Pattern Recognition Letters 000 (2015) 1–11

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Needs and challenges in human computer interaction for processing social emotional information ☆

Anna Esposito [a,b,*], Antonietta M. Esposito [c], Carl Vogel [d]

[a] Seconda Università di Napoli, Dipartimento di Psicologia, Viale Ellitco 31, Caserta 81100, Italy
[b] International Institute for Advanced Scientific Studies (IIASS), Via Pellegrino 19, Vietri sul Mare 84019, Italy
[c] Osservatorio Vesuviano, Viale Claudio, Napoli, Italy
[d] Centre for Computing and Language Studies School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Emotional expressions
Multimodal communication
Needs and challenges in emotionally and believable ICT interfaces

ABSTRACT

Demand for and delivery so far of sophisticated computational instruments able to recognize, process and store relevant interactional signals, as well as interact with people, displaying suitable autonomous reactions appropriately sensitive to environmental changes, have produced great expectations in Information Communication Technology (ICT). Knowing what an appropriate continuation of an interaction is depends on detecting the addresser's register and a machine interface unable to assess differences will have difficulty managing interactions. Progress toward understanding and modeling such facets is crucial for implementing behaving Human Computer Interaction (HCI) systems that will simplify user access to future, profitable, remote and nearby social services. This paper raises new research questions via new means for socio-behavioral and emotional investigations, and suggests the gathering of new experimental data and theories across a spectrum of research concepts, in order to develop new psychological and computational approaches crucial for implementing believable and trustable HCI systems which exploit synthetic agents, robots, and sophisticated humanlike interfaces.

1. Introduction

Emotional feelings permeate our everyday experience, consciously or unconsciously driving our daily activities and constraining our perception, actions and reactions.

During daily body-to-body interactions, our ability to encode and decode emotional expressions plays a vital role in creating social linkages, producing cultural exchanges, influencing relationships and communicating meanings. Emotional information is simultaneously transmitted through verbal (the semantic content of a message) and nonverbal (facial and vocal expressions, gestures, paralinguistic information and so on) communicative tools and relations and exchanges are highly affected by the way this information is coded/decoded by/from the addresser/addressee as well as by the contextual instance and environmental conditions. Research devoted to the understanding of the perceptual and cognitive processes involved in the encoding/decoding of emotional states during interactional exchanges is particularly relevant for building and maintaining human relationships as well as developing friendly and emotionally colored assistive technologies.

The role of emotional communication in the context of Human Computer Interaction (HCI) is even more relevant and challenging when considering the varied technological application areas, from e-teaching and e-learning to office applications, entertainment, remote health care and assistive technologies, robotics, advertising and design of new technological products. Continual areas of interest within such research include the automatic detection, as well as the synthesis of affect and emotion through facial, bodily and vocal signals.

Given the complexity and the multimodal nature of the phenomenon, there has been a branching of engineering approaches toward the improvement and development of automatic video–audio processing, detection and synthesis techniques [23,61] with the goal of developing advanced mathematical models and algorithms for encoding/decoding emotional states from faces [27,65,102,130], speech [2,5,6,11,131] and/or body movements [16,24,77,79]. In order to succeed, the above research lines are seeking for deeper investigations and analyses of human interactional behaviors. The aims are to detect relevant verbal and nonverbal behavioral features making successful and cooperative human interactions in order to detail the emotional encoding/decoding process exploited by humans and devise mathematical models to algorithmically implement socially and believable ICT interfaces.

☆ This paper has been recommended for acceptance by Friedhelm Schwenker.

* Corresponding author. Tel.: +39 338 1829988; fax: +39 089 761189.
E-mail address: anna.esposito@unina2.it, iiass.annaesp@tin.it (A. Esposito).

http://dx.doi.org/10.1016/j.patrec.2015.02.013
0167-8655/

In this context, the term "expression" is used to indicate nonverbal emotional information that can be captured and analyzed in a dyadic/group multi-sensory interaction and is restricted to the sensory appearance of emotional states, i.e. facial, vocal and gestural expressions of emotions.

A long research tradition has tackled the study of emotional expressions and related perceptual cues to encode and decode them, through separate investigations into the expressive domains of facial/vocal expressions and body movements. The following three sections (Sections 2–4) summarize the current main findings on the perceptual cues supposed to be exploited by humans to encode/decode facial, vocal and gestural expressions of emotions in these three domains, highlighting the strength and weakness of the related research. The reported considerations lead to the conclusion that emotions cannot be separated from communication, and communicative signals that bring meanings to the interactional instance, also bring emotional contents. Therefore, assuming that communication is always driven by emotional and motivational urges which allow humans to rapidly and continuously adapt to the dynamic changes it produces on our knowledge acquisition and use, Section 4 will highlight the need and challenges it poses on HCI. Section 5 is dedicated to conclusions.

2. On the emotional side

2.1. Facial displays or universal facial emotional expressions

Facial expressions play significant communicative functions. Changes in facial muscles are brought into play for disambiguating meanings, controlling the conversational flow, expressing emotional states, and guiding the speaker/listener's intentions. Among other things, faces can signal interest/disinterest, distress, psychopathologies, cognitive activities, intentions, physical efforts, as well as evoke sympathy and empathy, and describe different personalities. Bavelas and Chovil [9, p. 334] affirmed that "*facial displays … are actively symbolic components of integrated messages*" (including words, intonation and gestures) suggesting that they semantically and pragmatically contribute to the communicative plan. Slocombe et al. [116] and Parr et al. [91] brought evidence of facial communicative functions also for non-human primates. There has been a strong commitment by some authors [9,10,30] in distinguishing between facial and affective displays distinguishing the communicative facial expressions from the affective ones. It has been emphasized that facial displays "*do not express inner emotions; rather, they convey meaning to another person*" (Bavelas, web.uvic.ca/psyc/bavelas/Facial_displays.htm – last verified December 2014).

Communicative facial expressions[1] are therefore assumed to be different from "affective facial displays" that are biologically based on emotional mechanisms[2] clearly implying intercultural universality of the emotional facial expressive capacity. Accordingly a small, fixed number of discrete ("basic") emotions drive our primary motivational system acting as amplifiers of the incoming external, as well as inner signals [30,58]. Each basic emotion can vary in intensity and consists of an "innate affective program" whose triggering produces all the various detectable and undetectable manifestations of such an emotion, including changes in the facial musculature and therefore its corresponding facial emotional expression.

This small set of emotional categories includes happiness, anger, sadness, and fear, which can be reliably associated to basic survival problems such as nurturing offspring, earning food, competing for resource, avoiding and/or facing dangers. In this context, basic emotions are brief, intense and adapted reactions to urgent and demanding survival issues. These reactions to goal-relevant changes in the environment require "readiness to act" and "prompting of plans" in order to appropriately handle (under conditions of limited time) the incoming event, producing suitable mental states, physiological changes, feelings and expressions.

Both the categorization and the psychological universality of the emotional coding and decoding processes are currently debated among researchers. Alternative to the "basic emotion assumption", dimensional models have been proposed [106]. Such models assume that emotions can be decomposed into a finite set of primary features (dimensions) and suggest that different combinations of such features can produce different affective states. These generative models, bringing the dimensional concept to an extreme, suggest that, the number of primary features can be extended along a continuum, and generate a (possibly) infinite number of affective states.

This idea, even though intriguing, clashes with the economy principle that seems to rule natural dynamic systems and raises the question of whether it is possible to perceive a continuum of affective states. Categorization in fact allows humans to make fast associations and perform an efficient and rapid decision making process facilitating the handling of unexpected events. Emotional categories may therefore be favored in order to avoid excessive processing time. In addition, some of the suggested dimensions are seen to behave more as categories (see for example the "valence" concept that can be considered as positive or negative) than to vary along a continuum. Furthermore, the discrete basic emotion evolutionary perspective has been supported by several findings such as the discovery of an emotion specific Autonomic Nervous System's (ANS) activity [70] and of distinct brain regions tuned to handle basic emotions [26], as well as by the other mammalians' ability to code/decode basic emotional expressions [89] and the observation that some emotional expressions (such as smiling, amusement, and irritability) are universally expressed by infants, adults, blind and sighted independently of race and culture [29,58].

For this small set of primary emotions, Ekman [29,30] and Izard [58] provided a detailed description of how the contraction of each facial muscle (taken singularly and/or in combination with others) modifies facial appearance when such an emotion is experienced. These detailed descriptions lead to the definition of several anatomically based facial coding systems among these the Facial Action Coding System (FACS, [28]) and the AFFEX [57] coding system. The proposed coding systems measure facial muscle changes and identify for each primary emotion a facial anatomical configuration expressed as a combination of facial muscle modifications. However, despite the universal idea driving their definition, these coding systems are different both in the set of primary emotions they included (only six for Ekman – happiness, sadness, anger, fear, surprise, and disgust; 10 for Izard that added to the six contempt, guilt, interest and shame) and in their corresponding muscle changes they use to code the same primary emotional facial expression.

On the other hand, demonstrations of above-chance accuracy in decoding emotional facial expressions that support the idea of universally shared emotional facial expressions do not seem to be pervasive. Some authors maintain that expressions of emotions (in particular faces) are learned to efficiently and effectively express intentions and negotiate relations [42,43,105]. In addition facial expressions are dependent upon context, personal motives and the character and direction the ongoing social exchanges are taking. Thus they vary across cultures. This hypothesis was further supported by the fact that sophisticated measurements, such as facial electromyography [19,109] assessing facial muscle changes (even when emotional information was not visually perceptible) proved that distinction among primary, and more generally, among negative and positive emotions (debating also the dimensional universality suggested by Russell [105]) was not possible.

Recent theoretical models have attempted to account for both universality and cultural variations by specifying which particular

[1] Inquisitive eyebrow lifting is an example.

[2] The oral–nasal nausea-sympathetic response of disgust is quintessentially biological in basis [62].

Please cite this article as: A. Esposito et al., Needs and challenges in human computer interaction for processing social emotional information, Pattern Recognition Letters (2015), http://dx.doi.org/10.1016/j.patrec.2015.02.013