Spatiotemporal structure and substates in emotional facial expressions

Hélio Clemente José Cuve[1*], Sophie Sowden-Carvalho[2], Jennifer Cook[2]

[1]School of Psychological Science, University of Bristol, Bristol, United Kingdom

[2]School of Psychological Science, University of Birmingham, Birmingham, United Kingdom

[*] Corresponding author

e-mail: helio.cuve@bristol.ac.uk

School of Psychological Science

University of Bristol

12a, Priory Road

Bristol BS8 1TU

United Kingdom

Abstract

Facial expressions are tools for modulating social interaction, from the display of overt expressions to subtle cues like a raised eyebrow that accompanies speech. How this complex signalling capacity is achieved remains only partially explored. An overlooked factor is the dynamic form of facial expressions signalling, particularly how facial actions combine and recombine over time to produce nuanced expressions and enrich speech. Previous research has focused largely on the static aspects of facial behaviour, reflecting the theoretical and methodological challenges in modelling the production and perception of dynamic social signals. Drawing on theories of motor control we leveraged facial-motion tracking and spatiotemporal dimensionality reduction to investigate the structure and function of facial signalling dynamics. We show that despite the complexity of facial expressions, their emotion-signalling function is achieved through a few fundamental dynamic patterns with subtle yet diagnostic kinematic differences. Classification analysis further show that spatiotemporal components reliably differentiate emotion signals in *Expressions only* and *Emotive-speech* conditions. The underlying spatiotemporal structure represents an efficient encoding strategy, optimising the transmission and perception of facial social signals. These insights have implications for understanding normative and atypical face-to-face signalling and inform the optimal design of non-verbal expressive capabilities in artificial social agents. This work also contributes new methods for analysing facial movements, applicable to broader aspects human multimodal social communication.

Keywords: Facial expressions, movement, spatiotemporal, dimensionality reduction, emotion, signalling.

Author Summary

In face-to-face interactions, facial movements play a crucial role in conveying both emotional and speech-related information. Yet, we don't fully understand how such complex signalling is achieved. Previous research has mainly focused on static signals, such as a single moment during a smile or frown. This study instead explores how facial movements change and evolve over time, particularly in relation to producing communicative emotions. Using advanced techniques to measure and analyse moment-to-moment changes in dozens of facial muscles during facial expression and speech production, we identified fundamental movement patterns that differentiate emotional content. These dynamic patterns enable us to communicate both isolated emotional signals through facial expressions, but also to simultaneously communicate emotions during speech, by adding subtle cues like a raised eyebrow or a brief smile. Our findings have implications to understand typical and atypical face-to-face interactions abilities and could enhance the design of biologically inspired and human-like facial expressions for social robots and virtual agents. Additionally, we contribute new methods for analysing facial movements, applicable to broader aspects of human multimodal social communication.

Introduction

The human face is an incredibly versatile tool for social communication [1] capable of producing a rich array of non-verbal expressions through the coordination of basic motor facial action units (AUs) [2]. For example, frowning (AU4) combined with pursed lips (AU24) can convey a prototypical facial expression of anger; while raising the cheeks (AU6) and pulling back the lip corners (AU12) can signal a smile. In addition to ostensive non-verbal emotional cues, in real-life social interaction, facial expressions simultaneously modulate verbal communication. A slight tightening of the eyelids (AU7) for example, can make a “neutral” verbal utterance appear slightly angrier. The mechanisms that afford such flexible expressive capabilities to facial expressions remain only partially understood.

One overlooked factor in previous research is the dynamic nature of facial expressions. Most previous research has focused on the static aspects of facial behaviour, or aggregates of expressions that average out temporal dynamics [3]. Previous studies also largely ignore conditions that may carry a dynamic advantage, such as conversational or speech related facial expressions where dynamics simultaneously transmit visual cues for emotion and verbal communication [3–6]. Indeed, recent studies have started to reveal the importance of dynamics in both production and perception of facial expressions. Psychophysics inspired approaches have combined generative synthesis of facial movements [7] and reverse correlation to parametrically map the spatiotemporal features of AUS to perceptual representations of emotion categories [8], valence-arousal dimensions [9] and even social traits like dominance [10]. Similarly, a recent production and perception study recorded participants' facial movements during expression of anger, happiness, and

sadness, as well as during spoken sentences conveying the same emotions [11]. Kinematic analysis revealed faster movements for high-arousal emotions (happiness, anger) compared to slower movements for low-arousal emotions (sadness). Thus, suggesting that dynamic information in facial movements carries a diagnostic signal for differentiating facial expressions.

However, a direct exploration of the spatiotemporal structure of facial expressions—that is, how AUs dynamically combine and unfold over time and space - and how these patterns relate to their emotion signalling function remains lacking. Notably, one significant challenge to studying the role of spatiotemporal structure in facial expression production is the lack of a suitable theoretical and methodological framework. Such a framework should ideally integrate the dynamic nature of facial expression signals with their socio-communicative functions and provide an appropriate modelling strategy. In this work, we draw on the observation that facial expressions, like other facial and bodily movements, are controlled by an underlying motor system [12,13]. We argue that the motor control literature provides a valuable theoretical and analytical framework to investigate spatiotemporal structure and function of dynamics in facial expressions.

A motor control perspective provides two useful insights that are applicable to studying the spatiotemporal structure and function of facial expressions. First, a longstanding motor control hypothesis suggests that body movements are generated by a low-dimensional mechanism that underpins their effective neural control and physical execution [14–16]. Such low-dimensionality of motor control and execution might reflect the brain's solution to the problem of controlling numerous muscles and

components of the motor system whilst minimising computational demands - often referred to as the "degrees of freedom" problem [17–19]. Crucially, such low dimensionality has been shown to describe motion patterns during walking [20], whole body pointing [21], eye movements [22] and articulatory speech movements [23]. Indeed, while not primarily serving a communicative function, these movements are known to also convey emotion-related information [24–26], therefore suggesting shared control mechanisms between social and non-social aspects of movement. Applied to facial expressions, this raises the possibility that dynamics underlying facial expressions, and their ability to multiplex information (e.g., visually signal emotion and speech), may be structured around flexible low-dimensional spatiotemporal patterns. Such low dimensional patterns could, for example, reflect how AUs combine and evolve over time during facial expression production. Such spatiotemporal structure, can be explored using data-driven models that learn a reduced dimensionality from high-dimensional AU movement data [27]. Thus, this perspective also suggests a methodological framework to model the spatiotemporal nature of movement, which is rare in the study of facial expression production.

A second insight from the motor-control perspective is the notion that although outwardly fluid, movements can often be segmented into distinct 'substates' often seen in differential modulation of kinematics (e.g., speed, displacement of m motor actions). For instance, walking movements can be segmented into 'swing' or 'stance' phases [20] or eye movements into 'fixations' and 'saccades' [28], and even more fine-grained states (sub movements) [29]. Facial expressions are likely to have substates related to the biomechanics of how they are produced, encompassing 'transitions'—where action units are initiated or inhibited, - and 'stability' – where actions units are sustained

[30,31]. However, to date facial expressions substates have not been formally characterised, and whether and how they relate to emotional signalling is unknown Similar to spatiotemporal components, substates can be explored in a data-driven fashion, using methods such as spatiotemporal clustering [27,32,33] without *apriori* defined kinematic profiles.

The characterisation of underlying spatiotemporal dimensions and substates in facial expressions would enable a richer characterisation of facial dynamics that is both holistic, i.e., captures overall spatiotemporal patterns, but also focal as it can capture micro spatiotemporal patterns in movement data. Crucially, we hypothesise that these underlying patterns are not merely incidental; rather, they are diagnostic of emotional content, reflecting how facial movements are optimised for social signalling.

The current study therefore aims to (1) describe the spatiotemporal structure underlying facial expression production; (2) explore and characterise potential facial expression substates, and 3) evaluate the diagnostic value of spatiotemporal structure and substates for emotion signalling. To achieve this, we videoed participants while they produced facial movements in a) *Expression only* and b) *Emotive speech* conditions. We used motion tracking and automated Facial Action Coding System (FACS) to describe the spatiotemporal patterns of the facial action units (e.g., mouth widening, eyebrow raise, etc). We then used Non-Negative Matrix factorization (NMF) [34,35] a part-based dimensionality reduction approach to explore whether high-dimensional spatiotemporal patterns (e.g., AU weights over time) can be compressed into a set of diagnostic low-dimensional latent spatiotemporal patterns (groups of coactivated AUs and salient temporal patterns) – see Methods. To precis our results,

we find that facial expression dynamics can be distilled into a set of three fundamental spatiotemporal components that form the basis for flexible signalling. Classification and regression analysis (see Methods) show that emotion (happy, angry, sad) and conditions (*Expression only* and *Emotive speech*) are differentiated by synergistic mixing of components and subtle substate differences related to expression transitions. These results suggests that the complex and flexible signalling of emotional information through facial movements is achieved through variations in a small number of spatiotemporal components and substates.

Results

Spatiotemporal components for facial expression production.

<u>Spatiotemporal patterns during facial *Expression only* condition</u>

We first looked at the spatiotemporal components identified via the NMF, which represent the underlying spatiotemporal patterns (e.g., groups of AUs with common activation timecourses and salient temporal patterns). The optimal number of spatiotemporal components (*k*) optimising for various fit indices (see Methods) – was consistently three (see Fig 1 – top panel) (see also *Expression only* condition - NMF validation analysis in Supplementary Information – Fig S1-3).

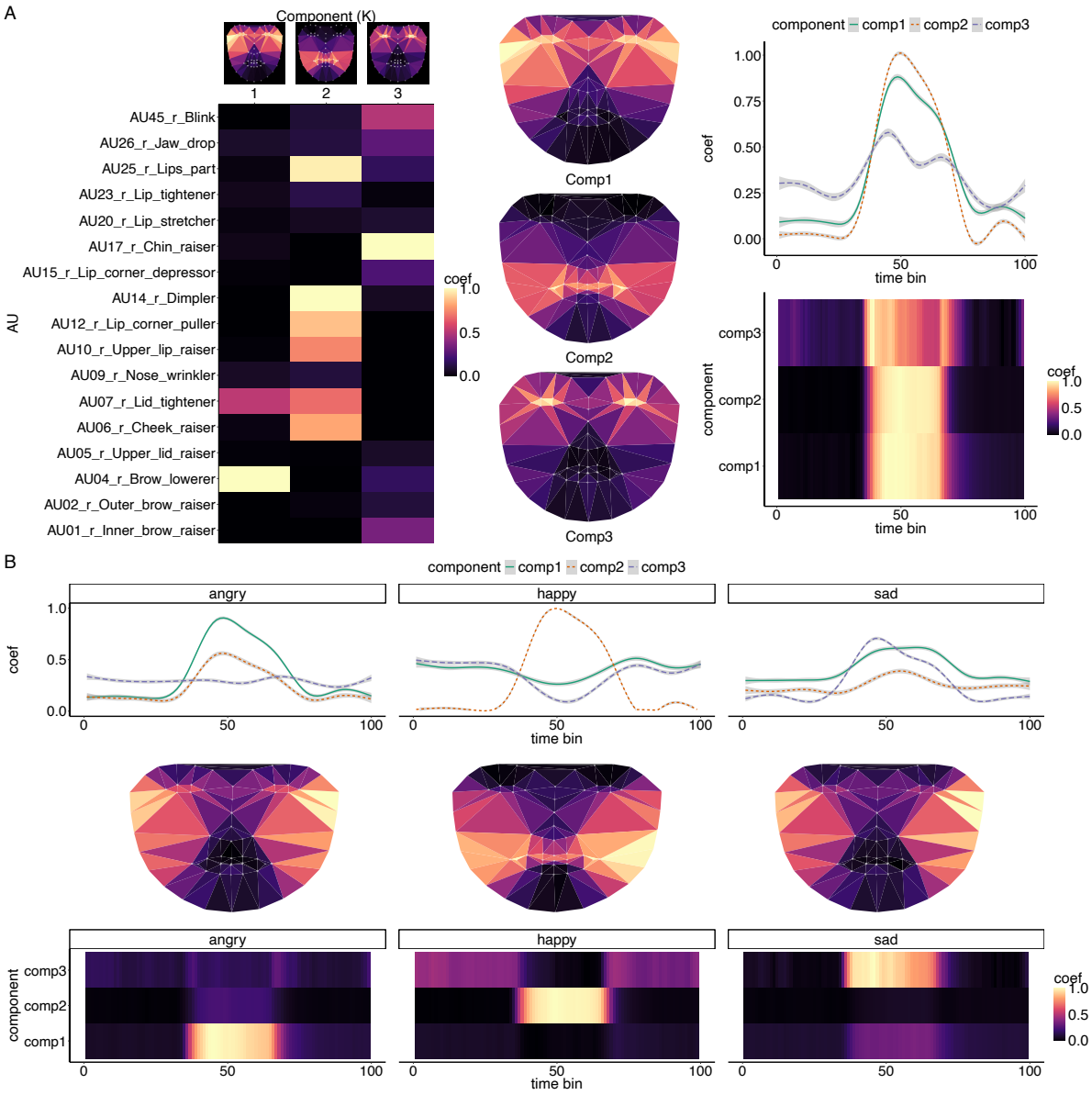

Figure 1. Spatiotemporal components for facial expressions only production. The top panel (A) shows that expression production can be summarised by three components, capturing spatiotemporal patterns for upper face AUs (component 1), lower face AUs (component 2) and a combination of both lower and upper face AUS (component 3). Top - middle and right columns show approximate spatial distribution of components in an average face model and how it changes over time, respectively. The bottom panel (B) highlight spatiotemporal profiles for angry, happy and sad expressions. Coef = coefficient from the NMF, brighter colours indicate stronger activation. Note while heatmaps are drawn to scale, face maps are drawn for visualisation purposes only.

Evaluating the spatiotemporal patterns revealed that component 1 (C1) had stronger activation of the upper face-related AUs (e.g., AU4 brow lowerer, AU7 lid tightener)

compared to other AUs. Component 2 (C2) had strong activation of both lower and upper face-related AUs (e.g., AU25 lips part, AU14 dimpler). Component 3 (C3) was mostly dominated by lower face actions (e.g., AU17 chin raiser) with some small activation of upper face AUs (e.g. AU45 blink). We name these components, upper face AUs (C1), lower face AUs (C2) and lower-to-upper face AUs (C3). The temporal profile of these components is shown in Fig 1 B. C1 (upper face AUs) and C2 (lower face AUS) had a common inverted-U shape with a clear singular peak, whereas C3 (lower to upper face AUs) had a less pronounced but more dynamic pattern (e.g. bimodal peaks).

Next, we assessed the predictive value of the spatiotemporal components, by analysing how they differed between expressions and validated it in a held-out test set (see Methods - *Multivariate Timeseries Classification*). When assessing spatiotemporal components' ability to classify expressions we achieved significantly above chance accuracy of over 90% (ACC= 0.91 [BalACC = .93], Kappa = 0.87; AUCmean = 0.98, $p < .001$ see Fig 2). These results indicate strong agreement between prediction and ground-truth – see also SI1 – Table S1-2

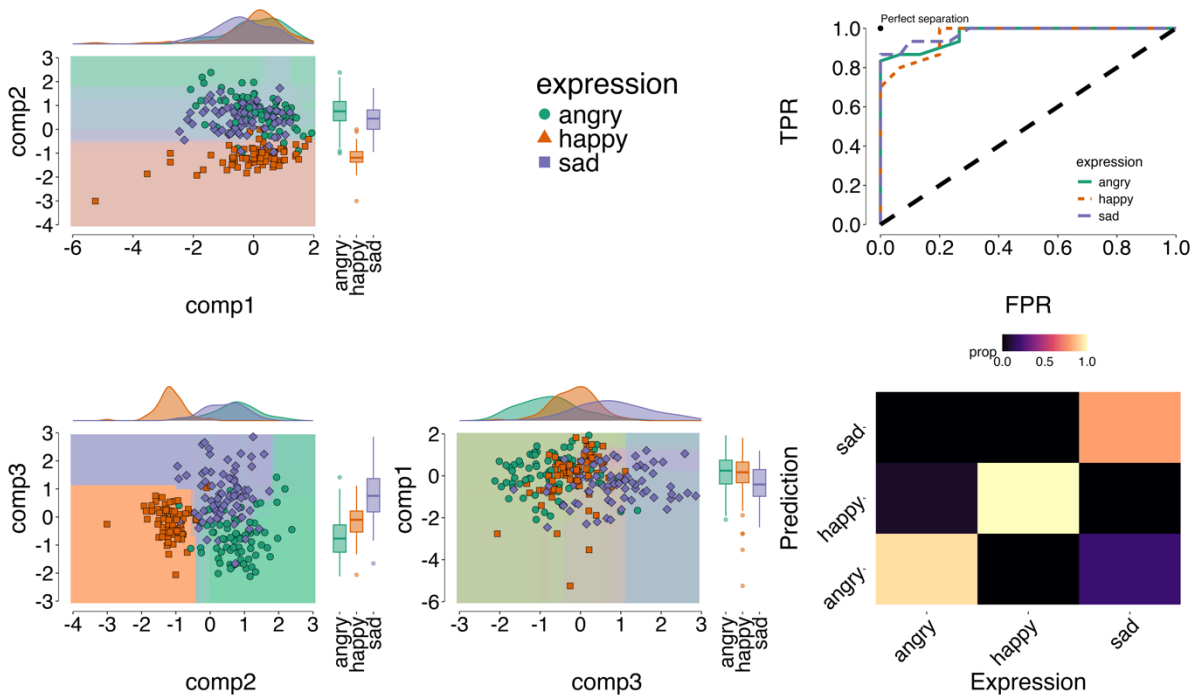

Figure 2. Spatiotemporal classification of facial expressions only production. *Notes:* The left panels show decision boundaries (filled backgrounds) across spatiotemporal components (Comp1: upper AUs, comp2: lower face AUs; comp3: lower to upper face AUs) estimated on a training set to predict emotion classes (angry, happy, sad). Coloured dots represent individual ground-truth expressions from both training and held-out test data. Dots within a matching background indicate correct classifications, while mismatches indicate misclassifications. The results demonstrate robust classification, with at least one component combination separating an expression from the rest. e.g. comp1 vs 2 and comp-2 vs 3 robustly distinguishes happy expressions. The top right panel presents Receiver Observer Characteristics (ROC) curves, showing good model performance (higher True Positive Rate [TPR] and lower False Positive Rate [FPR]) with the diagonal dashed line representing chance). The bottom right panel displays a confusion matrix. Brighter colours indicate higher accuracy along the diagonal elements, with off-diagonal showing little to no confusions (darker colours).

We also found distinct temporal profiles for each emotion. Angry expressions (BalACC = .91, AUC=.97, $p < .001$) were characterised by strong temporal coactivation of C1 (upper face AUS) and C2 (lower face AUS). Happy expressions (BalACC = .98, AUC = .99, $p < .001$) were dominated by C2 (lower face AUs); and sad (BalACC = .98, AUC = .90, $p<.001$) characterised by moderate changes in all three components. To further validate our model, we conducted a Fisher's exact test on the confusion matrix (see

Methods), which confirmed significant pairwise differentiation of all expressions on the basis of their spatiotemporal patterns (angry vs happy; angry vs sad; and happy vs sad: p.corrected < .001). These results reinforce the robustness of our classification model, demonstrating that spatiotemporal patterns differ between emotional expressions.

Spatiotemporal patterns for *Emotive speech* facial expressions

Similar to the *Expression only* condition, we found that three components effectively represented the spatiotemporal patterns of facial AUs in *Emotive speech* expressions (see SI1 – Fig S4-6). However, there were some notable differences in the spatial and temporal patterns, which likely reflect the fact that, in contrast with the *Expression only* videos, the *Emotive speech* videos comprise a combination of facial expression and speech AUs. Specifically, the *Emotive speech* components comprised a mix of both lower and upper face AUs, in comparison to mostly lower vs upper face AUs for *Expression only* conditions (see Fig 3).

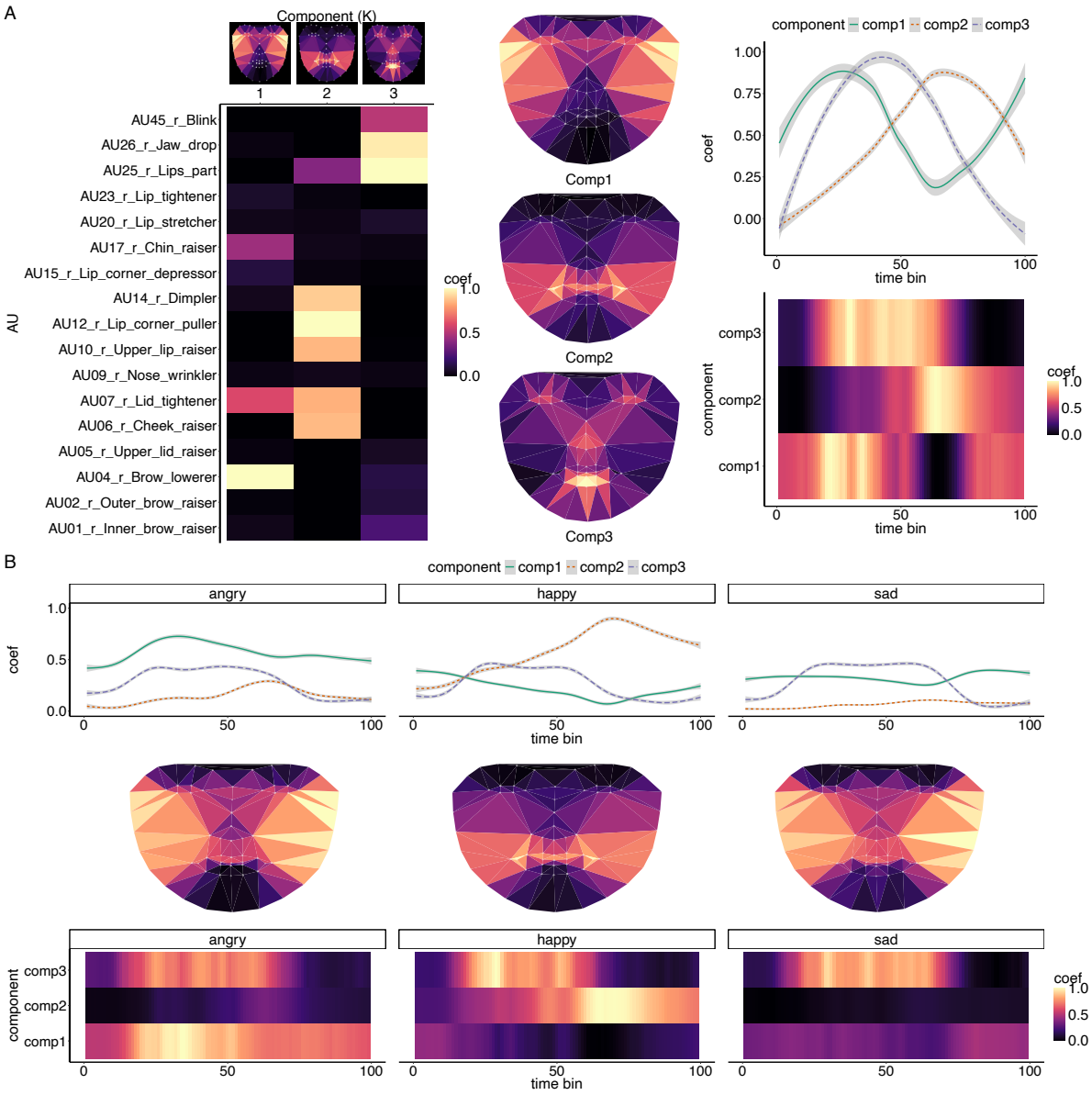

Figure 3. Spatiotemporal components for *Emotive speech* facial expression production. The top panel (A) shows that *Emotive speech* facial expression production can be summarised by three underlying components, capturing synergies between upper to lower face AUs (Component 1), lower-mid-upper AUs (component 2) and lower to upper face AUs (component 3). Top - middle and right columns show approximate spatial distribution of components in an average face model and how it changes over time, respectively. The bottom panels (B) show how components change over time, highlighting different profiles for angry, happy and sad expressions. Coef = coefficient from the NMF, brighter colours indicate stronger activation. Note while heatmaps are drawn to scale, face maps are not a one-to-one map and are drawn for visualisation purposes only.

Component 1 included strong activation of upper and lower face Aus (e.g., AU4 - brow lowerer; AU17 - chin raiser) peaking at both early and late time points. Component 2

(lower-mid-upper face AUs) combined strong activation of AUs in the lower (e.g., AU25 - lips part), middle (e.g., AU14 - dimpler) and upper face areas (e.g., AU7 - lid tightener). Component 3 primarily featured activation of lower face AUs, such as AU25 (lips part) and AU26 (jaw drop), plus additional moderate activation for AU 45 (blink). Henceforth component 3 is referred to as lower to upper face AUs (see Fig 3). C2 (lower-mid-upper face AUS) and C3 (lower to upper face AUs) had an inverted-U shape with C3 peaking earlier than C2 (Fig 3C).

As with facial expression-only production, *Emotive speech* expressions were also characterised by distinct spatiotemporal patterns. That is, a classifier trained on timeseries-based features of the spatiotemporal components (e.g., curvature, entropy) achieved significant accuracy in classifying expressions on a held-out test set, with performance at 70% (ACC = 0.69, 95% CI: 0.53, 0.81; BalAcc = .76 Kappa = 0.53, $p < 0.001$). These results indicate moderate to strong agreement between prediction and ground-truth emotional expression labels.

Class-wise, the spatiotemporal components significantly distinguished all expressions. Happy had the highest accuracy (BalACC = 0.91, AUC = 0.97, $p < .001$), suggesting a clear and distinct spatiotemporal profile, notable in the increased activation in C2 (lower-mid-upper face AUs) over time (see Fig 4).

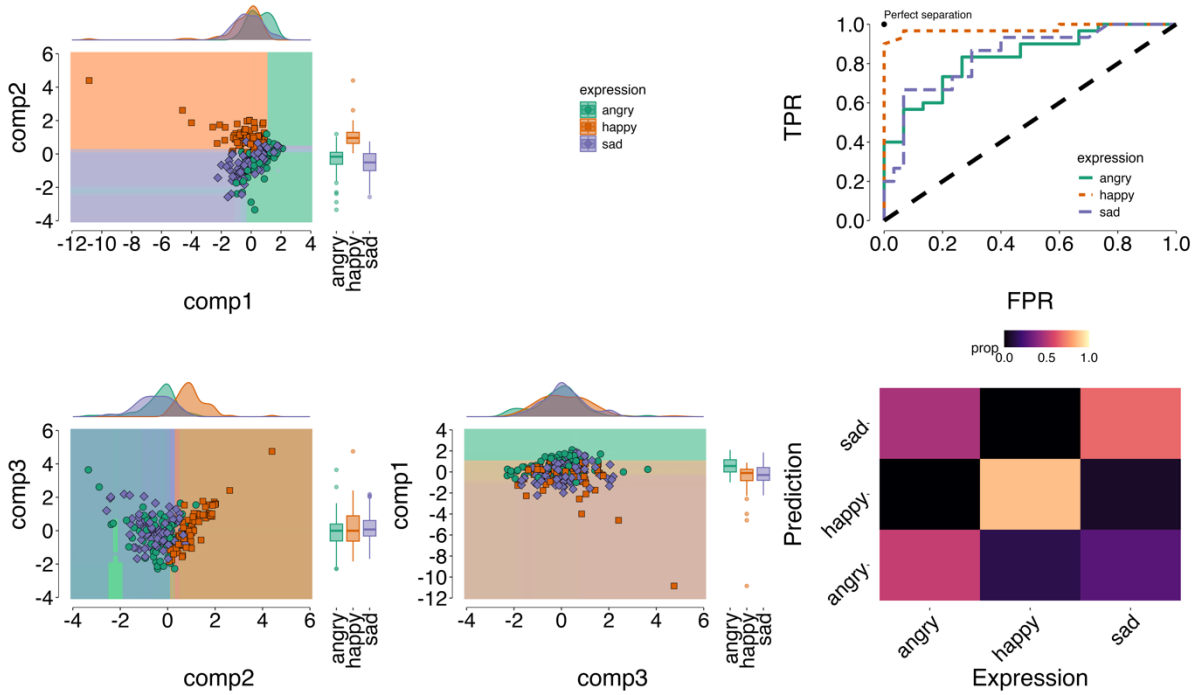

Figure 4. Spatiotemporal classification of *Emotive speech* facial expressions. The left panels show decision boundaries (filled backgrounds) across spatiotemporal components (comp1: upper to lower face AUs, comp2: lower to upper face AUs; comp3: upper-to-lower face AUs) estimated on a training set to predict emotion classes (angry, happy, sad). Coloured dots represent individual ground-truth expressions from both training and unseen test data. Dots within a matching background indicate correct classifications, while mismatches indicate misclassifications. The results demonstrate robust classification, with component combinations effectively separating the expressions in at least one combination.
The top right panel presents ROC curves, showing good model performance (higher True Positive Rate [TPR] and lower False Positive Rate [FPR]), with the diagonal dashed line representing chance. The bottom right panel displays a confusion matrix. Brighter colours indicate higher accuracy along the diagonal elements, and confusion off-diagonal (e.g., happy shows high accuracy and low confusion, whereas angry shows higher confusion than all expressions).

Angry had the lowest classification accuracy compared to all other expressions, yet still significantly above chance (BalACC = 0.67, AUC = 0.83, p = .01), reflecting a more complex and less distinct pattern across all components. Similar to happy, sad expressions were classified reliably above chance (BalACC = 0.72, AUC = 0.84, p = 0.001). Fisher's exact test on the confusion matrix confirmed significant pairwise differentiation between happy and sad expressions (p < .001) and between angry and

happy ($p < .001$). However, it suggested confusions between and angry and sad ($p = 0.12$) (see Fig 4; SI1 - Table S3-4).

In summary, we found three latent spatiotemporal components that reliably described the dynamics of facial expressions during *Expression-only* and *Emotive speech* conditions. Robust train-test classification on unseen held-out data, indicates that spatiotemporal components carry information that is diagnostic of emotional content. While we reported results for *Expression only* and *Emotive speech* expressions separately for simplicity, these patterns were consistent when analysed together.

Facial expression substate results

The second objective of this study was to identify and characterise facial expression substates, such as transition and sustain periods, and their diagnostic value for emotion signalling (i.e., whether they differ between emotions: happy, angry and sad expressions). We identified three primary substates through spatiotemporal clustering that segmented periods with robust local dynamics in the spatiotemporal components (see Methods: *Spatiotemporal clustering for substates analyses*). Visual representation of the kinematic profiles of each substate (Fig 5A-B) suggest the substates correspond to relaxed, transition, and sustain periods for spatiotemporal components. The overall pattern of substates is also consistent across participants (see Fig 5 C-D).

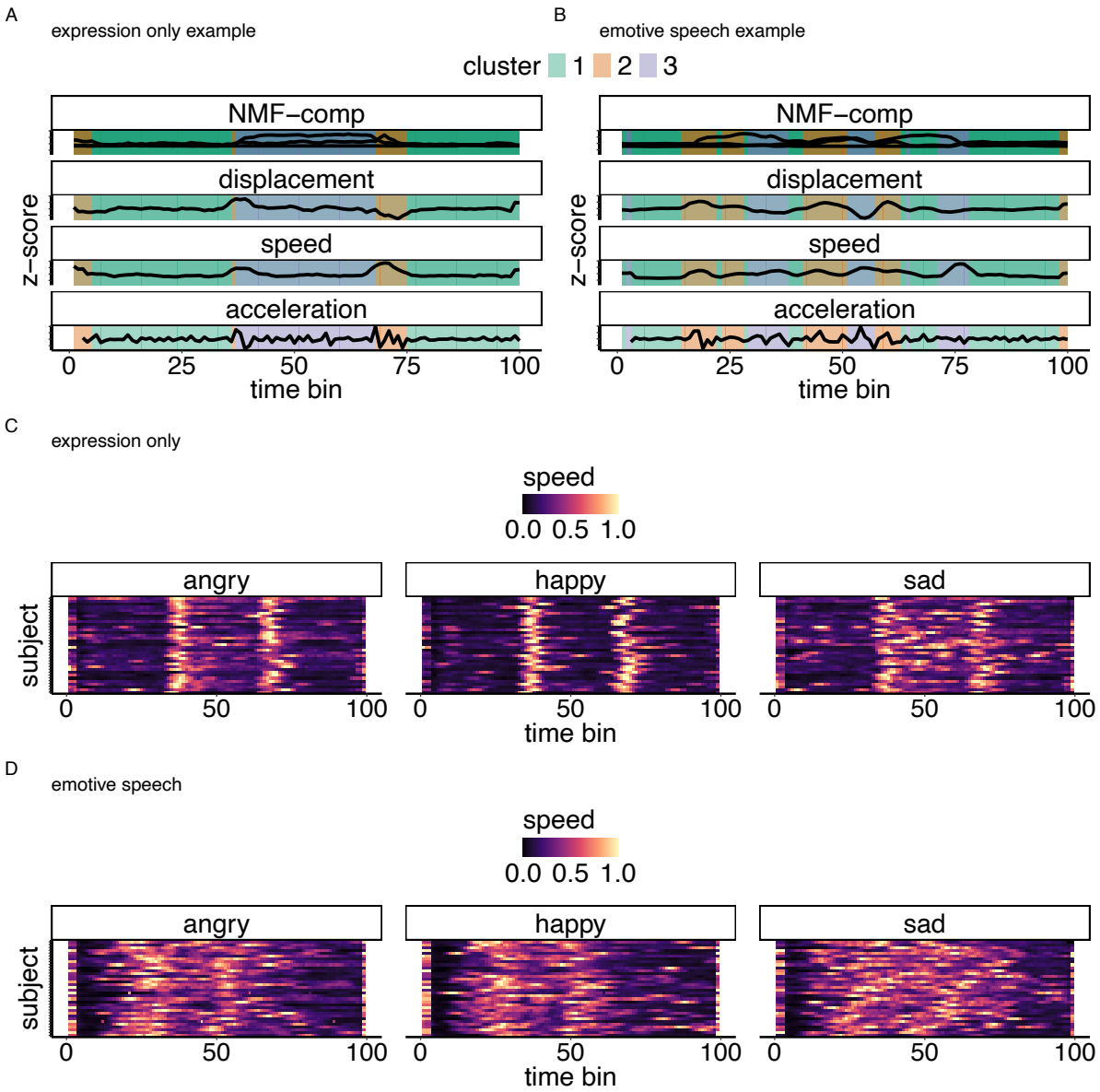

Figure 5. Illustration of facial expression substate profiles.
A-B. Example of substate clusters (1 (green) = relaxed, 2 (peach) = transition; 3 (purple) = sustain). C-D show overall the speed distribution for *Expression only* and *Emotive speech* expressions for all subjects.

To compare whether the pattern of facial expression substates differs between emotions (anger, sadness and happy) and conditions (*Expression only vs Emotive speech*), we computed two metrics: (1) speed within the substate hence providing a kinematic characterisation of each substate, and (2) entropy, which provides a view of overall dynamics of substates (e.g., how they change over time) (Fig 5) – see also

Methods. Similar measures have been shown to capture diagnostic information about facial expression movement patterns [8,11].

Transition substate speeds differentiate emotional expressions

A mixed-effects model predicting speed from emotion (happy, angry, sad), condition (*Expression only, Emotive speech*), and substate (relaxed, transition, sustain) showed strong total explanatory power (conditional R2 = 0.6 (66%)) with the fixed effects alone (marginal R2) accounting for 0.63 (63%) of the variance. All main effects were significant. First, there was a small significant effect of emotion (F(2,1990) = 12.89, p < .001; η_p^2= 0.02, 95% CI [0.009, 0.10]). Pairwise tests showed that angry expressions were significantly slower than happy expressions (angry < happy; beta = -0.028, SE = 0.008, t(1207) = -3.398, p = 0.002). Happy expressions were faster than sad expressions (happy > sad; beta = 0.041, SE = 0.008, t(1207) = 4.92, p < 0.001). There were no significant differences in speed between angry and sad expressions (beta = 0.012, SE = 0.008, t(1207) = 1.53, p.corrected = 0.38). There was a significant main effect of condition (F(1, 1190) = 369.41, p < .001; η_p^2 = 0.24, 95% CI [0.20, 1.00]) with *Expression only* condition moving faster than spoken (beta = 0.13, SE = 0.007, t(1207) = 19.08, p < .001). Finally, there was a main effect of substate (F(2, 1190) = 687.90, p < .001; η_p^2 = 0.54, 95% CI [0.51, 1.00]) where sustain had higher speed than relaxed (beta = -0.16, SE = 0.008, t(1207) = -19.52, p.corrected < .001); transition had higher speed than relaxed (beta = -0.307, SE = 0.008, t(1207) = -36.80, p.corrected < .001) and transition had higher speed than sustain (beta = -0.144, SE = 0.008, t(1207) = -17.28, p.corrected < .001).

All three two-way interactions were significant. First, there was a significant interaction between emotion and substate ($F(4, 1190) = 15.24$, $p < .001$; $\eta_p^2 = 0.05$, 95% CI [0.03, 1.00]). Pairwise tests showed that differences in speed between happy, angry and sad emotional expressions were statistically significant for the transition substate only (happy>anger, beta = -0.077 SE = 0.014, t(1207) = -5.36, p.corrected < .001; sad< angry, beta = 0.053, SE = 0.014, t(1207) = 3.64, p.corrected < 0.001; sad< happy, beta = 0.130, SE = 0.014, t(1207) = 9.005, p.corrected < .001). There were no speed differences between emotions for the relaxed substate (angry vs. happy, beta = -0.005, SE = 0.014, t(1207) = -0.348, p.corrected = 1.0; angry vs. sad, beta = 0.007, SE = 0.014, t(1207) = 0.498, p.corrected = 1.0; happy vs. sad, beta = 0.012, SE = 0.014, t(1207) = 0.846, p.corrected = 1.0) nor the sustain substate (angry vs. happy, beta = -0.003, SE = 0.014, t(1207) = -0.177, p.corrected = 1.0; angry vs. sad, beta = -0.022, SE = 0.014, t(1207) = -1.501, p = 0.40, p.corrected; happy vs. sad, beta = -0.01913, SE = 0.0144, t(1207) = -1.325, p.corrected = 0.5567). These results suggest that transition substate kinematics (but not relaxation or sustainment kinematics) carry diagnostic information about the emotional content of facial expressions (Fig 6).

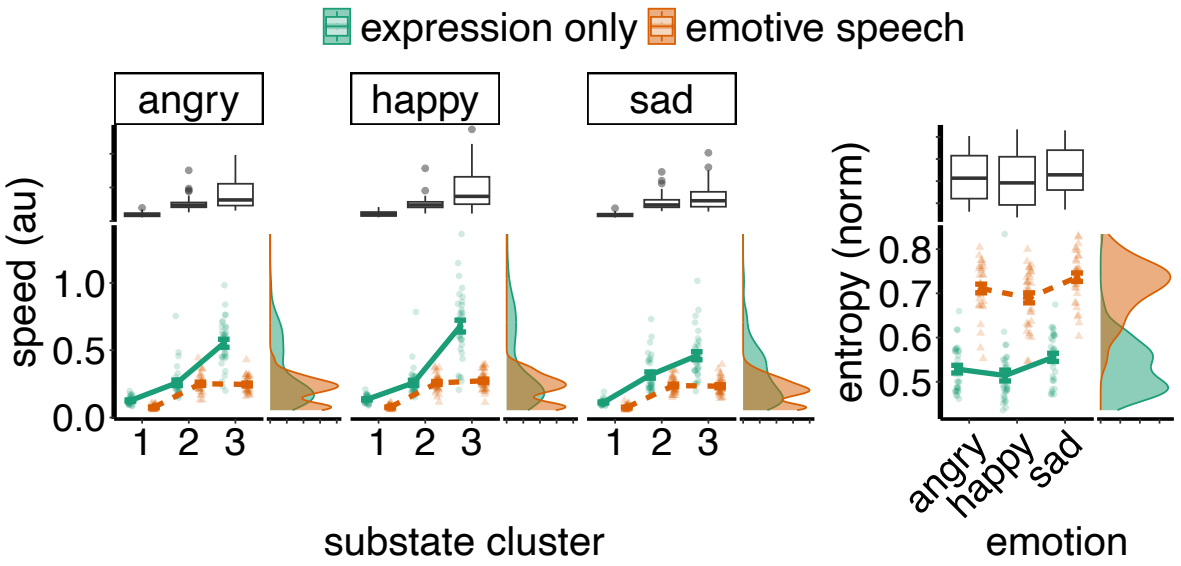

Figure 6: Summary of speed and entropy of substate patterns.

Substate cluster: 1-relaxed, 2-sustained, 3-transition. *Expression only* condition has faster than *Emotive speech*, with transition substates being faster; Entropy mainly differentiates *Expression only* and vs. *Emotive speech* condition, with the latter exhibiting more complex patterns. Entropy also differed between happy and all other expressions, with substate patterns being more predictable for former.

In addition, there was a significant two-way interaction between emotion and condition ($F(2) = 3.54$, $p = 0.029$; $\eta_p^2 = 0.006$, 95% CI [3.15e-04, 1.00]). For the *Expression only* condition, there were significant speed differences between two out of the three emotion expression pairs: specifically, happy > angry (beta = -0.045, SE = 0.012, t(1207) = -3.84, p.corrected < 0.001); happy > sad (beta = 0.06171, SE = 0.0118, t(1207) = 5.232, p.corrected < .001); and no differences between angry vs. sad (beta = 0.016, SE = 0.012, t(1207) = 1.39, p.corrected = 0.492). In contrast, for the *Emotive speech* condition, there were no significant differences in speed between the emotional expressions (angry versus happy, beta = -0.011, SE = 0.012, t(1207) = -0.97, p.corrected = 1.0; angry versus sad, beta = 0.009, SE = 0.012, t(1207) = 0.77, p.corrected = 1.0; happy versus sad, beta = 0.020, SE = 0.012, t(1207) = 1.73, p.corrected = 0.25).

Finally, there was a significant two-way interaction between condition and substate ($F(2) = 183.65$, $p < .001$; $\eta_p^2 = 0.24$, 95% CI [0.20, 1.00]). For the *Expression only* condition, there were significant differences in speed between all substates such that sustain was faster than relaxed (relaxed < sustain, beta = -0.153, SE = 0.012, t(1207) = -12.96, p.corrected < .001); transition was faster than relaxed (relaxed < transition, beta = -0.439, SE = 0.012, t(1207) = -37.23, p.corrected < .001), and faster than sustain (beta = -0.286, SE = 0.0118, t(1207) = -24.271, p.corrected < .001). These patterns were overall similar for the *Emotive speech* condition where relaxed periods were significantly slower than sustain (beta = -0.173, SE = 0.0118, t(1207) = -14.647,

p.corrected < .001) and transition periods (beta = -0.175, SE = 0.012, t(1207) = -14.82, p < .001). However, there was no statistically significant difference between sustain and transition speeds (beta = -0.002, SE = 0.012, t(1207) = -0.169, p.corrected = 1.00). In sum, transition substates, carry speed information that differs between expressions and also between *Expression only* and *Emotive speech* conditions.

Substate dynamics differentiate emotion content

Entropy is an information-theoretic measure of complexity, directly related to state dynamics within a system [36]. Here entropy is used as a metric to statistically describe the dynamics of facial expression substates in terms of how temporally structured/predictable they are (see Methods). High entropy therefore equates with high complexity in substate patterns, and low entropy with low complexity. Fig 5, for example, illustrates that substate transitions are more complex (i.e., high entropy, less predictable) in the *Emotive speech* condition where visual pattern differences between emotions are harder to discern than in the *Expression only* condition (see Methods: *Spatiotemporal clustering for substates analyses*).

The entropy mixed model explained 67% of the variance ($R^2 = 0.67$). The fixed effects alone (emotion, and expression condition) accounted for 57% of the variance (marginal $R^2 = 0.57$). The main effect of emotion was statistically significant ($F(2, 340) = 12.27$, $p < 0.001$; $\eta p^2 = 0.07$, 95% CI [0.03, 1.00]). Pairwise contrasts revealed that sad expressions had a significantly higher entropy (more unpredictable substate sequences) compared to happy expressions (estimate = -0.027, SE = 0.0055, t(344) = -4.897, p < 0.0001). There were no statistically significant differences in entropy between angry and happy expressions (estimate = 0.011, SE = 0.006, t(344) = 1.996,

$p = 0.14$), but there was a significant difference between angry and sad expressions (estimate = -0.016, SE = 0.0055, $t(344) = -2.90$, $p = 0.012$).

The largest effect was the main effect of condition ($F(1, 34) = 429.10$, $p < 0.001$; $\eta p^2 = 0.93$, 95% CI [0.89, 1.00]), with the *Expression only* condition demonstrating significantly lower entropy than *Emotive speech* expressions ($\beta = -0.208$, SE = 0.017, $t(35) = -12.594$, $p < 0.001$). This suggests more complex movement dynamics in the *Emotive speech* compared to the *Expression only* condition (see heatmaps in Fig 5). There was no significant interaction between emotion and condition ($F(2, 340) = 0.03$, $p = 0.970$; $\eta p^2 = 0.0002$, 95% CI [0.00, 1.00]). In summary, *Expression only* substates are more temporaly structured than those in *Emotive speech* expressions. Nonetheless, substates patterns differentiate expressions with sad expressions having a more complex patern of substates than happy and sad expressions.

To summarise, in addition to the low-dimensional patterns of spatiotemporal components that are diagostic of emotion and expression condition, we also identified three substates reflecting periods of relaxation, transition, and sustainment of facial actions. These substates are distinguished by their unique kinematic and dynamic (complexity) profiles which differ between emotional expressions and *Expression only* versus *Emotive speech* conditions.

Discussion

The current study explored how the underlying dynamics of facial movements relate to their emotional signalling function. We employed a motor-theoretic inspired approach to characterise the spatiotemporal structure of facial movements in

Expression only and *Emotive speech* conditions. Specifically, we used data-driven spatiotemporal dimensionality reduction to identify spatiotemporal components that characterise how the activation of different action units evolves and combines over time. We consistently identified a small number of components that reliably described the time courses and covariance patterns of over a dozen action units. The identification of these components suggests that, despite the apparent complexity of facial expressions, their emotion signalling function is achieved by just a few fundamental dynamic patterns; combinations of these dynamic patterns differentiate angry, happy and sad expressions. Second, we showed that these spatiotemporal patterns contain meaningful substates, which reflect periods of relaxation, transition, and sustainment of expressions. The kinematic profiles of these substates differed by emotion category as well as between *Expression only* and *Emotive speech* conditions.

The low dimensionality of the spatiotemporal patterns of facial expressions, that we have uncovered, is intriguing given the various combinations of action units that can contribute to the same expressions [3] as well as the variability in how people produce expressions [37]. A low-dimensional spatiotemporal structure may reflect flexibility for signalling, as a variety of expressions can be rapidly produced by relying on just a few interchangeable dynamic patterns. Such flexibility is likely beneficial in face-to-face interaction, where facial expressions may need to be quickly adapted according to interaction demands [3,38] or where it is necessary to accentuate emotional information during speech [4,39]. Interestingly, our findings are consistent with the observation of low-dimensional spatiotemporal patterns in other body movements [17,19,21] which also have the capacity to signal emotion information [24,40,41]. Thus,

we extend existing findings from the bodily motor control literature to encompass facial expression movement.

An important question concerns whether our results are representative of facial expression signalling in general, or simply an outcome of the particular methodology we employed. For example, it could be argued that we identified a small number of spatiotemporal components because we asked participants to demonstrate relatively simplistic expressions which can be signalled by activating just a few action units. We argue that this is not the case because a) the *Emotive speech* condition was specifically designed to dissuade participants from relying on overly simple and caricatured expressions, and b) visualisations of the non-reduced patterns of facial expressions, as well as their reconstruction from the components show variability in the AU patterns. It could also be argued that our results are simply a consequence of dimensionality reduction, as any set of data can in theory be represented by a lower-dimensional structure. Yet, as we showed in a series of validation and sensitivity analyses, the components learned carry diagnostic information about emotions, which is robust to comparisons against appropriate randomised dynamic patterns and generalises to unseen data. Thus, we have not generated a “non-sense” low dimensional structure but rather a meaningful way to represent the data.

Here we provide evidence for a low-dimensional spatiotemporal structure for facial expressions of emotion. An intriguing question concerns why this low-dimensional structure exists. One possibility is that receivers only attend to a reduced subset of spatiotemporal features and so, correspondingly, signallers learn to only move a reduced subset of features, or to move them in specific ways. In support of this, studies

have found that even when observers are presented with dozens of biologically plausible face movements, their perceptual representations suggest that they attend to only a reduced subset [35]. Since human infants learn to produce and recognise facial expressions throughout a protracted period of development [42], it is possible that humans learn to optimise signals based on how these are optimally perceived and responded to, creating a reciprocal dynamic interaction between these processes. Thus, the low-dimensional organisation of facial expressions may reflect the optimisation of social signals according to perceptual demands and these perceptual demands, themselves, may be influenced by the low-dimensional organisation of facial expressions [43]. An alternative possibility is that there are savings (e.g., energy/computational savings) to be had from moving the face according to a few spatiotemporal patterns. Future research that integrates both perception and production of facial expressions, particularly in interactive contexts, will be crucial for elucidating the bidirectional influences.

The identification of substates in spatiotemporal patterns of facial expressions adds a nuanced view of the dynamics of facial expression production. On the one hand substates likely reflect the biomechanics of producing facial expressions, which inevitably result in periods of relaxation, transition and sustainment of facial motor actions [44]. Yet, the fact that the patterns of these substates (e.g., speed, entropy) vary with different emotions and expressive conditions, suggests that the local dynamics are sensitive to the communication of different emotions.

In real-life interaction facial movements subserving verbal and nonverbal expressions are often seamlessly integrated. Our findings in the *Emotive speech* condition raise

533 the possibility that facial movements underlying verbal and non-verbal emotion
534 information are integrated by modulating different spatiotemporal components. For
535 example, during speech we can modulate the lip movement to produce the word while
536 accentuating movements such as open mouth and eyebrow raise to signal emotion.
537 Indeed, the observation of mostly part-based profiles (e.g., lower vs upper face AUs)
538 in *Expression only* conditions, contrasting the involvement of both mouth and non-
539 mouth-related components in *Emotive speech* expressions supports this assertion.
540
541 Despite the robustness of our results, there are three important caveats that need to
542 be addressed. First is that the finding of low-dimensional spatiotemporal structure for
543 emotion signalling does not imply that *three* spatiotemporal components are an
544 appropriate representation of *all* possible facial expressions. The discretisation of
545 movement using FACS with more action units than those used in this study, may
546 require more dimensions, as might different emotional expressions (e.g., surprise,
547 fear, disgust). Furthermore, while FACS AUs are anatomically based robust
548 descriptors of facial behaviour, facial movements for speech are more complex and
549 not exhaustively described by AUs. Future studies should explore different coding
550 systems (e.g., geometrical/landmark-based approaches) and include systems
551 optimised for the complexities of speech. Second, a valid question concerns whether
552 we would have found more/less substates had we employed a different methodology.
553 However the number of substates and the degree to which they are an optimal
554 representation of dynamics depends on the level of spatial and temporal resolution in
555 the underlying data and that which is desired in the analysis. A different number of
556 substates might be discovered using physiological approaches, such as
557 electromyography (EMG). EMG is sensitive to subtle motor potentials and therefore

may capture more substates including preparatory or corrective movement impulses. Lastly, our approach to quantifying spatiotemporal structure of facial behaviour focuses on exhaustively describing spatiotemporal features that account for variance in the signal such as curvature patterns, entropy, or others. While our method highlights that the underlying dynamic patterns are robust and diagnostic of emotion expression, it does not necessarily isolate specific features (or groups) nor does it compare their individual diagnostic value in classifying emotion. While such analysis is beyond the scope of this study, it can be easily integrated with our approach.

Despite these caveats, our results and approach contribute new insights for future research on facial expression signalling. Spatiotemporal structure and substates might be important to better understand the mechanisms underlying differences in social communication. This may provide useful insight into clinical conditions such as autism and depression. While emotion recognition differences have typically been focused upon with respect to these conditions, an underexplored dimension is the potential role of atypical facial expression production in contributing to social interaction difficulties [45,46]. A compelling extension of this work would therefore investigate whether conditions such as autism and depression are characterised by unique spatiotemporal patterns and substates during social signalling. The insights gained from the current study also have practical applications in fields such as social robotics and human-computer interaction. A significant challenge in these fields relates to endowing social agents (robots, digital agents) with convincing and adaptable repertoires of social behaviour, including facial signalling [47]. Spatiotemporal components and substates underlying facial expression production provide actionable targets to enhance the design of flexible realistic and expressive agents without the

need to painstakingly craft each signal feature into the architecture of these systems. Similarly, the incorporation of spatiotemporal cues and substates could significantly aid understanding how spatiotemporal structure influences emotion recognition, in both humans and machine recognition systems.

Conclusion

In conclusion, we revealed a low-dimensional spatiotemporal structure with distinct substates that underpin facial expression production in *Expression only* and *Emotive speech* conditions. We provide a methodological framework to advance the development of novel theoretical and experimental accounts of the structure and function of facial expression dynamics. This framework can be extended to other modalities and has practical implications for understanding emotional signalling and perception in human and machine systems.

Materials and Methods

Participants

43 healthy volunteers from the university community (24 female; mean age = 26), took part in a study for course credit or monetary incentive. All participants gave informed consent, and all procedures of the study were approved by the institutional ethics committee. The dataset in this study was collected as part of a larger project investigating facial expression production kinematics and perception – see [11]. However, all the variables and analyses reported here are novel and specific to this study and not discussed in previous papers.

608 **Experimental procedure**

609 In two separate sessions, across two days, participants' facial movements were
610 captured during a facial expression production task. Participants were asked to
611 produce three emotion expressions: anger, happiness, and sadness in two
612 counterbalanced conditions. An *Expression only* condition where participants simply
613 produced facial movements to signal each expression, and an *Emotive speech*
614 condition where they produce a fixed spoken sentence in emotional fashion (e.g., "My
615 name is John and I'm a scientist"). The conditions are motivated by the fact that in
616 social interaction, facial expressions can be signalled both in isolation and embedded
617 within *Emotive speech*, meaning that facial movements might signal information
618 relevant to emotion and speech simultaneously, yet how this is achieved in unclear.

619

620 Facial movements were recorded with participants seated with their head at the centre
621 of a polystyrene frame to ensure central positioning towards the camera, positioned 1
622 meter from a tripod-mounted video camcorder, sampling at 60Hz f/s (Sony Handicap
623 HDR-CX240E). The start and end points were signalled using an audible beep, which
624 allowed us to align the recordings of facial expressions. See the schematic illustration
625 in Fig 7.

626

627 **Analytical approach**

628 *Data processing.* Facial expression videos were processed using OpenFace [48], an
629 open-source suite for automated face-tracking, landmarking and facial action unit
630 detection based on the facial action coding system (FACS) [2]. We focused on Action
631 Units (AUs), which code for the observable presence and intensity of basic face
632 actions (e.g., blinking, mouth widening, eyebrow-raising) that can be used to produce

almost any facial expression. We estimated the temporal activation weights for 18 AUs. Only videos with a tracking accuracy of over 90% and with participant data across two sessions were used, resulting in a total of 532 videos with approximately 240 time points each. Based on visual inspection of the movement profiles, we applied a moving average (window = 3 frames) to smooth out the slight jitter inherent in face tracking and AU estimation, while preserving the underlying trends and dynamic range. To ensure an equal temporal window for spatiotemporal analyses, AU timeseries were reduced into 100-time bins for each video. Binning was preferred over more computationally expensive alternatives such as dynamic time-warping (DTW) for practicality since the overall timeseries duration was very similar across videos, and visual comparisons between binning and DTW tests on our data showed negligible differences.

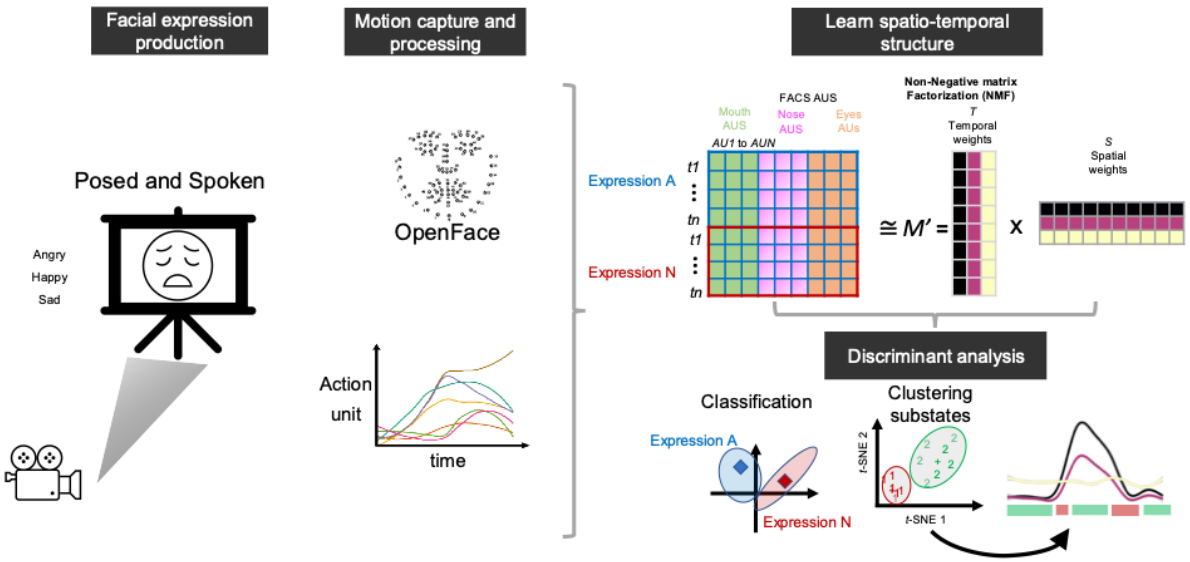

Figure 7. Schematic illustration of methodological and analytical approach. Left panel illustrates the facial expression production task, video capture and OpenFace processing for landmark and action unit (AU) estimation. The right panel summarises the procedure for the concatenation of AU timeseries data and the application of Non-Negative Matrix Factorization (NMF) to learn an underlying low-dimensional spatiotemporal representation of the data. The accuracy and effectiveness of this lower-dimensional representation were validated through a series of sensitivity and

classification analysis (lower right panel). Time series clustering was used to identify distinct substates or 'phases' in facial movement components.

Spatiotemporal analysis

Timeseries of AU weights for the full facial expression production task were analysed using Non-Negative Matrix Factorization (NMF) [34]. NMF is a multivariate dimensionality reduction method that decomposes a high-dimensional dataset into two non-negative matrices that form a part-based representation of the data (i.e., the original data is achieved by combining/adding components). The non-negativity constraint is particularly well suited for representing movement and action unit data, which is typically positive, and the two-matrix part-based decomposition is well suited to represent spatial vs time-varying aspects in the data – see Fig 7 (see also: [18,34,35]). The NMF takes the concatenated AU timeseries as input, denoted by a matrix ***M,*** which is a product of two low-dimensional matrices (***T*** and ***S***), ***T*** represents the temporal weights (basis values for time bins), and ***S*** represents the spatial weights (face action activation weights). The product of ***T*** *and* ***S*** matrices denoted by ***M'*** represents an approximation of the full original matrix such that ***M~M' = T * S.*** The algorithm works by initialising the values for ***T*** and ***S*** (e.g., with random values) and using an iterative multiplicative update rule to sequentially update *T* and *S* by a factor that minimises the approximation error (i.e., residuals). While the iterative nature of NMF means that the solution is non-deterministic (i.e., you can learn multiple equally effective part-based representations), these update rules have been shown to guarantee convergence (i.e., stability of the cost function) after a relatively low number of iterations [49]. The optimal number of spatiotemporal components (*k*) was assessed computationally by fitting a range of values (*k* = 2 to 6) and choosing a *k* that optimised various model fit metrics (e.g., residual sum of squares, silhouette for clustering

680 cohesion, cophenetic coefficient, see below). We also evaluated the final solution
681 against NMF results on a permutated dataset, using block-based time shuffling. This
682 resulted in an appropriate null dataset that destroyed the inherent spatiotemporal
683 pattern in the data but retained some temporal ordering. We then compared fit metrics
684 between original and permutated NMF (see also *Classification and clustering*
685 *analysis*). We used the NMF decomposition functions in the NMF package for R [50].
686 To visualise our NMF results in an intuitive manner, we trained a custom model to
687 predict aligned Facial Landmark displacements (after correcting rotation, translation
688 and looming) from AUs, based on original data using partial least squares regression.
689 Note however that the AU to Landmark face model is not a perfect one to one map
690 between NMF results and visualisation, given the many-to-one relationship between
691 facial landmarks and AUs.
692
693 *Multivariate timeseries classification*
694 We used a supervised classification approach to validate the diagnostic value of
695 spatiotemporal components for emotion expression. This involved in five steps.
696 **(1) Data Splitting**: We began by splitting the original pre-NMF data into training (80%)
697 and testing (20%) sets, ensuring that each set contained complete AU time series for
698 each video. This split was performed before any further analyses, and all subsequent
699 steps maintained this separation ensuring further processing and model training was
700 only done on the training set, and then applying results to the test set. Hence any
701 mention of “training”, always refers to operations on the respective training set. This
702 approach was taken to prevent information leakage between the training and test sets,
703 ensuring that the classification was conducted on truly unseen data [51,52].

704 (**2) NMF re-training and projection:** We then re-trained the NMF model using only
705 the training set data and projected the test set onto the trained NMF model space to
706 ensure consistency of components for classification.
707 (**3) Feature extraction:** Because standard classification approaches do not account
708 for inherent spatiotemporal structure of the NMF components, we opted for an
709 approach that represents the spatiotemporal structure via interpretable theoretic time
710 series features (e.g., curvature, entropy, predictability, peak, trend) [53] consistent with
711 features suggested to capture diagnostic information for classification of facial
712 expressions and body movements [8,11,54,55]. Feature extraction was done
713 separately for train and test sets. Then, a Principal Component Analysis (PCA) was
714 used to reduce the features that explained the majority of the variance in the train set
715 while retaining a number of dimensions equal to the number of NMF components for
716 consistency. The test set features were then transformed using the PCA model fitted
717 on the training set features to be used in the next step. The component scores were
718 used as inputs for classification analysis in the next step.
719 **(4) Classification training and testing.** We the trained a classifier on spatiotemporal
720 component features derived from step 3 and applied to predict emotion expression in
721 a held-out test set. Specifically we used a Random Forests (RF) Classifier - a
722 supervised learning method that learns a discriminant model between a set of input
723 variables (e.g., spatiotemporal components) and a set of target classes (expressions).
724 Note, that we achieved similar results with different methods (e.g., Support Vector
725 Machine, Linear Discriminant Analysis), thus our results are not a function of the
726 analytical approach. The Random Forests Classifier has the advantage of being
727 relatively more robust to overfitting and able to capture non-linear relationships and
728 therefore was preferred. We used cross-validation during training, where multiple

729 training sets are created, and the data not included in the test set is used to test the
730 accuracy and tune the model parameters. Reliable classification on the test set would
731 indicate that the underlying spatiotemporal components (identified by the NMF
732 analysis) carry diagnostic information for differentiating expressions. We derived
733 significance tests for our classification results using binomial tests to determine
734 whether the model's accuracy on the test set was significantly greater than what would
735 be expected by random guessing (by comparing it to a no information rate). These
736 binomial p-values were compared with permutation-based p-values, which were
737 generated by comparing the observed accuracy to the distribution of accuracies
738 obtained from shuffling the labels (expressions) in the test set (N = 1000). The
739 negligible differences between the binomial and permutation-based p-values for our
740 dataset validated the robustness of our results, so we primarily report the binomial p-
741 values. Additionally, we used Fisher's Exact Test to assess the statistical significance
742 of the confusion matrix, by comparing pairwise expression differentiation with
743 adjustments for multiple comparisons. A low p-value in Fisher's exact test indicates
744 statistically significant differences between pairs, meaning the pattern in the confusion
745 matrix is unlikely to have occurred by chance.
746
747 *Spatiotemporal clustering for substates analyses*
748 We investigated whether reliable substates could be identified from the spatiotemporal
749 patterns of facial expressions. We considered substates as local intervals where the
750 components exhibit salient signatures that are consistent across expression
751 timeseries. Therefore, this challenge can be formulated as segmenting a multivariate
752 timeseries into finer temporal segments. To do this we first derived speed and
753 displacement for spatiotemporal components, as these features are sensitive to

segment characteristic of other body movements (e.g., swing and stance signatures in walking, fixation and saccades in eye movements) [19,28,29,56]. Second, we applied spatiotemporal clustering using the *dtwclust* package in R. This was motivated by the fact that, unlike other movements with well-established kinematic profiles, we have no *a priori* knowledge regarding the characteristics of facial expression substates. We tested cluster sizes of 2 to 5 based on visualisations which suggested (Fig 5) two to three interpretable substates that described transition and stability phases in spatiotemporal components of facial expressions. To be able to compare substates and whether they differentiate emotion, we derived measures to describe them. To this end, we computed two metrics: (1) speed within the substate and (2) entropy, both which have been shown to capture diagnostic information about facial expression movement patterns [8,11]. Specifically, visualisations indicated that speed profiles varied with substate clusters, therefore we used speed as measure to assess the characteristics of each substate (see Fig 5). Entropy, an information-theoretic measure of complexity, is directly related the dynamics of state transitions (or sequences) within a system [36]. In this context, entropy provides a suitable summary of the dynamics of substates in a single statistic, describing how complex or unpredictable the sequence of these substates is (e.g., how much information about the current substate can predict the next substate). This statistic aimed at quantifying, for example, the apparent differences in the temporal patterns of substates between expressions and conditions (see Fig 5). We computed entropy using the transition entropy function (for state change entropy) from the *GrpString* package for R [57] but for simplicity we refer it as entropy. To improve interpretability, entropy was normalised by dividing the computed entropy by the maximum theoretical entropy for a system with the same number of substates.

Both speed and entropy measures were computed for each individual expression and analysed using a mixed-effects model to test the effect of emotion (happy, angry, sad), condition (*Expression only*, *Emotive speech*), and substate (relaxed, transition, sustain) as fixed effects and subject (face id) as a random effect. Entropy was log transformed to correct for non-normal residuals. Linear mixed models were fitted using the lme4 package for R [58]. Linear mixed models were used as opposed to the classification approach due to the larger number of fixed effects and nested conditions which makes classification results overly complex.

Data availability statement

Supporting data is available in the Open Science Framework (OSF): https://osf.io/3nzk7/?view_only=db7d9467423f42b794385e549b6aa96a

All code for pre-processing and models is available at https://github.com/hcuve/spat-tmp-emo-faces

Acknowledgments

We would like to thank Professor Rachael Jack for her invaluable insights during the initial discussions on facial signalling dynamics and dimensionality reduction.

We would also like to thank Katherine Lane and Tahniyyat Bokhari, for their assistance with data collection for a portion of the study.

Author Contributions (CRediT)

HCC, Conceptualisation, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Software, Validation, Visualisation, Writing – Original Draft Preparation, Writing – Review & Editing

SSC, Methodology, Validation, Writing – Review & Editing, Resources

JC, Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing, Resources

Funding Statement

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under ERC-2017-STG Grant Agreement No 757583. HCJC was funded by an Experimental Psychology Society Postdoctoral Fellowship

References

1. Jack RE, Schyns PG. The human face as a dynamic tool for social communication. Curr Biol. 2015;25: R621-34. doi:10.1016/j.cub.2015.05.052

2. Ekman P, Friesen WV. Facial action coding system. Environmental Psychology & Nonverbal. 1978.

3. Krumhuber EG, Skora LI, Hill HCH, Lander K. The role of facial movements in emotion recognition. Nat Rev Psychol. 2023. doi:10.1038/s44159-023-00172-1

4. Scherer KR, Ellgring H. Multimodal expression of emotion: affect programs or componential appraisal patterns? Emotion. 2007;7: 158–171. doi:10.1037/1528-3542.7.1.158

5. Ghazanfar AA, Takahashi DY. Facial expressions and the evolution of the speech rhythm. J Cogn Neurosci. 2014;26: 1196–1207. doi:10.1162/jocn_a_00575

6. Chandrasekaran C, Trubanova A, Stillittano S, Caplier A, Ghazanfar AA. The natural statistics of audiovisual speech. PLoS Comput Biol. 2009;5: e1000436. doi:10.1371/journal.pcbi.1000436

7. Jack RE, Schyns PG. Toward a social psychophysics of face communication. Annu Rev Psychol. 2017;68: 269–297. doi:10.1146/annurev-psych-010416-044242

8. Jack RE, Garrod OGB, Schyns PG. Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. Curr Biol. 2014;24: 187–192. doi:10.1016/j.cub.2013.11.064

9. Liu M, Duan Y, Ince RAA, Chen C, Garrod OGB, Schyns PG, et al. Facial expressions elicit multiplexed perceptions of emotion categories and dimensions. Curr Biol. 2022;32: 200-209.e6. doi:10.1016/j.cub.2021.10.035

10. Gill D, Garrod OGB, Jack RE, Schyns PG. Facial movements strategically camouflage involuntary social signals of face morphology. Psychol Sci. 2014;25: 1079–1086. doi:10.1177/0956797614522274

11. Sowden S, Schuster BA, Keating CT, Fraser DS, Cook JL. The role of movement kinematics in facial emotion expression production and recognition. Emotion. 2021;21: 1041–1061. doi:10.1037/emo0000835

12. Burrows AM. The facial expression musculature in primates and its evolutionary significance. Bioessays. 2008;30: 212–225. doi:10.1002/bies.20719

13. Sherwood CC. Comparative anatomy of the facial motor nucleus in mammals, with an analysis of neuron numbers in primates. Anat Rec A Discov Mol Cell Evol Biol. 2005;287: 1067–1079. doi:10.1002/ar.a.20259

14. Todorov E. Optimality principles in sensorimotor control. Nat Neurosci. 2004;7: 907–915. doi:10.1038/nn1309

15. Pang R, Lansdell BJ, Fairhall AL. Dimensionality reduction in neuroscience. Curr Biol. 2016;26: R656-60. doi:10.1016/j.cub.2016.05.029

16. Latash ML, Scholz JP, Schöner G. Toward a new theory of motor synergies. Motor Control. 2007;11: 276–308. doi:10.1123/mcj.11.3.276

17. Chiovetto E, Salatiello A, d'Avella A, Giese MA. Toward a unifying framework for the modeling and identification of motor primitives. Front Comput Neurosci. 2022;16: 926345. doi:10.3389/fncom.2022.926345

18. Chiovetto E, Curio C, Endres D, Giese M. Perceptual integration of kinematic components in the recognition of emotional facial expressions. J Vis. 2018;18: 13. doi:10.1167/18.4.13

19. Labaune O, Deroche T, Teulier C, Berret B. Vigor of reaching, walking, and gazing movements: on the consistency of interindividual differences. J Neurophysiol. 2020;123: 234–242. doi:10.1152/jn.00344.2019

20. Ivanenko YP, Poppele RE, Lacquaniti F. Five basic muscle activation patterns account for muscle activity during human locomotion. J Physiol (Lond). 2004;556: 267–282. doi:10.1113/jphysiol.2003.057174

21. Chiovetto E, Berret B, Pozzo T. Tri-dimensional and triphasic muscle organization of whole-body pointing movements. Neuroscience. 2010;170: 1223–1238. doi:10.1016/j.neuroscience.2010.07.006

22. Zangrossi A, Cona G, Celli M, Zorzi M, Corbetta M. Visual exploration dynamics are low-dimensional and driven by intrinsic factors. Commun Biol. 2021;4: 1100. doi:10.1038/s42003-021-02608-x

23. Ramanarayanan V, Goldstein L, Narayanan SS. Spatio-temporal articulatory movement primitives during speech production: extraction, interpretation, and validation. J Acoust Soc Am. 2013;134: 1378–1394. doi:10.1121/1.4812765

24. Barliya A, Omlor L, Giese MA, Berthoz A, Flash T. Expression of emotion in the kinematics of locomotion. Exp Brain Res. 2013;225: 159–176. doi:10.1007/s00221-012-3357-4

25. Montepare JM, Goldstein SB, Clausen A. The identification of emotions from gait information. J Nonverbal Behav. 1987;11: 33–42. doi:10.1007/BF00999605

26. Dalmaijer ES, Lee A, Leiter R, Brown Z, Armstrong T. Forever yuck: Oculomotor avoidance of disgusting stimuli resists habituation. J Exp Psychol Gen. 2021;150: 1598–1611. doi:10.1037/xge0001006

27. Mackevicius EL, Bahle AH, Williams AH, Gu S, Denisenko NI, Goldman MS, et al. Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience. eLife. 2019;8. doi:10.7554/eLife.38471

28. Rucci M, Poletti M. Control and Functions of Fixational Eye Movements. Annu Rev Vis Sci. 2015;1: 499–518. doi:10.1146/annurev-vision-082114-035742

29. Torricelli F, Tomassini A, Pezzulo G, Pozzo T, Fadiga L, D'Ausilio A. Motor invariants in action execution and perception. Phys Life Rev. 2023;44: 13–47. doi:10.1016/j.plrev.2022.11.003

30. Krumhuber EG, Scherer KR. Affect bursts: dynamic patterns of facial expression. Emotion. 2011;11: 825–841. doi:10.1037/a0023856

31. Schmidt KL, Cohn JF, Tian Y. Signal characteristics of spontaneous facial expressions: automatic movement in solitary and social smiles. Biol Psychol. 2003;65: 49–66. doi:10.1016/S0301-0511(03)00098-X

32. Aghabozorgi S, Seyed Shirkhorshidi A, Ying Wah T. Time-series clustering – A decade review. Inf Syst. 2015;53: 16–38. doi:10.1016/j.is.2015.04.007

33. Aggarwal CC. An introduction to cluster analysis. In: Aggarwal CC, Reddy CK, editors. Data clustering: algorithms and applications. Chapman and Hall/CRC; 2018. pp. 1–28. doi:10.1201/9781315373515-1

34. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature. 1999;401: 788–791. doi:10.1038/44565

35. Delis I, Chen C, Jack RE, Garrod OGB, Panzeri S, Schyns PG. Space-by-time manifold representation of dynamic facial expressions for emotion categorization. J Vis. 2016;16: 14. doi:10.1167/16.8.14

36. Shannon CE, Weaver W. The mathematical theory of communication. books.google.com; 1998.

37. Le Mau T, Hoemann K, Lyons SH, Fugate JMB, Brown EN, Gendron M, et al. Professional actors demonstrate variability, not stereotypical expressions, when portraying emotional states in photographs. Nat Commun. 2021;12: 5037. doi:10.1038/s41467-021-25352-6

38. Parkinson B. Social Functions. Heart to heart: how your emotions affect other people. Cambridge University Press; 2019. pp. 215–240. doi:10.1017/9781108696234.006

39. Siegman AW, Feldstein S, editors. Multichannel integrations of nonverbal behavior. Psychology Press; 2014. doi:10.4324/9781315802305

40. Gobel MS, Kim HS, Richardson DC. The dual function of social gaze. Cognition. 2015;136: 359–364. doi:10.1016/j.cognition.2014.11.040

41. Wolpert DM, Doya K, Kawato M. A unifying computational framework for motor control and social interaction. Philos Trans R Soc Lond B Biol Sci. 2003;358: 593–602. doi:10.1098/rstb.2002.1238

42. Bell SM, Ainsworth MD. Infant crying and maternal responsiveness. Child Dev. 1972;43: 1171–1190. doi:10.2307/1127506

43. Rowe C. Receiver psychology and the evolution of multicomponent signals. Anim Behav. 1999;58: 921–931. doi:10.1006/anbe.1999.1242

44. Duchenne de Boulogne GB. The mechanism of human facial expression. Cuthbertson RA, editor. Cambridge University Press; 1990. doi:10.1017/CBO9780511752841

45. Brewer R, Biotti F, Catmur C, Press C, Happé F, Cook R, et al. Can neurotypical individuals read autistic facial expressions? atypical production of emotional facial expressions in autism spectrum disorders. Autism Res. 2016;9: 262–271. doi:10.1002/aur.1508

46. Haque A, Fei-Fei L, Guo M, Miner AS. Measuring Depression Symptom Severity from Spoken Language and 3D Facial Expressions. Machine Learning for Health (ML4H) Workshop at NeurIPS 2018, Montréal, Canada. 2018.

47. Rawal N, Stock-Homburg RM. Facial emotion expressions in human–robot interaction: A survey. Int J Soc Robot. 2022;14: 1583–1604. doi:10.1007/s12369-022-00867-0

48. Baltrušaitis T, Robinson P. Openface: an open source facial behavior analysis toolkit. 2016 IEEE Winter. 2016.

49. Fu X, Huang K, Sidiropoulos ND, Ma W-K. Nonnegative matrix factorization for signal and data analytics: identifiability, algorithms, and applications. IEEE Signal Process Mag. 2019;36: 59–80. doi:10.1109/MSP.2018.2877582

50. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics. 2010;11: 367. doi:10.1186/1471-2105-11-367

51. Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci. 2009;12: 535–540. doi:10.1038/nn.2303

52. Varoquaux G, Colliot O. Evaluating machine learning models and their diagnostic value. In: Colliot O, editor. Machine learning for brain disorders. New York, NY: Humana; 2023. doi:10.1007/978-1-0716-3195-9_20

53. Baldán FJ, Benítez JM. Multivariate times series classification through an interpretable representation. Inf Sci (Ny). 2021;569: 596–614. doi:10.1016/j.ins.2021.05.024

54. Dobs K, Bülthoff I, Breidt M, Vuong QC, Curio C, Schultz J. Quantifying human sensitivity to spatio-temporal information in dynamic faces. Vision Res. 2014;100: 78–87. doi:10.1016/j.visres.2014.04.009

55. Furl N, Begum F, Sulik J, Ferrarese FP, Jans S, Woolley C. Face space representations of movement. Neuroimage. 2020;212: 116676. doi:10.1016/j.neuroimage.2020.116676

56. Miranda JGV, Daneault JF, Vergara-Diaz G, Torres Â, Quixadá AP, Fonseca ML, et al. Complex Upper-Limb Movements Are Generated by Combining Motor Primitives that Scale with the Movement Size. Sci Rep. 2018;8: 12918. doi:10.1038/s41598-018-29470-y

57. Hooge I, Camps G. Scan path entropy and arrow plots: capturing scanning behavior of multiple observers. Front Psychol. 2013;4: 996. doi:10.3389/fpsyg.2013.00996

58. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:14065823. 2014.