Biologically Inspired Cognitive Architectures 24 (2018) 107–114

Contents lists available at ScienceDirect

Biologically Inspired Cognitive Architectures

journal homepage: www.elsevier.com/locate/bica

# An embodied virtual agent platform for emotional Stroop effect experiments: A proof of concept

A. Oker[a,*], N. Glas[a], F. Pecune[b], C. Pelachaud[c]

[a] Institut Mines Télécom, ParisTech, Paris, France
[b] ArticuLab, Carnegie Mellon University, Pittsburgh, United States
[c] The Institute for Intelligent Systems and Robotics, UMR7222 CNRS, Pierre and Marie Curie University, Paris, France

ABSTRACT

The need for experimentation of facial expression recognition in a more ecological manner necessitates the use of multimodal, interactive experimental stimuli. At the same time, the prerequisite of reproducibility of results and controlled conditions is still mandatory. An embodied conversational agent (ECA) is a pertinent framework that meets all these requirements. The VIB (Virtual Interactive Behavior) Platform is a SAIBA compliant system which supports the real-time generation of multimodal behavior for interacting with socio-emotional virtual agents. We created a new feature for this platform, namely VIB-Ex, which can be used for presenting real-time facial expressions and recording the user's reaction time and interaction while exporting data for statistical purposes. In this paper, we present our proof of concept study in which a 3D male virtual character has been used to convey joyful or sad facial expressions. At the same time, the same character pronounced joyful or sad words in congruence or incongruence with its facial expression in order to trigger an emotional Stroop effect. Only 12 adults were sufficient in order to obtain an emotional Stroop effect within our virtual agent. The results of this study confirmed that the VIB-Ex platform can replicate a robust effect of psychological phenomena concerning recognition of facial expressions. VIB-Ex proves itself to be a suitable and a pertinent tool to perform experiments on a human's automatic process of facial expression recognition. Finally, we discuss the possible future research topics with VIB-Ex to carry out other type of experiments in the field of social cognition.

## Introduction

In cognitive sciences, like any other empirical research, reductionism is the main investigation method. Every aspect of human behavior is reduced to its smaller scales and modules in order to investigate its function on the central nervous system. Historically, every cognitive task or process is deliberately dissected from the 'whole' system and used in this manner in experimental paradigms. The same logic applies for the modality of stimuli; although richer modality can be used in cognitive sciences, the major trend remains largely the use of unimodal stimuli in experimentation. From behaviorism of the early 19th century, this approach has the merits of discovering most structural and functional properties of the cognitive systems, such as memory, human language or vision, as well as providing most intriguing insights to some neurological (hemispatial neglect, Alzheimer's…) and psychiatric conditions (schizophrenia, autism…).

However, following Frith's works in the early 90's regarding the theory of mind (ToM; Frith, 1992) and the discovery of mirror neuron system (MNS) by Rizolatti and collaborators (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996), the very unique idea of social cognition as being qualitatively different to any other neurocognitive process has been brought to the daylight. In parallel, the field of emotion recognition research from the facial expressions has been developed (Ekman & Friesen, 1978; Russell, 1994). These three research subjects (ToM, MNS and recognition of facial expressions) have been generally admitted to constitute the core of the social cognition (Van Overwalle, 2009). Thus, the social cognition is what refers "to the mental operations underlying social interactions, which include processes involved in perceiving, interpreting, and generating responses to the intentions, dispositions, and behaviors of others" (Green, Olivier, Crawley, Penn, & Silverstein, 2005).

This is why, following the emergence of social cognition as a whole new field of research, some authors advocated for the need that stimuli used in experiments should be similar to real life as much as possible. In cognitive science, the term "ecological" is used for stimuli which is more multimodal, interactive and similar to real-life experience rather than the unimodal, non-interactive and representational stimuli used in the laboratory environment. As a matter of fact, in 1979 James J.

* Corresponding author.
*E-mail address:* Dr.alioker@gmail.com (A. Oker).

https://doi.org/10.1016/j.bica.2018.04.011
Received 8 December 2017; Received in revised form 25 April 2018; Accepted 26 April 2018

Gibson already used the term "ecological psychology" for the study of how perception is built by prior visual experience and how perception involves the transmission of sensory stimulation to a passive observer to the idea of *perception–action* systems exploring an information-rich environment. Starting with this view of the perceiver as embodied agent, and perceiving as the functioning of perceptual systems, Gibson explores the questions of what is the nature of the environment that is perceived and how are its functionally significant properties detected. Thus, the word "ecological" is commonly used by cognitive science researchers whom use real life-like stimuli in their experiments.

In this paper,

1) We will show that our embodied conversational agent platform the VIB – Virtual Interactive Behavior – (Pecune, Cafaro, Chollet, Philippe, & Pelachaud, 2014) is biologically inspired thanks to its Saiba compliance.
2) We will present a new module namely the "VIB-Ex" in this architecture which can be used for procedural experimentations in social cognition such as emotional Stroop effect.
3) We will present how emotional Stroop effect can be measured in a more ecological way using embodied conversational agents using our platform. The field of facial recognition is one of the most investigated research subjects in social cognition therefore new technologies as the embodied conversational agents can be a real asset in new research avenues. In order to study a facial expression phenomenon, we have chosen emotional Stroop effect because it forces necessarily and automatically participants to recognize facial expressions. Thus, finally we will present our proof of concept which is an empirical study used to see whether our VIB-Ex module can trigger automatic recognition of a virtual agent's facial expressions evaluated by the means of quantitative and inferential statistics.

Our purpose is to provide arguments that an embodied conversational agent (The VIB, Pecune et al., 2014) module we've developed in order to perform behavioral experiments the "VIB-Ex" can provide more naturalistic experimental settings. To do so, we have conducted a validation study for replicating robust psychological phenomena already being brought to the daylight in cognitive sciences: an emotional Stroop effect. Therefore, our primary objective is to conduct a proof of concept of the VIB-Ex module while creating a Stroop effect within an embodied conversational agent. Before presenting the experiment, firstly we will present our platform, then why Stroop effect is a good candidate to evaluate the ability to perceive facial expression automatically and why Stroop effect can be explained by mirror neuron system triggered by a biologically inspired virtual agent architecture.

## Facial expression recognition as a complex cognitive process

The ability to understand other's emotion and inner states (the theory of mind) and emotions require recognition of facial expressions undoubtedly. As Frijda (1988) pointed out that while "emotions and feelings are often considered the most idiosyncratic of psychological phenomena…. It can be described in terms of a set of laws" (pg. 349). Following this idea, Ekman (1992) and Izard (1992) propose in accordance with an evolutionary psychology perspective that recognition and expression of an emotion are fundamental in the social behavior of human species and they can be described and characterized. Consequently, information processing demands made by social cognition are different from those made by non-social cognition. According to Adolphs (2001), "Compared to the physical environment in general, the social environment is more complex, less predictable, and, critically, more responsive to one's own behavior." Thus, as a communicative intent or as an expression of inner state, human expressions plays a fundamental and complex role. By the neuroscientific perspective, neural structures involved in social cognition, even particularly in facial expression recognition are multiple and dynamic. For instance, amygdala plays an important role for the appraisal of threat; it is activated when normal subjects view facial expressions of fear but equally in processing the direction of gaze of others (Baron-Cohen et al., 1999) while the insular cortex is involved in disgust and self-consciousness (Phillips & et al., 1997). However, amygdala and the insular cortex does not act alone. Regarding the faces; inferotemporal cortex; particularly superior temporal sulcus and gyrus fusiform area are primarily selective for facial expressions and face detection but also for intended action recognition based on those (Allison, Puce, & McCarthy, 2000). Finally, another region of great interest; the inferior parietal lobule and the caudal sector of the inferior frontal gyrus and the adjacent part of the premotor cortex, has been highlighted for involving neural mechanisms which allow us to directly understand the meaning of the actions and emotions of others by internally replicating ('simulating') them without any explicit reflective mediation (Gallese, Keysers, & Rizzolatti, 2004). According to these authors, when we witness the disgusted facial expressions of someone else, this network helps to activate that part of our insula that is also active when we experience disgust for example. While the cognitive neuroscience of facial expression recognition is not the purpose of our work, it seems clearer that this is a highly demanding cognitive process and researchers in social cognition are demanding the experimental tasks which can help them to investigate these networks involving facial expressions.

In this field, the literature seems to agree that there would be seven fundamental facial expressions: Joy, fear, anger, surprise, sadness, disgust and contempt. Ekman and Friesen (1978) introduced the idea that there would be a prototypical law of features for these fundamental expressions which would be also universal (Ekman, 1994). Of great importance for embodied conversational agents that we present in this paper; Ekman and Friesen (1978) also proposed a taxonomic system of facial muscles named FACS (Facial Action Coding System) which would standardize physical expression of emotions. These codes called 'Action Units' is what is being used to animate most conversational virtual agents today, but not all.

The photography of stimuli provided by Ekman (1993) started mostly to study facial expression recognition impairments of several social cognition disorders in psychiatry. For instance, schizophrenia patients seem to be very impaired to this task (Edwards, Jackson, & Pattison, 2002; Kohler, Walker, Martin, Healey, & Moberg, 2009), as well as autism spectrum disorders (Harms, Martin, & Wallace, 2010) and in natural of pathological aging individuals (Chaby & Narme, 2009). However, most of these studies use unimodal stimulus (e.g. pictures) and lack to investigate ecological nature of facial expression. On this subject, Keysers and Gazzola (2007) noted that "much of the debate in social cognition might result from choosing tasks that isolate the processes of just one route in the laboratory. However, it is essential to start designing tasks that reflect the complexity of social life to test how the social brain forms an integrated whole." Therefore, we believe that Embodied Conversational Agents (ECA) are a pertinent tool for facial recognition research.

## The pertinence of the use of ECA's on facial expression experiments

The pertinence of the use of ECA's in social cognition research has been already highlighted (Gratch, 2014). According to Gratch, "virtual humans aspire to simulate the cognitive abilities of people, but also many of the "embodied" aspects of human behavior, more traditionally studied in fields outside of cognitive psychology, such as nonverbal behavior recognition and production".

The need to investigate social cognition in a more ecological manner necessitates the use of multimodal, interactive experimental stimuli (Oker, Courgeon et al., 2015, Oker, Prigent et al., 2015) called naturalistic settings in psychology experiments. In the late 1980s, Ickes and coworkers already suggested the use of the term naturalistic social cognition for studies in which the experimental conditions were socially