



Exploring Mental Prototypes by an Efficient Interdisciplinary Approach: Interactive Microbial Genetic Algorithm

Sen Yan, Catherine Soladie, Renaud Segulier

► To cite this version:

Sen Yan, Catherine Soladie, Renaud Segulier. Exploring Mental Prototypes by an Efficient Interdisciplinary Approach: Interactive Microbial Genetic Algorithm. 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG), Jan 2023, Waikoloa Beach, United States. pp.1-8, 10.1109/FG57933.2023.10042515 . hal-04050608

HAL Id: hal-04050608

<https://hal.science/hal-04050608>

Submitted on 29 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exploring Mental Prototypes by an Efficient Interdisciplinary Approach: Interactive Microbial Genetic Algorithm

Sen Yan, Catherine Soladié, Renaud Segulier
CentraleSupélec, IETR, UMR CNRS 6164, Rennes, France

Abstract—Facial expression-based technologies have flooded our daily lives. However, most technologies are limited to Ekman’s basic facial expressions and rarely deal with more than ten emotional states. This is not only due to the lack of prototypes for complex emotions but also the time-consuming and laborious task of building an extensive labeled database. To remove these obstacles, we were inspired by a psychophysical approach for affective computing, so-called the reverse correlation process (RevCor), to extract mental prototypes of what a given emotion should look like for an observer. We proposed a novel, efficient, and interdisciplinary approach called Interactive Microbial Genetic Algorithm (IMGA) by integrating the concepts of RevCor into an interactive genetic algorithm (IGA). Our approach achieves four challenges: online feedback loop, expertise-free, velocity, and diverse results. Experimental results show that for each observer, with limited trials, our approach can provide diverse mental prototypes for both basic emotions and emotions that are not available in existing deep-learning databases. Our work is available at <https://yansen0508.github.io/Interactive-Microbial-Genetic-Algorithm/>.

I. INTRODUCTION

Facial expressions (FEs) are a natural and universal means of conveying social information [7]. Nowadays, FE-based technologies are ubiquitous, such as facial expression manipulation technologies in TikTok, Instagram, Zoom, or other social media [32], and automatic facial expression recognition technologies in social robots [21]. However, there are still major challenges. First, most FE-based systems are based on the prototypes defined by Ekman et al., which are supposed to be universally perceived by humans [8], [10]. However, the universality of Ekman’s prototypes is now being challenged by a growing number of psychologists [16], [27], [1]. This indicates that there may be many prototypes for one emotional label. Moreover, most systems can only deal with six basic facial expressions: happiness, sadness, anger, disgust, fear, and surprise. Yet, as research in cognitive science covers, there are more than 4000 labels of emotions [29]. To address more critical domains such as psychotherapy or the service industry, machine learning applications should develop tools to describe, understand or detect more emotions in real life. For instance, one can imagine training such as self-confidence before a job interview or dealing with anxiety in a therapeutic context [12]. Nowadays, most computer science tools require a large amount of reliable labeled data. Although some databases exist for some specific emotions, e.g., compound emotions [11], they are far from

the 4000 emotional labels [29]. Indeed, creating such a database comes with many concerns: lack of labeled data (e.g., self-confidence), requiring expertise for some labeling tasks, and time-consuming and labor-intensive human annotations.

To tackle these challenges, we propose a different way of thinking from the traditional machine learning process. We get inspired by the psychophysical reverse correlation (RevCor) process, typically employed for affective computing in psychology [23], [3]. Based on data-driven methods, RevCor is used to extract the mental prototypes of what a given basic or non-basic emotion should look like for an observer (or called participant). Yet, RevCor has imperfections that should be solved to be largely applied in various scenarios, e.g., a digital coach for the online interview and a digital mirror treating psychiatric disorders of emotion. First, RevCor is based on the assumption that there is one, and only one, mental prototype for one affective state that exists in one or a group of observers. This unicity can be questioned. Moreover, each observer needs to perform massive trials randomly generated by domain experts. Time-consuming experiments lead to user fatigue, and the reliance on expertise also hinders expansion to other areas.

Hence, a system that aims to extract mental prototypes should fulfill the following requirements. 1) The system should find prototypes that are not limited to basic emotions. The emotion prototypes that are not available in existing deep-learning databases should also be found. 2) The entire process should not require expert knowledge. 3) The system should consider user fatigue [15], [2], which means no longer than 15 minutes. 4) For one emotion, the system should obtain multiple prototypes, thus being closer to reality.

In this paper, we created an efficient interdisciplinary approach called IMGA for Interactive Microbial Genetic Algorithm (illustrated in Fig. 1) that meets the mentioned requirements. The first originality of our approach is that we integrated the psychophysical reverse correlation (RevCor) process into the interactive genetic algorithm (IGA). The strengths of this integration are as follows. 1) **Online feedback loop.** Unlike the traditional RevCor that generates massive trials randomly, in our approach, based on the observer’s feedback, automatically updated trials can contain more valuable information (closer to the mental prototypes of observers). 2) **Expertise-free.** Everyone can use our approach to extract their own mental prototypes of a given emotion. The categories of emotions are not limited to those provided in existing deep-learning databases. Our

This work was supported by Randstad and ANR REFLETS.

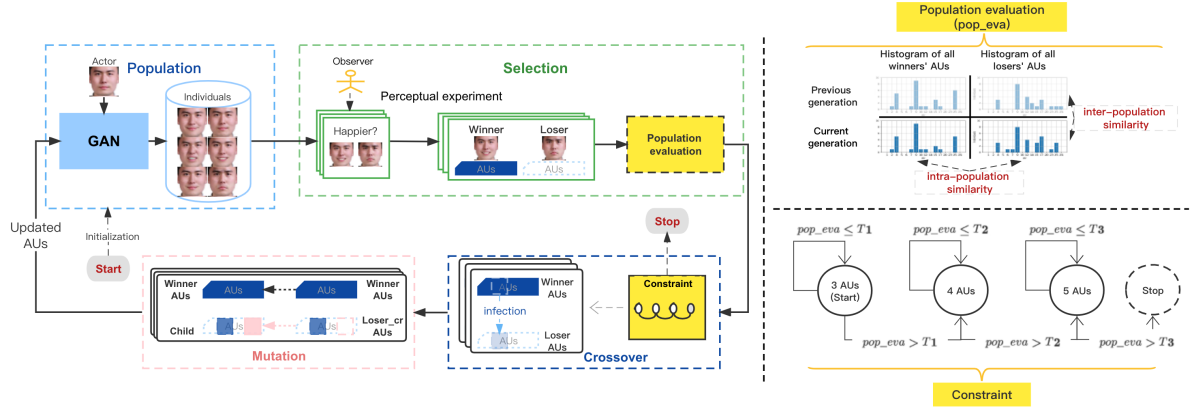


Fig. 1. Framework of our interactive microbial genetic algorithm (IMGA). An efficient interdisciplinary approach integrates the psychophysical reverse correlation process (RevCor) into an interactive genetic algorithm (IGA). For the genetic algorithm module within IGA, we adopt the microbial genetic algorithm (MGA) that can obtain various mental prototypes and accelerate the system’s convergence. To monitor the convergence of the system and evaluate the quality of the entire population with limited trials, we add a population evaluation module. We also add a constraint automaton to limit the manipulation of facial expressions and to determine the termination of the system. For the tool to generate different facial expressions, we employ GANimation [26] (marked by “GAN”) controlled by facial action units [9].

approach only requires the observer’s perception (i.e., subjective judgment) rather than the observer’s expertise (e.g., no expert knowledge in affective computing, psychology, or certified FACS coders [9]). 3) **Velocity**. The number of trials required for mental prototype computations is thereby reduced. 4) **Diverse results**. For one emotion, IMGA can provide multiple mental prototypes to each observer. This is closer to reality. The second originality of our approach is that within IGA, we adopt the microbial genetic algorithm (MGA) [14] to increase the converging speed of the system. The third originality of our approach is that, differing from the traditional genetic algorithm that needs to acquire the fitness values of all individuals, we added a population evaluation module that evaluates the quality of the entire population with limited trials. In addition, we added a three-state constraint automaton to gradually increase the number of activated facial action units (AUs) [9] for each face and determine the process’s termination.

II. RELATED WORKS

Nowadays, defining and understanding facial expression prototypes can be done via the psychophysical reverse correlation process (RevCor). We overview the related work using RevCor for mental prototype computation. Then, we overview interactive genetic algorithms (IGA) for optimizing the system based on subjective judgments.

A. Psychophysical reverse correlation process

According to observers’ judgments of a large quantity of randomly-varied stimuli, RevCor is able to reverse-engineer what perceptual representations subtend these judgments [3]. This process is widely used in cognitive science to study the perception of faces [16], [33], [5], speech [25], [4], and bodies [17], [22]. In an influential example, Jack et al. [16] randomly generated 4800 trials. Each trial consists of one dynamic facial animation created by a 3D morphing tool

[33]. In the perceptual experiment, 15 Western Caucasian and 15 East Asian observers were asked to categorize the animations into six basic emotion categories. For each cultural group, after collecting all feedback, the authors aggregated one mental representation corresponding to each emotion and concluded that these representations were not culturally universal.

It is less efficient for one observer to perform numerous trials and obtain one mental prototype. Indeed, a speech intonation study using RevCor indicated that for some tasks, researchers could have reached the same precision with fewer trials [4]. An online feedback loop can be more efficient in updating trials containing more valuable information by iteration. Hence, interactive genetic algorithms (IGA) can be a good idea.

B. Interactive genetic algorithm

IGA optimizes the system to fit the user preference based on subjective judgments [30]. The process of IGA is similar to the traditional genetic algorithm (GA). Yet, the main difference is that the fitness function is the subjective judgments of users rather than a mathematically objective fitness function. IGA is widely used in several areas: geology [24], design [19], [31], [28], and image processing, such as image retrieval [6], [20] and 3D facial animations [13]. To the best of our knowledge, there is no literature using IGA to manipulate AUs for affective computing. Like RevCor, IGA considers subjective judgments, but contrary to RevCor, it can offer multiple solutions. Hence, RevCor can be embedded into an IGA system to generate more user-preferred trials closer to the mental prototypes, thus reducing the workload of human observers and obtaining multiple mental prototypes.

Since an elitist GA can converge fast [18], we adopted the microbial genetic algorithm (MGA) [14] as the GA module within IGA. To the best of our knowledge, MGA

has never been employed in an interactive system. We named our approach interactive microbial genetic algorithm (IMGA), which is a novel process integrating RevCor into an MGA-based IGA. Such an approach not only improves the efficiency of RevCor but also brings multiple mental prototypes.

III. METHODOLOGY

Fig. 1 describes our IMGA process. Similar to the traditional Genetic Algorithm (GA), it is an iterative process that repeats four steps: Population (initialization in Section III-A and update in Section III-E), Selection (in Section III-B), Crossover (in Section III-C), and Mutation (in Section III-D). For the glossary of our IMGA, the individuals are facial expressions that evolved by iteration. The population in each iteration is called a generation. The entire system, especially the interaction between the human (observer) and the machine (GA system), is detailed by a video demonstration¹.

A. Population: initialization

We can employ any tool to generate facial expressions defined by low-level attributes. Here, we choose GANimation [26] controlled by facial action units (AUs) [9], i.e., the low-level attributes. This module can manipulate facial expressions with a relatively fine control. Thus, more facial expressions can be produced by different combinations of AU, regardless of whether these expressions belong to a certain type of emotion or not. In this procedure, GANimation (thereafter, *Gan*) takes as input a colored image of the actor's face s (e.g., captured with an emotionally neutral expression) and a n -dimensional binary vector v of AUs to create a deformed face (i.e., individual): $I = Gan(s, v)$.

GANimation is capable of manipulating $n = 16$ AUs from the list $\mu = \{1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26\}$. We define as $v = [\lambda_1, \dots, \lambda_n]$, the binary AU vector where each component λ_i represents the activation ($\lambda_i = 1$) or deactivation ($\lambda_i = 0$) of $AU_{\mu[i]}$ (the i_{th} element in the AU list μ). For instance, $\lambda_3 = 1$ represents that AU4 (brow lowerer) is activated, $\lambda_9 = 0$ represents that AU12 (lip corner puller) is deactivated. See the literature [9] for a complete list of AUs.

While GANimation can, in principle, simultaneously activate AUs, activating too many AUs typically create visual artifacts. Moreover, the state-of-the-art (SOTA) facial expression prototypes of Ekman et al. [9] indicate that most facial expressions have between 3 and 5 AUs activated. Therefore, we initialize the individuals by activating 3 AUs: $\forall v, \sum_{i=1}^{16} \lambda_i = c$, where $c = 3$. There can be $C_{16}^3 = 560$ possible AU vectors $\mathcal{V} = \{v_1, v_2, \dots, v_{560}\}$. Based on the actor's face s , we randomly choose N out of the 560 AU vectors to initialize a population of N individuals. Fig.1 displays some examples of individuals from the initial population.

B. Selection

Selection has two parts: the perceptual experiment of RevCor and the population evaluation module.

Perceptual experiment. In the perception experiment, we group the population of N individuals into $N/2$ pairs. Note that each pair of individuals is displayed only once for each iteration. In each trial of the perceptual experiment, a pair of individuals is displayed. Observers are asked to choose which individual best corresponds to the target expression (e.g., "which of these two faces looks happier?"). Note that each observer conducts $N/2$ trials for each generation. According to the answer from the observer, each pair of individuals are annotated by "winner" and "loser". Next, we use the set of $N/2$ winners (W) and the set of $N/2$ losers (L) to evaluate the quality of the current generation.

Population evaluation. In the traditional GA, the computer can easily assign each individual a fitness value and rank all individuals from the best fit to the worst fit based on a mathematical fitness function. With the winner-loser strategy, $N(N-1)/2$ trials are required to rank the individuals. In our case, since the fitness function is the subjective judgments of observers, we cannot afford so many trials. That is why we only get $N/2$ trials, and we add a population evaluation module to evaluate the entire population.

For the population evaluation module (*pop_eva* of Fig. 1), we compute the similarity between the losers in the previous generation and the current generation, so-called inter-population similarity (*inter_pop*), and the similarity between winners and losers in the current generation, so-called intra-population similarity (*intra_pop*).

$$pop_eva = \alpha \cdot inter_pop + \beta \cdot intra_pop \quad (1)$$

$$inter_pop = corr(H[L_g], H[L_{g-1}]) \quad (2)$$

$$intra_pop = corr(H[W_g], H[L_g]) \quad (3)$$

α and β are positive constants. The similarity is computed by the Pearson correlation $corr(\cdot, \cdot)$. $H[\cdot]$ represents the histogram that counts how many times each AU (from the list μ) occurs in the corresponding set. W_g and L_g denote the AUs of winners W and losers L in the g_{th} generation, respectively. When the inter-population similarity increased, we can infer that there were fewer changes between the successive generations. When the intra-population similarity increased, we can infer that the losers became closer to the winners. Overall, we maximize *pop_eva* to ensure the convergence of the system.

C. Crossover

Three-state constraint automaton. Here, we add the three-state constraint automaton to gradually increase the number of activated AUs from 3 to 5 during the crossover. Details are shown in Constraint of Fig. 1. The entire population goes through three restricted states: initially 3-AU activation state, then 4-AU activation state, and finally 5-AU activation state (denoted by "3 AUs", "4 AUs", and "5 AUs"), and three thresholds are given as $[T1, T2, T3]$ accordingly.

¹<https://youtu.be/GwvC2u9r01o>

The population is initialized by activating only 3 AUs for each individual, i.e., $c=3$ defined in Section III-A. During the first state, i.e., 3-AU activation state, each individual can not have more than c AUs activated after the crossover. Once pop_eva exceeds the first threshold, i.e., $T1$, the system goes to the next state, i.e., 4-AU activation state. Accordingly, the threshold is updated to $T2$ and $c=4$ for the 4-AU activation state. As shown in Constraint of Fig. 1, the procedures in the following states are similar to that in the 3-AU activation state. Finally, the system stops when pop_eva exceeds the last threshold, i.e., $T3$.

Infection. The infection operator is the same as the literature of MGA [14], i.e., uniform infection. The binary AU vector of the loser is infected by that of the winner. Thus, each element in the binary AU vector of the loser can be replaced by the corresponding element of the winner (illustrated in Fig. 1). The crossover rate is defined by cr . If the loser after the infection has more than c AUs activated, we randomly deactivate the excess. Note that the AU vectors of winners are unchanged, and the AU vectors of losers after infection are named *Loser_cr* in Fig. 1.

D. Mutation

The mutation operator is the same as the literature on MGA [14], i.e., bit mutation. As shown in Fig. 1, each element of *Loser_cr* has the same mutation rate mr to change its binary value. Note that due to the infection and the mutation, individuals can have or less than 3 AU activated, even though they are allowed to have more than 3 AU activated.

E. Population: update

In order to keep the size of the population unchanged, we only replace AU vectors of the losers with their offspring and keep the winners unchanged. In our approach, we employ the same tool, i.e., GANimation, to create the next generation by the updated low-level attributes, i.e., AU vectors.

IV. EXPERIMENTS

The goal of our experiments is to validate our approach that can efficiently generate multiple prototypes corresponding to a given emotion, even a complex emotion. We first list the implementation details of our process in Section IV-A, then detail the experimental protocol in Section IV-B, and finally analyze the evolution of the population in Section IV-C.

A. Implementation details

GANimation model. We use the code of GANimation [26] released by its authors. All settings are unchanged.

GA parameter settings. We took the suggestion from the original literature on MGA [14]: $cr = 0.5$ and $mr = 0.03$. The other GA parameters, i.e., the population size N , the constants α , β (defined in Eq. 1), and the thresholds for the three-state constraint automaton, were calibrated empirically. For this purpose, we simulated perceptual experiments by replacing the real observers shown in Fig. 1 with an

automatic facial expression recognition system. According to the simulations, we set the population size as $N = 20$, the constants of population evaluation $\alpha = 0.5, \beta = 0.5$, and thresholds of the three-state constraint automaton as $[T1=0.9, T2=0.9, T3=0.95]$.

B. Experimental protocol

Observer demography. To validate our approach, we decided to recruit 12 observers since related works proposing new tools for perceptual experiments recruited a limited number of participants, from 8 to 12 [33], [4]. The 12 observers we recruited are adults (mean age: 34.7 yo) from five cultural groups: Algeria (1), China (1), Brazil (1), France (8), and Russia (1). Only two of the 12 observers have experience in affective computing, whereas nobody is a certified coder in Facial Action Coding System [9]. Each observer signed informed consent, and the experimental data were anonymous.

Perceptual experiment. To illustrate the efficiency of our approach, we chose three basic emotions (happiness, sadness, and anger) that existed in deep-learning databases and one complex emotion (confidence) that is not available in existing deep-learning databases. Each of the 12 observers participated in four different experimental tasks to find his/her mental prototypes of happiness, sadness, anger, and confidence. In each task, the question is fixed and unique. For example, “Which of these two faces looks happier?” The order of the four experimental tasks was counterbalanced among observers, and all experimental tasks used the same actor’s photograph. Based on the three-state constraint automaton, the experimental task was automatically terminated. However, if the population has evolved for 50 generations, we forcibly stopped the current experimental task. Between the experimental tasks, observers had a 5-minute rest. All experiments were conducted in a quiet room in the laboratory, using a custom computer graphic interface from PsychoPy.

C. Results: evolution of the population

For all experimental results, we can observe the gradual evolution of the population. Fig.2 illustrates this evolution through the experimental results of anger from observer#4. In Fig.2(a), we monitored the individuals according to the first, second, and third frequently occurring AUs and presented the representative prototypes. Fig.2(b) illustrates the corresponding values computed by the population evaluation module.

The AUs related to the observer’s mental prototypes survived, while unrelated AUs gradually disappeared. Based on the subjective judgment from observer #4 on “which of these two faces looks angrier?” AU9 was the most frequently occurring AU in the last generation. AU25 and AU4 were the second and the third. They are more relevant than the other AUs to anger for this observer. If we look at the evolution from the first to the last generation, we observe that the related AUs (AU9, AU25, and AU4) spread throughout the population, and the individuals without

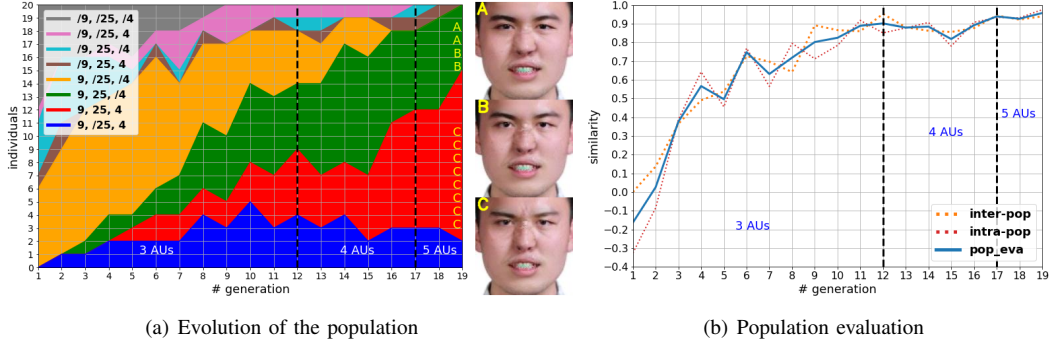


Fig. 2. Experimental results of anger from observer #4. 2(a) left: evolution of the population over generations. The x-axis represents the generation number of the population. The y-axis represents the individuals of the current population. The legend lists 8 classes of individuals based on the activation or deactivation (marked by “/”) of the related AUs (AU9, AU25, and AU4). For instance, “9, /25, 4” (blue) denotes all the individuals who had AU9 and AU4 activated and had AU25 deactivated. 2(a) right: three representative prototypes, i.e., the individuals with the same AU vectors in the last generation, where A: AU7, AU9, AU25; B: AU9, AU25; C: AU4, AU9, AU25. 2(b): Population evaluation. We draw the curves of the similarities computed by the population evaluation module. The vertical dotted lines in 2(a) and 2(b) indicate changing the constraint in the 12_{th} and the 17_{th} generation.

the related AUs activated were gradually eliminated (the gray area of Fig. 2(a)). Therefore, more individuals with the related AUs activated appeared in subsequent trials.

The related AUs combined with each other. As the experiment proceeded, we noticed the growths of the green, red and blue areas and the disappearance of the other areas. This indicates that observer #4 prefers to combine AU9 with AU4 and/or AU25 than the other AUs. During the 3-AU activation state, the system gradually converged. Indeed, 18 of 20 individuals had AU9 activated in the 12_{th} generation. At the end of the experiment, 100% of the population had AU9 activated (green, red, and blue areas), 90% of the population had the combination of AU9 and AU25 (green and red areas), 75% of the population had the combination of AU9 and AU4 (red and blue areas), and 65% of the population had activated both AU4, AU9, and AU25 (red area).

Fig. 2(b) reflects the convergence of the system by illustrating the similarities computed by the population evaluation module. Our approach considers not only the inter-population similarity but also the intra-population similarity. During the 3-AU activation state, the system gradually converged. Once changing the constraint, the system searched for results in a broader space and then converged again. That is why the curves dropped and re-converged during the 4-AU activation state and the 5-AU activation state. See the video demonstration for extra information: 1) the AU histograms of winners’ and losers’ AUs in the previous and the current generations, 2) winners and losers of the current generation and 3) the computer graphic interface for the experiments.

We define the **representative prototypes** as the individuals with the same AU vectors in the last generation. In Fig. 2(a), there are three different representative prototypes of anger from observer #4 denoted by “A” (2 individuals), “B” (2 individuals), and “C” (8 individuals). Next, we quantitatively and subjectively evaluate the representative prototypes.

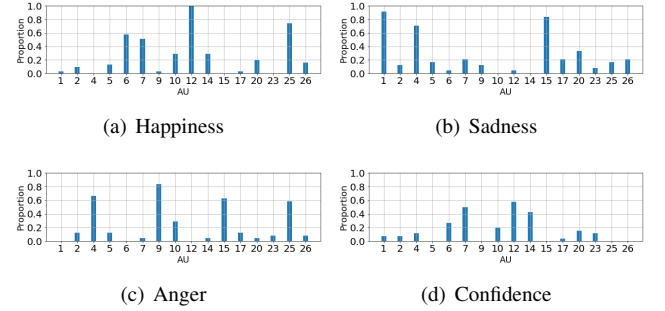


Fig. 3. Proportion of each AU in the representative prototypes. For the basic emotions, some AUs reveal universality.

V. EVALUATIONS AND COMPARISON

For each emotion category, we collected all representative prototypes of observers. In this section, first, we analyze representative prototypes from two perspectives: action units (in Section V-A) and prototypes (in Section V-B). Second, we present our subjective evaluation process (in Section V-C and Section V-D) for two purposes: 1) to validate that our representative prototypes can reflect observers’ mental prototypes and 2) to subjectively compare with the SOTA prototypes. Third, we discuss the efficiency by comparing our approach with the related works using RevCor for affective computing (in Section V-E).

A. Quantitative evaluation: Action units

Our IMGA-generated prototypes are compatible with state-of-the-art prototypes. Our findings indicate all representative prototypes generally convey similar emotional expressions. Fig. 3 presents the proportion of each AU that appears in our representative prototypes. We find all AUs from SOTA prototypes [9], [33] in our representative prototypes, except AU43 (eyes closed) for sadness and AU16 (lower lip depressor) for anger. Note that these AUs are not included in the list of editable AUs μ by GANimation aforementioned in Section III-A. To consult the AUs of

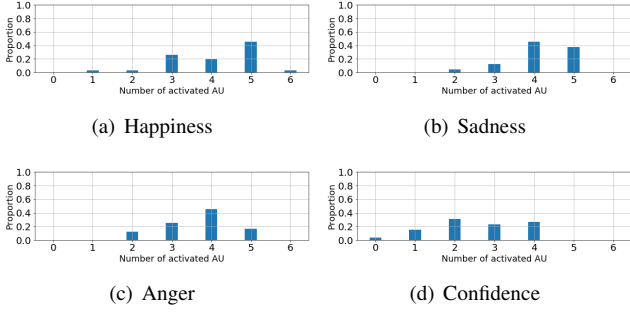


Fig. 4. The proportion of prototypes that have different numbers of AUs activated. There is a discrepancy between basic emotions and confidence.

SOTA prototypes, please see the literature [9], [33].

Within the scope of basic emotions, some AUs reveal universality. Some common AUs can be found in the representative prototypes of basic emotions. In Fig. 3, for happiness, 100% of representative prototypes have AU12 (lip corner raiser) activated. For sadness, more than 80% of representative prototypes have AU1 (inner brow raiser) and AU15 (lip corner depressor) activated. For anger, 83% of representative prototypes have AU9 (nose wrinkler) activated. For confidence, the proportions of AUs are not as prominent as those in the basic emotions: 57.7% of representation prototypes have AU12 (lip corner raiser) activated, and 50% of representative prototypes have AU7 (lid tightener) activated.

There is a discrepancy between basic emotions and confidence in terms of the number of activated AUs. Fig. 4 summarizes the proportion of prototypes that have different numbers of AUs activated. Typically, most representative prototypes of basic emotions have at least 3 AUs activated. For confidence, although our system initialized with the constraint of 3-AU activation, approximately 50% of the representative prototypes have less than 3 AUs activated.

B. Quantitative evaluation: Prototypes

The diversity of mental prototypes not only exists within observers, but also between observers. Table I indicates a great variety of prototypes. From 12 observers, we obtained 31, 24, 24, and 26 different representative prototypes of happiness, sadness, anger, and confidence. On average, multiple representative prototypes are acquired per observer. This indicates the diversity of mental prototypes within observers. Furthermore, only a small proportion of these different representative prototypes coexist in at least two observers. Most representative prototypes are different between observers. It can also indicate the diversity of the mental prototypes between observers. Given that neither the SOTA prototypes of Ekman et al. [9] (except sadness) nor Yu et al. [33] could be found in our representative prototypes, this implies that the SOTA prototypes need to be refined. The number of prototypes should be enlarged.

TABLE I
NUMBER OF DIFFERENT REPRESENTATIVE PROTOTYPES IN ALL OBSERVERS. AMONG THESE DIFFERENT REPRESENTATIVE PROTOTYPES, WE ALSO LIST THE PROPORTION OF COEXISTING PROTOTYPES BETWEEN DIFFERENT OBSERVERS.

	H	Sa	A	C
Number of proto	31	24	24	26
Proportion	22.6%	25%	16.7%	15.4%

TABLE II
PROPORTION OF OBSERVERS WHO STILL CHOOSE AT LEAST ONE OF THEIR REPRESENTATIVE PROTOTYPES. WE COMPUTE THE CORRESPONDING BASELINE BY RANDOM SELECTIONS.

	H	Sa	A	C
Ours	83.3%	75%	66.7%	66.7%
Baseline	41.6%	44.3%	39.7%	42.2%

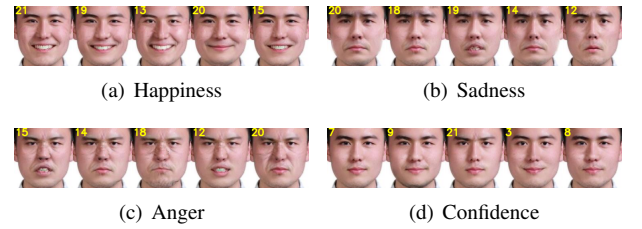


Fig. 5. We display the top-5 most selected prototypes by the 12 observers. The names of the prototypes are marked in yellow. There is no state-of-the-art prototype appearing in the top-5 prototypes.

TABLE III
EXPERIMENT TIME (IN MINUTES) OF OUR APPROACH.

	Happiness	Sadness	Anger	Confidence
mean	11.1	12.4	8.6	11.3
std	3.8	3.6	2.9	3.3

C. Subjective evaluation: Protocol and measurements

Here, we present our subjective evaluation. We asked the same 12 observers to participate in the subjective evaluation. The subjective evaluation process was divided into four tasks corresponding to the four facial expressions: happiness, sadness, anger, and confidence. In each task, we created an evaluation set including all representative prototypes of observers and the SOTA prototypes [9], [33] (except confidence, for which no SOTA prototype is available). All prototypes in each evaluation set were presented in shuffled order. In the tasks of happiness, sadness, anger, and confidence, observers were asked to select five faces that were the happiest / saddest / angriest / most confident, respectively.

We applied two measurements for the subjective evaluation. First, we count the proportion of observers who still choose at least one representative prototype of theirs. Second, in more detail, we ranked the representative prototypes from the most selected prototype to the least selected prototype by the 12 observers.

TABLE IV

COMPARISON BETWEEN OUR IMGA AND THE RELATED WORKS USING REVERSE CORRELATION PROCESS FOR AFFECTIVE COMPUTING [33], [16], [5], [25], [4]. WE LIST IN THE FIRST COLUMN: THE STIMULI CATEGORY, THE REVERSE CORRELATION PARADIGM, THE NUMBER OF AFFECTIVE STATES, THE NUMBER OF TRIALS PERFORMED BY ONE OBSERVER FOR ALL AFFECTIVE STATES, THE NUMBER OF TRIALS PERFORMED BY ONE OBSERVER FOR ONE AFFECTIVE STATE AND THE NUMBER OF MENTAL PROTOTYPES FOR ONE OBSERVER.

	IMGA	Yu et al. [33]	Jack et al. [16]	Chen et al. [5]	Ponsot et al. [25]	Burred et al. [4]
stimuli	face	face	face	face	speech	speech
paradigm	2-AFC	7-AFC	7-AFC	3-AFC	2-AFC	2-AFC
states	4	6	6	2	2	1
trials/obs		2400	4800	3600		
trials/obs/state	330	(400)	(800)	(1800)	~700	700
proto/obs	multiple	single	single	single	single	single

D. Subjective evaluation: Results

Most observers still selected at least one of their mental prototypes. Table II lists the proportion of observers who still choose at least one representation prototype of theirs. The baseline is derived from random selections of each evaluation task. Compared with the baselines, a larger proportion of observers still selected at least one of their representative prototypes. This can indicate that our representative prototypes can reflect the observers’ mental prototypes.

Observers less preferred state-of-the-art prototypes. We first sorted all the prototypes according to the proportion selected by observers. Then, we displayed the top-5 prototypes with the highest proportions in Fig. 5. We noticed that there is no SOTA prototype appearing in the top-5 prototypes of the three basic emotions.

E. Comparison with related works using RevCor: Efficiency

First, we present the experiment time of our approach. Then, in order to discuss the efficiency, we compare the experiment time and the number of mental prototypes for each observer between our IMGA and the related works using RevCor for affective computing.

Converging speed of our IMGA. In Table III, we present the duration for observers to perform the perceptual experiments. On average, observers performed the perceptual experiments on anger faster than the experiments on the other facial expressions. By calculating the average time for all perceptual experiments, it takes about 10.8 minutes (with 330 trials) for an observer to obtain mental prototypes using our IMGA.

Comparison with related works. Since all the related works did not provide the experiment time, we compared the number of trials for the perceptual experiment. Table IV illustrates the results of the related work using RevCor, three for facial expression and two for speech. Due to the different paradigms, the numbers of trials are presented in two different ways, i.e., “trials/obs” and “trials/obs/state”. By comparing our IMGA (“330”) with the works using the two-alternative forced choice (2-AFC) paradigm, (“~700”, and “700” for [25], and [4]), our work reduced the number of required trials (per observer, for one affective state) by approximately a factor of two.

We cannot directly compare with the works [33], [16], [5] that employed different paradigms, since “trials/obs/state” is

unknown in the original literature. By calculating “trials/obs” / “number of affective states” to compare these works (“(400)”, “(800)”, and “(1800)” for [33], [16], and [5]) with ours (“330”), our approach still needs fewer trials than these works.

In summary, compared with related works using RevCor, our approach has two strengths. First, our approach shrinks the experiment time. Second, only our approach can obtain multiple mental prototypes for each observer.

VI. CONCLUSION

In this paper, we proposed an efficient interdisciplinary approach: Interactive Microbial Genetic Algorithm (IMGA). Such an interdisciplinary approach that integrated the psychological reverse correlation process (RevCor) into an interactive genetic algorithm (IGA) efficiently explored diverse mental prototypes in a broader range of facial expressions (basic emotion and non-basic emotion) for each observer. Our IMGA considered real-time feedback from observers to update subsequent trials. Differing from the traditional genetic algorithm, we added a population evaluation module to evaluate the quality of the entire population with limited trials and a three-state constraint module to limit the manipulation of facial expressions and determine the termination of the system.

Compared with the SOTA prototypes [9], [33], we observe that the diversity of mental prototypes exists not only within observers but also between observers. Thus, the prototypes of a given emotion should be enlarged. Furthermore, our approach can extract the emotions that are not available in existing deep-learning databases. Compared with the related works using RevCor [33], [16], [5], [25], [4], our approach is more efficient at two-fold: faster and obtaining multiple mental prototypes. However, the limitation of the related works and our IMGA is that all the stimuli are unimodal. Multimodal stimuli (e.g., video and audio) should be employed to enrich affective computing studies in the future. Another limitation comes from the facial-expression manipulation tool we chose. Indeed, GANimation [26] provides only 16 editable AUs. In the future, GANimation can be replaced by the other facial-expression-manipulation tool that can edit more AUs.

Overall, we hope our approach can pave the way for further scientific studies not only in cognitive science but also

in computer science, such as non-basic emotion recognition (for lack of labeled databases) using mental prototypes. We also expect this approach can be customized for audiences in different application domains, e.g., a digital coach for the online interview and a digital mirror treating psychiatric disorders of emotion.

VII. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the contribution of reviewers' comments.

REFERENCES

- [1] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest*, 20(1):1–68, 2019.
- [2] R. Berto, M. R. Baroni, A. Zainaghi, and S. Bettella. An exploratory study of the effect of high and low fascination environments on attentional fatigue. *Journal of environmental psychology*, 30(4):494–500, 2010.
- [3] L. Brinkman, A. Todorov, and R. Dotsch. Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, 28(1):333–361, 2017.
- [4] J. J. Burred, E. Ponsot, L. Goupil, M. Liuni, and J.-J. Aucouturier. Cleese: An open-source audio-transformation toolbox for data-driven experiments in speech and music cognition. *PloS one*, 14(4):e0205943, 2019.
- [5] C. Chen, C. Crivelli, O. G. Garrod, P. G. Schyns, J.-M. Fernández-Dols, and R. E. Jack. Distinct facial expressions represent pain and pleasure across cultures. *Proceedings of the National Academy of Sciences*, 115(43):E10013–E10021, 2018.
- [6] S.-B. Cho and J.-Y. Lee. A human-oriented image retrieval system using interactive genetic algorithm. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 32(3):452–458, 2002.
- [7] C. Darwin. *The expression of the emotions in man and animals*. University of Chicago press, 2015.
- [8] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124, 1971.
- [9] P. Ekman and W. V. Friesen. Facial action coding system. *Environmental Psychology & Nonverbal Behavior*, 1978.
- [10] P. Ekman, W. V. Friesen, M. O'sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti, et al. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, 53(4):712, 1987.
- [11] C. Fabian Benitez-Quiroz, R. Srinivasan, and A. M. Martinez. Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5562–5570, 2016.
- [12] N. Guerouaou, G. Vaiva, and J.-J. Aucouturier. The shallow of your smile: the ethics of expressive vocal deep-fakes. *Philosophical Transactions of the Royal Society B*, 377(1841):20210083, 2022.
- [13] M. Hailemariam, B. Goertzel, and T. Yohannes. Evolving 3d facial expressions using interactive genetic algorithms. In *International Conference on Advances of Science and Technology*, pages 492–502. Springer, 2018.
- [14] I. Harvey. The microbial genetic algorithm. In *European conference on artificial life*, pages 126–133. Springer, 2009.
- [15] F.-C. Hsu and P. Huang. Providing an appropriate search space to solve the fatigue problem in interactive evolutionary computation. *New Generation Computing*, 23(2):115–127, 2005.
- [16] R. E. Jack, O. G. Garrod, H. Yu, R. Caldara, and P. G. Schyns. Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 109(19):7241–7244, 2012.
- [17] K. L. Johnson, M. Iida, and L. G. Tassinari. Person (mis) perception: Functionally biased sex categorization of bodies. *Proceedings of the Royal Society B: Biological Sciences*, 279(1749):4982–4989, 2012.
- [18] S. Katoch, S. S. Chauhan, and V. Kumar. A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, 80(5):8091–8126, 2021.
- [19] H.-S. Kim and S.-B. Cho. Application of interactive genetic algorithm to fashion design. *Engineering applications of artificial intelligence*, 13(6):635–644, 2000.
- [20] C.-C. Lai and Y.-C. Chen. A user-oriented image retrieval system based on interactive genetic algorithm. *IEEE transactions on instrumentation and measurement*, 60(10):3318–3325, 2011.
- [21] S. Li and W. Deng. Deep facial expression recognition: A survey. *IEEE transactions on affective computing*, 2020.
- [22] D. J. Lick, C. M. Carpinella, M. A. Preciado, R. P. Spunt, and K. L. Johnson. Reverse-correlating mental representations of sex-typed bodies: the effect of number of trials on image quality. *Frontiers in psychology*, 4:476, 2013.
- [23] R. F. Murray. Classification images: A review. *Journal of vision*, 11(5):2–2, 2011.
- [24] A. D. Piemonti, M. Guizani, M. Babbar-Sebens, E. Zhang, and S. Mukhopadhyay. Exploration and visualization of patterns underlying multistakeholder preferences in watershed conservation decisions generated by an interactive genetic algorithm. *Water Resources Research*, 57(5):e2020WR028013, 2021.
- [25] E. Ponsot, J. J. Burred, P. Belin, and J.-J. Aucouturier. Cracking the social code of speech prosody using reverse correlation. *Proceedings of the National Academy of Sciences*, 115(15):3972–3977, 2018.
- [26] A. Pumarola, A. Agudo, A. M. Martinez, A. Sanfeliu, and F. Moreno-Noguer. Ganimation: Anatomically-aware facial animation from a single image. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.
- [27] J. A. Russell and L. F. Barrett. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology*, 76(5):805, 1999.
- [28] T. Souaille, J.-F. Petiot, M. Lagrange, and N. Misdariis. Extracting design recommendations from interactive genetic algorithm experiments: Application to the design of sounds for electric vehicles. *Proceedings of the Design Society*, 1:1567–1576, 2021.
- [29] K. Sweeney and C. Whissell. A dictionary of affect in language: I. establishment and preliminary validation. *Perceptual and motor skills*, 59(3):695–698, 1984.
- [30] H. Takagi. Interactive evolutionary computation: Fusion of the capabilities of ec optimization and human evaluation. *Proceedings of the IEEE*, 89(9):1275–1296, 2001.
- [31] T. Wang and M. Zhou. A method for product form design of integrating interactive genetic algorithm with the interval hesitation time and user satisfaction. *International Journal of Industrial Ergonomics*, 76:102901, 2020.
- [32] M. Westerlund. The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9(11), 2019.
- [33] H. Yu, O. G. Garrod, and P. G. Schyns. Perception-driven facial expression synthesis. *Computers & Graphics*, 36(3):152–162, 2012.