions on Affective Computing*, *10*(3), 325–347. https://doi.org/10.1109/TAFFC.2017.2731763

Matsumoto, D. (1999). American-Japanese Cultural Differences in Judgements of Expression Intensity and Subjective Experience. *Cognition and Emotion* 13, 201–218. https://doi.org/10.1080/026999399379339.

Matsumoto, D., & Ekman, P. (2010). Subjective Experience and the Expression of Emotion in Man. In *Encyclopedia of Behavioral Neuroscience* (pp. 342–348). Elsevier. https://doi.org/10.1016/B978-0-08-045396-5.00097-X

Matsumoto, D., & Hwang, H. C. (2014). Judgments of subtle facial expressions of emotion. *Emotion*, *14*(2), 349–357. https://doi.org/10.1037/a0035237

Matsumoto, D., & Hwang, H. C. (2017). Methodological Issues Regarding Cross-Cultural Studies of Judgments of Facial Expressions. *Emotion Review*, *9*(4), 375–382. https://doi.org/10.1177/1754073916679008

Matsumoto, D., & Hwang, H. S. (2011). Judgments of facial expressions of emotion in profile. *Emotion*, *11*(5), 1223–1229. https://doi.org/10.1037/a0024356

Matsumoto, D., Consolacion, T., Yamada, H., Suzuki, R., Franklin, B., Paul, S., Ray, R., & Uchida, H. (2002). American-Japanese cultural differences in judgements of emotional expressions of different intensities. *Cognition and Emotion*, *16*(6), 721–747. https://doi.org/10.1080/02699930143000608

Matsumoto, D., Olide, A., Schug, J., Willingham, B., & Callan, M. (2009). Cross-cultural judgments of spontaneous facial expressions of emotion. *Journal of Nonverbal Behavior*, *33*(4), 213–238. https://doi.org/10.1007/s10919-009-0071-4

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), 151–160. https://doi.org/10.1109/T-AFFC.2013.4

Mayer, C., Eggers, M., & Radig, B. (2014). Cross-database evaluation for facial expression recognition. *Pattern Recognition and Image Analysis*, *24*(1), 124–132. https://doi.org/10.1134/S1054661814010106

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*(3), 424–453. https://doi.org/10.1037/0033-2909.126.3.424

McDuff, D., Amr, M., & Kaliouby, R. E. (2019). AM-FED+: An Extended Dataset of Naturalistic Facial Expressions Collected in Everyday Settings. *IEEE Transactions on Affective Computing*, *10*(1), 7–17. https://doi.org/10.1109/TAFFC.2018.2801311

McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J., & Kaliouby, R. E. (2016). AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 3723–3726. https://doi.org/10.1145/2851581.2890247

McLellan, T., Johnston, L., Dalrymple-Alford, J., & Porter, R. (2010). Sensitivity to genuine versus posed emotion specified in facial displays. *Cognition & Emotion*, *24*(8), 1277–1292. https://doi.org/10.1080/02699930903306181

Meeten, F., Ivak, P., Dash, S. R., Knowles, S., Duka, T., Scott, R., Kaiser, J., & Davey, G. C. L. (2015). The Effect of Facial Expressions on the Evaluation of Ambiguous Statements.

Journal of Experimental Psychopathology, *6*(3), 253–263. https://doi.org/10.5127/jep.039613

Mehu, M., & Scherer, K. R. (2015). Emotion categories and dimensions in the facial communication of affect: An integrated approach. *Emotion*, *15*(6), 798–811. https://doi.org/10.1037/a0039416

Mehu, M., Little, A. C., & Dunbar, R. I. M. (2007). Duchenne smiles and the perception of generosity and sociability in faces. *Journal of Evolutionary Psychology*, *5*(1), 183–196. https://doi.org/10.1556/JEP.2007.1011

Miehlke, A., Fisch, U., Eneroth, C.-M., & Miehlke, A. (1973). *Surgery of the facial nerve* (2. enl. ed. of "Die Chirurgie des Nervus facialis"). Urban & Schwarzenberg.

Milders, M., Sahraie, A., & Logan, S. (2008). Minimum presentation time for masked facial expression discrimination. *Cognition and Emotion*, *22*(1), 63–82. https://doi.org/10.1080/02699930701273849

Minami, T., Nakajima, K., & Nakauchi, S. (2018). Effects of Face and Background Color on Facial Expression Perception. *Frontiers in Psychology*, *9*, 1012. https://doi.org/10.3389/fpsyg.2018.01012

Miolla, A., Cardaioli, M., & Scarpazza, C. (2022). Padova Emotional Dataset of Facial Expressions (PEDFE): A unique dataset of genuine and posed emotional facial expressions. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01914-4

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & for the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, *339*(1), b2535–b2535. https://doi.org/10.1136/bmj.b2535

Morecraft, R. J., Louie, J. L., Herrick, J. L., & Stilwell-Morecraft, K. S. (2001). Cortical innervation of the facial nucleus in the non-human primate. *Brain*, *124*(1), 176–208. https://doi.org/10.1093/brain/124.1.176

Morishima, S., Yotsukura, T., Yamada, H., Uchida, H., Tetsutani, N., & Akamatsu, S. (2001). Dynamic micro aspects of facial movements in elicited and posed expressions using high-speed camera. *Proceedings 10th IEEE International Workshop on Robot and Human Interactive Communication. ROMAN 2001 (Cat. No.01TH8591)*, 371–376. https://doi.org/10.1109/ROMAN.2001.981932

Motley, M. T., & Camden, C. T. (1988). Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Speech Communication*, *52*(1), 1–22. https://doi.org/10.1080/10570318809389622

Naab, P. J., & Russell, J. A. (2007). Judgments of emotion from spontaneous facial expressions of New Guineans. *Emotion*, *7*(4), 736–744. https://doi.org/10.1037/1528-3542.7.4.736

Namba, S., Kabir, R. S., Miyatani, M., & Nakao, T. (2018). Dynamic Displays Enhance the Ability to Discriminate Genuine and Posed Facial Expressions of Emotion. *Frontiers in Psychology*, *9*, 672. https://doi.org/10.3389/fpsyg.2018.00672

Namba, S., Kagamihara, T., Miyatani, M., & Nakao, T. (2017). Spontaneous Facial Expressions Reveal New Action Units for the Sad Experiences. *Journal of Nonverbal Behavior*, *41*(3), 203–220. https://doi.org/10.1007/s10919-017-0251-6

Namba, S., Makihara, S., Kabir, R. S., Miyatani, M., & Nakao, T. (2017). Spontaneous Facial Expressions Are Different from Posed Facial Expressions: Morphological Properties and Dynamic Sequences. *Current Psychology*, *36*(3), 593–605. https://doi.org/10.1007/s12144-016-9448-9

Namba, S., Sato, W., Osumi, M., & Shimokawa, K. (2021). Assessing automated facial action unit detection systems for analyzing cross-domain facial expression databases. *Sensors*, *21*(12), 4222. https://doi.org/10.3390/s21124222

Nandrino, J. L., Gandolphe, M. C., Saloppe, X., Daoudi, M., Moustafa, A. A., & El Haj, M. (2019). The face of memory: Experiential avoidance and facial expressions during the retrieval of autobiographical memories. *Journal of Cognitive Psychology*, *31*(5–6), 533–542. https://doi.org/10.1080/20445911.2019.1637879

Nelson, N. L., & Russell, J. A. (2013). Universality Revisited. *Emotion Review*, *5*(1), 8–15. https://doi.org/10.1177/1754073912457227

Neta, M., & Whalen, P. J. (2010). The Primacy of Negative Interpretations When Resolving the Valence of Ambiguous Facial Expressions. *Psychological Science*, *21*(7), 901–907. https://doi.org/10.1177/0956797610373934

Ngo, A. C., Liong, S.-T., See, J., & Phan, R. C.-W. (2015). Are subtle expressions too sparse to recognize? *2015 IEEE International Conference on Digital Signal Processing (DSP)*, 1246–1250. https://doi.org/10.1109/ICDSP.2015.7252080

Nguyen, H., Kotani, K., Chen, F., & Le, B. (2014). A Thermal Facial Emotion Database and Its Analysis. In R. Klette, M. Rivera, & S. Satoh (Eds.), *Image and Video Technology* (Vol. 8333, pp. 397–408). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-53842-1_34

Nishiyama, M., Kawashima, H., Hirayama, T., & Matsuyama, T. (2005). Facial Expression Representation Based on Timing Structures in Faces. In W. Zhao, S. Gong, & X. Tang (Eds.), *Analysis and Modelling of Faces and Gestures* (Vol. 3723, pp. 140–154). Springer Berlin Heidelberg. https://doi.org/10.1007/11564386_12

Nummenmaa, L., & Calvo, M. G. (2015). Dissociation between recognition and detection advantage for facial expressions: A meta-analysis. *Emotion*, *15*(2), 243–256. https://doi.org/10.1037/emo0000042

Nummenmaa, T. (1988). The recognition of pure and blended facial expressions of emotion from still photographs. *Scand J Psychol* 29, 33–47. https://doi.org/10.1111/j.1467-9450.1988.tb00773.x.

O'Toole, A. J., An, X., Dunlop, J., Natu, V., & Phillips, P. J. (2012). Comparing face recognition algorithms to humans on challenging tasks. *ACM Transactions on Applied Perception*, *9*(4), 1–13. https://doi.org/10.1145/2355598.2355599

Onal Ertugrul, I., Ahn, Y. A., Bilalpur, M., Messinger, D. S., Speltz, M. L., & Cohn, J. F. (2022). Infant AFAR: Automated facial action recognition in infants. *Behavior Research Methods*, *55*(3), 1024–1035. https://doi.org/10.3758/s13428-022-01863-y

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Palermo, R., & Coltheart, M. (2004). Photographs of facial expression: Accuracy, response times, and ratings of intensity. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 634–638. https://doi.org/10.3758/BF03206544

Pampouchidou, A., Pediaditis, M., Kazantzaki, E., Sfakianakis, S., Apostolaki, I. A., Argyraki, K., Manousos, D., Meriaudeau, F., Marias, K., Yang, F., Tsiknakis, M., Basta, M., Vgontzas, A. N., & Simos, P. (2020). Automated facial video-based recognition of depression and anxiety symptom severity: Cross-corpus validation. *Machine Vision and Applications*, *31*(4), 30. https://doi.org/10.1007/s00138-020-01080-7

Pantic, M. (2009). Machine analysis of facial behaviour: Naturalistic and dynamic behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3505–3513. https://doi.org/10.1098/rstb.2009.0135

Pantic, M., & Bartlett, M. S. (2007). *Machine analysis of facial expressions* (Vol. 558). INTECH Open Access Publisher

Pantic, M., & Patras, I. (2006). Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, *36*(2), 433–449. https://doi.org/10.1109/TSMCB.2005.859075

Pantic, M., & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(12), 1424–1445. https://doi.org/10.1109/34.895976

Pantic, M., & Rothkrantz, L. J. M. (2004). Facial Action Recognition for Facial Expression Analysis From Static Face Images. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, *34*(3), 1449–1461. https://doi.org/10.1109/TSMCB.2004.825931

Park, S., Lee, K., Lim, J.-A., Ko, H., Kim, T., Lee, J.-I., Kim, H., Han, S.-J., Kim, J.-S., Park, S., Lee, J.-Y., & Lee, E. C. (2020). Differences in facial expressions between spontaneous and posed smiles: Automated method by action units and three-dimensional facial landmarks. *Sensors*, *20*(4), 1199. https://doi.org/10.3390/s20041199

Parkinson, B. (2005). Do Facial Movements Express Emotions or Communicate Motives? *Personality and Social Psychology Review*, *9*(4), 278–311. https://doi.org/10.1207/s15327957pspr0904_1

Parkinson, B. (2013). Contextualizing Facial Activity. *Emotion Review*, *5*(1), 97–103. https://doi.org/10.1177/1754073912457230

Pei, E., Oveneke, M. C., Zhao, Y., Jiang, D., & Sahli, H. (2021). Monocular 3D Facial Expression Features for Continuous Affect Recognition. *IEEE Transactions on Multimedia*, *23*, 3540–3550. https://doi.org/10.1109/TMM.2020.3026894

Pfister, T., Li, X., Zhao, G., & Pietikainen, M. (2011). Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework. *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 868–875. https://doi.org/10.1109/ICCVW.2011.6130343

Pfister, T., Li, X., Zhao, G., & Pietikainen, M. (2011). Recognising spontaneous facial micro-expressions. *2011 International Conference on Computer Vision*, 1449–1456. https://doi.org/10.1109/ICCV.2011.6126401

Philippot, P., Schaefer, A., & Herbette, G. (2003). Consequences of specific processing of emotional information: Impact of general versus specific autobiographical memory priming on emotion elicitation. *Emotion*, *3*(3), 270–283. https://doi.org/10.1037/1528-3542.3.3.270

Piironen, J., & Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, *11*(2). https://doi.org/10.1214/17-EJS1337SI

Plouffe-Demers, M.-P., Fiset, D., Saumure, C., Duncan, J., & Blais, C. (2019). Strategy Shift Toward Lower Spatial Frequencies in the Recognition of Dynamic Facial Expressions of Basic Emotions: When It Moves It Is Different. *Front. Psychol.* 10, 1563. https://doi.org/10.3389/fpsyg.2019.01563.

Pollick, F. E., Hill, H., Calder, A., & Paterson, H. (2003). Recognising Facial Expression from Spatially and Temporally Modified Movements. *Perception*, *32*(7), 813–826. https://doi.org/10.1068/p3319

Powell, W. R., & Schirillo, J. A. (2009). Asymmetrical facial expressions in portraits and hemispheric laterality: A literature review. *Laterality: Asymmetries of Body, Brain and Cognition*, *14*(6), 545–572. https://doi.org/10.1080/13576500802680336

Qu, F., Wang, S.-J., Yan, W.-J., & Fu, X. (2016). CAS(ME)2: A Database of Spontaneous Macro-expressions and Micro-expressions. In M. Kurosu (Ed.), *Human-Computer Interaction. Novel User Experiences* (Vol. 9733, pp. 48–59). Springer International Publishing. https://doi.org/10.1007/978-3-319-39513-5_5

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

Recio, G., Schacht, A., & Sommer, W. (2014). Recognizing dynamic facial expressions of emotion: Specificity and intensity effects in event-related brain potentials. *Biological Psychology*, *96*, 111–125. https://doi.org/10.1016/j.biopsycho.2013.12.003

Recio, G., Sommer, W., & Schacht, A. (2011). Electrophysiological correlates of perceiving and evaluating static and dynamic facial emotional expressions. *Brain Research*, *1376*, 66–75. https://doi.org/10.1016/j.brainres.2010.12.041

Reddy, S. P. T., Karri, S. T., Dubey, S. R., & Mukherjee, S. (2019). *Spontaneous Facial Micro-Expression Recognition using 3D Spatiotemporal Convolutional Neural Networks*. https://doi.org/10.48550/ARXIV.1904.01390

Rehg, J. M., Abowd, G. D., Rozga, A., Romero, M., Clements, M. A., Sclaroff, S., Essa, I., Ousley, O. Y., Li, Y., Kim, C., Rao, H., Kim, J. C., Presti, L. L., Zhang, J., Lantsman, D., Bidwell, J., & Ye, Z. (2013). Decoding Children's Social Behavior. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 3414–3421. https://doi.org/10.1109/CVPR.2013.438

Reisenzein, R., Bördgen, S., Holtbernd, T., & Matz, D. (2006). Evidence for strong dissociation between emotion and facial displays: The case of surprise. *Journal of Personality and Social Psychology*, *91*(2), 295–315. https://doi.org/10.1037/0022-3514.91.2.295

Righart, R., & De Gelder, B. (2008). Recognition of facial expressions is influenced by emotional scene gist. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(3), 264–272. https://doi.org/10.3758/CABN.8.3.264

Rinn, W. E. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin*, *95*(1), 52–77. https://doi.org/10.1037/0033-2909.95.1.52

Robert W., L. (2007). Emotion Elicitation With Neurological Patients. In *Handbook of emotion elicitation and assessment* (pp. 158–168). Oxford University Press.

Rodger, H., Lao, J., & Caldara, R. (2018). Quantifying facial expression signal and intensity use during development. *Journal of Experimental Child Psychology*, *174*, 41–59. https://doi.org/10.1016/j.jecp.2018.05.005

Rosas, V., Mihalcea, R., & Morency, L.-P. (2013). Multimodal Sentiment Analysis of Spanish Online Videos. *IEEE Intelligent Systems*, *28*(3), 38–45. https://doi.org/10.1109/MIS.2013.9

Rosenberg, E. L., & Ekman, P. (Eds.). (2020). *What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS)* (Third edition). Oxford University Press.

Rotshtein, P., Richardson, M. P., Winston, J. S., Kiebel, S. J., Vuilleumier, P., Eimer, M., Driver, J., & Dolan, R. J. (2010). Amygdala damage affects event-related potentials for fearful faces at specific time windows. *Human Brain Mapping*, *31*(7), 1089–1105. https://doi.org/10.1002/hbm.20921

Russell, C. A., Swasy, J. L., Russell, D. W., & Engel, L. (2017). Eye-tracking evidence that happy faces impair verbal message comprehension: The case of health warnings in direct-to-consumer pharmaceutical television commercials. *International Journal of Advertising*, *36*(1), 82–106. https://doi.org/10.1080/02650487.2016.1196030

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Russell, J. A. (1993). Forced-choice response format in the study of facial expression. *Motivation and Emotion*, *17*(1), 41–51. https://doi.org/10.1007/BF00995206

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115*(1), 102–141. https://doi.org/10.1037/0033-2909.115.1.102

Ryumina, E., & Karpov, A. (2020). Facial Expression Recognition using Distance Importance Scores Between Facial Landmarks. *Proceedings of the 30th International Conference on Computer Graphics and Machine Vision (GraphiCon 2020). Part 2*, paper32-1-paper32-10. https://doi.org/10.51130/graphicon-2020-2-3-32

Ryumina, E., Dresvyanskiy, D., & Karpov, A. (2022). In search of a robust facial expressions recognition model: A large-scale visual cross-corpus study. *Neurocomputing*, *514*, 435–450. https://doi.org/10.1016/j.neucom.2022.10.013

Saganowski, S., Komoszyńska, J., Behnke, M., Perz, B., Kunc, D., Klich, B., Kaczmarek, Ł. D., & Kazienko, P. (2022). Emognition dataset: Emotion recognition with self-reports, facial expressions, and physiology using wearables. *Scientific Data*, *9*(1), 158. https://doi.org/10.1038/s41597-022-01262-0

Saha, P., Bhattacharjee, D., De, B. K., & Nasipuri, M. (2019). A Survey on Image Acquisition Protocols for Non-posed Facial Expression Recognition Systems. *Multimedia Tools and Applications*, *78*(16), 23329–23368. https://doi.org/10.1007/s11042-019-7596-2

Sandbach, G., Zafeiriou, S., Pantic, M., & Yin, L. (2012). Static and dynamic 3D facial expression recognition: A comprehensive survey. *Image and Vision Computing*, *30*(10), 683–697. https://doi.org/10.1016/j.imavis.2012.06.005

Hyunwoo Kim

Sannikov, K. A., Bashlikov, A. A., & Druki, A. A. (2017). Two-level algorithm of facial expressions classification on complex background. *2017 International Siberian Conference on Control and Communications (SIBCON)*, 1–5. https://doi.org/10.1109/SIBCON.2017.7998594

Sariyanidi, E., Gunes, H., & Cavallaro, A. (2017). Robust Registration of Dynamic Facial Sequences. *IEEE Transactions on Image Processing*, *26*(4), 1708–1722. https://doi.org/10.1109/TIP.2016.2639448

Sato, W., & Yoshikawa, S. (2004). BRIEF REPORT The dynamic aspects of emotional facial expressions. *Cognition & Emotion* 18, 701–710. https://doi.org/10.1080/02699930341000176.

Sato, W., & Yoshikawa, S. (2007). Enhanced Experience of Emotional Arousal in Response to Dynamic Facial Expressions. *Journal of Nonverbal Behavior*, *31*(2), 119–135. https://doi.org/10.1007/s10919-007-0025-7

Sato, W., Yoshikawa, S., Kochiyama, T., & Matsumura, M. (2004). The amygdala processes the emotional significance of facial expressions: An fMRI investigation using the interaction between expression and face direction. *NeuroImage*, *22*(2), 1006–1013. https://doi.org/10.1016/j.neuroimage.2004.02.030

Saumure, C., Plouffe-Demers, M.-P., Estéphan, A., Fiset, D., & Blais, C. (2018). The use of visual information in the recognition of posed and spontaneous facial expressions. *Journal of Vision*, *18*(9), 21. https://doi.org/10.1167/18.9.21

Sauter, D. A., & Fischer, A. H. (2018). Can perceivers recognise emotions from spontaneous expressions? *Cognition and Emotion*, *32*(3), 504–515. https://doi.org/10.1080/02699931.2017.1320978

Sayette, M. A., Cohn, J. F., Wertz, J. M., Perrott, M. A., & Parrott, D. J. (2001). [No title found]. *Journal of Nonverbal Behavior*, *25*(3), 167–185. https://doi.org/10.1023/A:1010671109788

Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition & Emotion*, *24*(7), 1153–1172. https://doi.org/10.1080/02699930903274322

Scherer, K. R. (2005). Appraisal Theory. In T. Dalgleish & M. J. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 637–663). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470013494.ch30

Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion*, *7*(1), 113–130. https://doi.org/10.1037/1528-3542.7.1.113

Scherer, K. R., Feldstein, S., Bond, R. N., & Rosenthal, R. (1985). Vocal cues to deception: A comparative channel approach. *Journal of Psycholinguistic Research*, *14*(4), 409–425. https://doi.org/10.1007/BF01067884

Scherer, K. R., Ladd, D. R., & Silverman, K. E. A. (1984). Vocal cues to speaker affect: Testing two models. *The Journal of the Acoustical Society of America*, *76*(5), 1346–1356. https://doi.org/10.1121/1.391450

Schmidt, K. L., Ambadar, Z., Cohn, J. F., & Reed, L. I. (2006). Movement Differences between Deliberate and Spontaneous Facial Expressions: Zygomaticus Major Action in Smiling. *Journal of Nonverbal Behavior*, *30*(1), 37–52. https://doi.org/10.1007/s10919-005-0003-x

Schmidt, K. L., Bhattacharya, S., & Denlinger, R. (2009). Comparison of deliberate and spontaneous facial movement in smiles and eyebrow raises. *Journal of Nonverbal Behavior*, *33*(1), 35–45. https://doi.org/10.1007/s10919-008-0058-6

Schmidt, K. L., Cohn, J. F., & Tian, Y. (2003). Signal characteristics of spontaneous facial expressions: Automatic movement in solitary and social smiles. *Biological Psychology*, *65*(1), 49–66. https://doi.org/10.1016/S0301-0511(03)00098-X

Schmidt, S., Tinti, C., Levine, L. J., & Testa, S. (2010). Appraisals, emotions and emotion regulation: An integrative approach. *Motivation and Emotion*, *34*(1), 63–72. https://doi.org/10.1007/s11031-010-9155-z

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal* 27, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Sheldon, K. M., Corcoran, M., & Sheldon, M. (2021). Duchenne Smiles as Honest Signals of Chronic Positive Mood. *Perspectives on Psychological Science*, *16*(3), 654–666. https://doi.org/10.1177/1745691620959831

Shi, Q., Eriksson, A., Van Den Hengel, A., & Shen, C. (2011). Is face recognition really a Compressive Sensing problem? *CVPR 2011*, 553–560. https://doi.org/10.1109/CVPR.2011.5995556

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X., & Yang, X. (2018). A Review of Emotion Recognition Using Physiological Signals. *Sensors*, *18*(7), 2074. https://doi.org/10.3390/s18072074

Siddiqui, N., Reither, T., Dave, R., Black, D., Bauer, T., & Hanson, M. (2022). A Robust Framework for Deep Learning Approaches to Facial Emotion Recognition and Evaluation. *2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML)*, 68–73. https://doi.org/10.1109/CACML55074.2022.00020

Siedlecka, E., & Denson, T. F. (2019). Experimental Methods for Inducing Basic Emotions: A Qualitative Review. *Emotion Review*, *11*(1), 87–97. https://doi.org/10.1177/1754073917749016

Skiendziel, T., Rösch, A. G., & Schultheiss, O. C. (2019). Assessing the convergent validity between the automated emotion recognition software Noldus FaceReader 7 and Facial Action Coding System Scoring. *PLOS ONE*, *14*(10), e0223905. https://doi.org/10.1371/journal.pone.0223905

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*(4), 813–838. https://doi.org/10.1037/0022-3514.48.4.813

Smith, C. A., McHugo, G. J., & Lanzetta, J. T. (1986). The facial muscle patterning of posed and imagery-induced expressions of emotion by expressive and nonexpressive posers. *Motivation and Emotion*, *10*(2), 133–157. https://doi.org/10.1007/BF00992252

Sneddon, I., McRorie, M., McKeown, G., & Hanratty, J. (2012). The Belfast Induced Natural Emotion Database. *IEEE Transactions on Affective Computing*, *3*(1), 32–41. https://doi.org/10.1109/T-AFFC.2011.26

Soussignan, R. (2002). Duchenne smile, emotional experience, and autonomic reactivity: A test of the facial feedback hypothesis. *Emotion*, *2*(1), 52–74. https://doi.org/10.1037/1528-3542.2.1.52

Soussignan, R., Dollion, N., Schaal, B., Durand, K., Reissland, N., & Baudouin, J.-Y. (2018). Mimicking emotions: How 3–12-month-old infants use the facial expressions and eyes of a model. *Cognition and Emotion*, *32*(4), 827–842. https://doi.org/10.1080/02699931.2017.1359015

Hyunwoo Kim

Stanciu, L., & Albu, A. (2019). Analysis on emotion detection and recognition methods using facial microexpressions. A review. *2019 E-Health and Bioengineering Conference (EHB)*, 1–4. https://doi.org/10.1109/EHB47216.2019.8969925

Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2018). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods*, *50*(4), 1446–1460. https://doi.org/10.3758/s13428-017-0996-1

Stratou, G., Ghosh, A., Debevec, P., & Morency, L.-P. (2011). Effect of illumination on automatic expression recognition: A novel 3D relightable facial database. *Face and Gesture 2011*, 611–618. https://doi.org/10.1109/FG.2011.5771467

Sun, B., Li, L., Zhou, G., & He, J. (2016). Facial expression recognition in the wild based on multimodal texture features. *Journal of Electronic Imaging*, *25*(6), 061407. https://doi.org/10.1117/1.JEI.25.6.061407

Tanaka, J. W., Kaiser, M. D., Butler, S., & Le Grand, R. (2012). Mixed emotions: Holistic and analytic perception of facial expressions. *Cognition & Emotion* 26, 961–977. https://doi.org/10.1080/02699931.2011.630933.

Tcherkassof, A., & Dupré, D. (2021). The emotion–facial expression link: Evidence from human and automatic expression recognition. *Psychological Research*, *85*(8), 2954–2969. https://doi.org/10.1007/s00426-020-01448-4

Tcherkassof, A., Dupré, D., Meillon, B., Mandran, N., Dubois, M., & Adam, J.-M. (2013). Dynemo: A Video Database of Natural Facial Expressions of Emotions. *The International Journal of Multimedia & Its Applications*, *5*(5), 61–80. https://doi.org/10.5121/ijma.2013.5505

Thomas, L. A., De Bellis, M. D., Graham, R., & LaBar, K. S. (2007). Development of emotional facial recognition in late childhood and adolescence. *Developmental Science*, *10*(5), 547–558. https://doi.org/10.1111/j.1467-7687.2007.00614.x

Tobin, A., Favelle, S., & Palermo, R. (2016). Dynamic facial expressions are processed holistically, but not more holistically than static facial expressions. *Cognition and Emotion*, 30, 1208–1221. https://doi.org/10.1080/02699931.2015.1049936.

Tong, Chen, J., & Ji, Q. (2010). A Unified Probabilistic Framework for Spontaneous Facial Action Modeling and Understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(2), 258–273. https://doi.org/10.1109/TPAMI.2008.293

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Tracy, J. L., & Robins, R. W. (2007). The prototypical pride expression: Development of a nonverbal behavior coding system. *Emotion*, *7*(4), 789–801. https://doi.org/10.1037/1528-3542.7.4.789

Turk, M., & Pentland, A. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, *3*(1), 71–86. https://doi.org/10.1162/jocn.1991.3.1.71

Uhrig, M. K., Trautmann, N., Baumgärtner, U., Treede, R.-D., Henrich, F., Hiller, W., & Marschall, S. (2016). Emotion Elicitation: A Comparison of Pictures and Films. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00180

Valentine, T., & Bruce, V. (1988). Mental rotation of faces. *Memory & Cognition*, *16*(6), 556–566. https://doi.org/10.3758/BF03197057

Valstar, M. F., Almaev, T., Girard, J. M., McKeown, G., Mehu, M., Yin, L., Pantic, M., & Cohn, J. F. (2015). FERA 2015—Second Facial Expression Recognition and Analysis challenge. *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 1–8. https://doi.org/10.1109/FG.2015.7284874

Valstar, M. F., Jiang, B., Mehu, M., Pantic, M., & Scherer, K. (2011). The first facial expression recognition and analysis challenge. *Face and Gesture 2011*, 921–926. https://doi.org/10.1109/FG.2011.5771374

Valstar, M. F., Mehu, M., Bihan Jiang, Pantic, M., & Scherer, K. (2012). Meta-analysis of the first facial expression recognition challenge. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *42*(4), 966–979. https://doi.org/10.1109/TSMCB.2012.2200675

Valstar, M. F., Pantic, M., Ambadar, Z., & Cohn, J. F. (2006). Spontaneous vs. posed facial behavior: Automatic analysis of brow actions. *Proceedings of the 8th International Conference on Multimodal Interfaces*, 162–170. https://doi.org/10.1145/1180995.1181031

Valstar, M. F., Sanchez-Lozano, E., Cohn, J. F., Jeni, L. A., Girard, J. M., Zhang, Z., Yin, L., & Pantic, M. (2017). FERA 2017—Addressing head pose in the third facial expression recognition and analysis challenge. *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 839–847. https://doi.org/10.1109/FG.2017.107

Van Der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion*, *11*(4), 907–920. https://doi.org/10.1037/a0023853

Van Erp, S., Oberski, D. L., & Mulder, J. (2019). Shrinkage priors for Bayesian penalized regression. *Journal of Mathematical Psychology*, *89*, 31–50. https://doi.org/10.1016/j.jmp.2018.12.004

Van Kleef, G. A. (2016). *The Interpersonal Dynamics of Emotion: Toward an Integrative Theory of Emotions as Social Information* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781107261396

Vanny, M., Ko, K.-E., Park, S.-M., & Sim, K.-B. (2013). Physiological Responses-Based Emotion Recognition Using Multi-Class SVM with RBF Kernel. *Journal of Institute of Control, Robotics and Systems*, *19*(4), 364–371. https://doi.org/10.5302/J.ICROS.2013.13.1879

Wagner, H. L. (1990). The spontaneous facial expression of differential positive and negative emotions. *Motivation and Emotion*, *14*(1), 27–43. https://doi.org/10.1007/BF00995547

Wagner, H. L. (1997). Methods for the study of facial behavior. In J. A. Russell & J. M. Fernández-Dols (Eds.), *The Psychology of Facial Expression* (1st ed., pp. 31–54). Cambridge University Press. https://doi.org/10.1017/CBO9780511659911.004

Wagner, H. L., MacDonald, C. J., & Manstead, A. S. (1986). Communication of individual emotions by spontaneous facial expressions. *Journal of Personality and Social Psychology*, *50*(4), 737–743. https://doi.org/10.1037/0022-3514.50.4.737

Wallhoff, F., Schuller, B., Hawellek, M., & Rigoll, G. (2006). Efficient Recognition of Authentic Dynamic Facial Expressions on the Feedtum Database. *2006 IEEE International Conference on Multimedia and Expo*, 493–496. https://doi.org/10.1109/ICME.2006.262433

Wallraven, C., Breidt, M., Cunningham, D. W., & Bülthoff, H. H. (2008). Evaluating the perceptual realism of animated facial expressions. *ACM Trans. Appl. Percept.* 4, 1–20. https://doi.org/10.1145/1278760.1278764.

Walther, J. B., & D'Addario, K. P. (2001). The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication. *Social Science Computer Review*, *19*(3), 324–347. https://doi.org/10.1177/089443930101900307

Wang, F., Li, J., Zhang, S., Qi, C., Zhang, Y., & Miao, D. (2021). A Dynamic 3D Spontaneous Micro-expression Database: Establishment and Evaluation. *ArXiv*. https://doi.org/10.48550/ARXIV.2108.00166

Wang, S., Liu, Z., Lv, S., Lv, Y., Wu, G., Peng, P., Chen, F., & Wang, X. (2010). A Natural Visible and Infrared Facial Expression Database for Expression Recognition and Emotion Inference. *IEEE Transactions on Multimedia*, *12*(7), 682–691. https://doi.org/10.1109/TMM.2010.2060716

Wang, S., Liu, Z., Wang, Z., Wu, G., Shen, P., He, S., & Wang, X. (2013). Analyses of a Multimodal Spontaneous Facial Expression Database. *IEEE Transactions on Affective Computing*, *4*(1), 34–46. https://doi.org/10.1109/T-AFFC.2012.32

Wang, S., Yu, R., Tyszka, J. M., Zhen, S., Kovach, C., Sun, S., et al. (2017). The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity. *Nat Commun* 8, 14821. https://doi.org/10.1038/ncomms14821.

Wang, Z., Wang, S., Zhu, Y., & Ji, Q. (2012). Bias Analyses of Spontaneous Facial Expression Database. *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*.

Weber, R., Soladié, C., & Séguier, R. (2018). A Survey on Databases for Facial Expression Analysis: *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 73–84. https://doi.org/10.5220/0006553900730084

Webster, P. J., Wang, S., & Li, X. (2021). Review: Posed vs. Genuine Facial Emotion Recognition and Expression in Autism and Implications for Intervention. *Frontiers in Psychology*, *12*, 653112. https://doi.org/10.3389/fpsyg.2021.653112

Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology*, *78*(1), 105–119. https://doi.org/10.1037/0022-3514.78.1.105

Wells, L. J., Gillespie, S. M., & Rotshtein, P. (2016). Identification of emotional facial expressions: Effects of expression, intensity, and sex on eye gaze. *PLOS ONE*, *11*(12), e0168307. https://doi.org/10.1371/journal.pone.0168307

Widen, S. C., & Russell, J. A. (2015). Do Dynamic Facial Expressions Convey Emotions to Children Better Than Do Static Ones? *Journal of Cognition and Development* 16, 802–811. https://doi.org/10.1080/15248372.2014.916295.

Wieser, M. J., & Brosch, T. (2012). Faces in Context: A Review and Systematization of Contextual Influences on Affective Face Processing. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00471

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (2016). Validation of the Amsterdam Dynamic Facial Expression Set – Bath Intensity Variations (ADFES-BIV): A Set of Videos Expressing Low, Intermediate, and High Intensity Emotions. *PLOS ONE*, *11*(1), e0147112. https://doi.org/10.1371/journal.pone.0147112

Xiao, N. G., Perrotta, S., Quinn, P. C., Wang, Z., Sun, Y.-H. P., & Lee, K. (2014). On the facilitative effects of face motion on face recognition and its development. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00633

Yan, W.-J., Li, X., Wang, S.-J., Zhao, G., Liu, Y.-J., Chen, Y.-H., & Fu, X. (2014). CASME II: An Improved Spontaneous Micro-Expression Database and the Baseline Evaluation. *PLoS ONE*, *9*(1), e86041. https://doi.org/10.1371/journal.pone.0086041

Yan, W., Wu, Q., Liu, Y., Wang, S., & Fu, X. (2013). CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces. *2013 10th IEEE*

International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 1–7. https://doi.org/10.1109/FG.2013.6553799

Yeasin, M., Bullot, B., & Sharma, R. (2006). Recognition of facial expressions and measurement of levels of interest from video. *IEEE Transactions on Multimedia*, *8*(3), 500–508. https://doi.org/10.1109/TMM.2006.870737

Yik Zhaolan Meng., James A. Russ, M. (1998). Brief Report- Adults' Freely Produced Emotion Labels for Babies' Spontaneous Facial Expressions. *Cognition & Emotion*, *12*(5), 723–730. https://doi.org/10.1080/026999398379529

Yitzhak, N., Giladi, N., Gurevich, T., Messinger, D. S., Prince, E. B., Martin, K., & Aviezer, H. (2017). Gently does it: Humans outperform a software classifier in recognizing subtle, nonstereotypical facial expressions. *Emotion*, *17*(8), 1187–1198. https://doi.org/10.1037/emo0000287

Yitzhak, N., Pertzov, Y., Guy, N., & Aviezer, H. (2022). Many ways to see your feelings: Successful facial expression recognition occurs with diverse patterns of fixation distributions. *Emotion* 22, 844–860. https://doi.org/10.1037/emo0000812.

Yoshikawa, S., & Sato, W. (2008). Dynamic facial expressions of emotion induce representational momentum. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(1), 25–31. https://doi.org/10.3758/CABN.8.1.25

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition* 63, 271–313. https://doi.org/10.1016/S0010-0277(97)00003-6.

Zavarez, M. V., Berriel, R. F., & Oliveira-Santos, T. (2017). Cross-Database Facial Expression Recognition Based on Fine-Tuned Deep Convolutional Network. *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 405–412. https://doi.org/10.1109/SIBGRAPI.2017.60

Zeng, J., Shan, S., & Chen, X. (2018). Facial Expression Recognition with Inconsistently Annotated Datasets. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018* (Vol. 11217, pp. 227–243). Springer International Publishing. https://doi.org/10.1007/978-3-030-01261-8_14

Zeng, Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 39–58. https://doi.org/10.1109/TPAMI.2008.52.

Zhalehpour, S., Onder, O., Akhtar, Z., & Erdem, C. E. (2017). BAUM-1: A Spontaneous Audio-Visual Face Database of Affective and Mental States. *IEEE Transactions on Affective Computing*, *8*(3), 300–313. https://doi.org/10.1109/TAFFC.2016.2553038

Zhang, L., Walter, S., Xueyao Ma, Werner, P., Al-Hamadi, A., Traue, H. C., & Gruss, S. (2016). "BioVid Emo DB": A multimodal database for emotion analyses validated by subjective ratings. *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–6. https://doi.org/10.1109/SSCI.2016.7849931

Zhang, T., Zheng, W., Cui, Z., Zong, Y., Yan, J., & Yan, K. (2016). A Deep Neural Network-Driven Feature Learning Method for Multi-view Facial Expression Recognition. *IEEE Transactions on Multimedia*, *18*(12), 2528–2536. https://doi.org/10.1109/TMM.2016.2598092

Zhang, T., Zong, Y., Zheng, W., Chen, C. L. P., Hong, X., Tang, C., Cui, Z., & Zhao, G. (2022). Cross-Database Micro-Expression Recognition: A Benchmark. *IEEE Transactions on Knowledge and Data Engineering*, *34*(2), 544–559. https://doi.org/10.1109/TKDE.2020.2985365

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), 692–706. https://doi.org/10.1016/j.imavis.2014.06.002

Zhao, K., Zhao, J., Zhang, M., Cui, Q., & Fu, X. (2017). Neural Responses to Rapid Facial Expressions of Fear and Surprise. *Frontiers in Psychology*, *8*, 761. https://doi.org/10.3389/fpsyg.2017.00761

Zloteanu, M., & Krumhuber, E. G. (2021). Expression Authenticity: The Role of Genuine and Deliberate Displays in Emotion Perception. *Frontiers in Psychology*, *11*, 611248. https://doi.org/10.3389/fpsyg.2020.611248

Zloteanu, M., Krumhuber, E. G., & Richardson, D. C. (2018). Detecting Genuine and Deliberate Displays of Surprise in Static and Dynamic Faces. *Frontiers in Psychology*, *9*, 1184. https://doi.org/10.3389/fpsyg.2018.01184

Zuckerman, M., Hall, J. A., DeFrank, R. S., & Rosenthal, R. (1976). Encoding and decoding of spontaneous and posed facial expressions. *Journal of Personality and Social Psychology*, *34*(5), 966–977. https://doi.org/10.1037/0022-3514.34.5.966

Zupan, B., & Eskritt, M. (2020). Eliciting emotion ratings for a set of film clips: A preliminary archive for research in emotion. *The Journal of Social Psychology*, *160*(6), 768–789. https://doi.org/10.1080/00224545.2020.1758016