Available online at www.sciencedirect.com
ScienceDirect
Computer Speech and Language 25 (2011) 45–62
COMPUTER SPEECH AND LANGUAGE
www.elsevier.com/locate/csl

Automatic inference of complex affective states

Tal Sobol-Shikler *

Computer Laboratory, University of Cambridge, UK

Received 5 May 2008; received in revised form 23 March 2009; accepted 24 December 2009
Available online 4 January 2010

Abstract

Affective states and their non-verbal expressions are an important aspect of human reasoning, communication and social life. Automated recognition of affective states can be integrated into a wide variety of applications for various fields. Therefore, it is of interest to design systems that can infer the affective states of speakers from the non-verbal expressions in speech, occurring in real scenarios. This paper presents such a system and the framework for its design and validation. The framework defines a representation method that comprises a set of affective-state groups or archetypes that often appear in everyday life. The inference system is designed to infer combinations of affective states that can occur simultaneously and whose level of expression can change over time. The framework considers also the validation and generalisation of the system. The system was built of 36 independent pair-wise comparison machines, with average accuracy (tenfold cross-validation) of 75%. The accumulated inference system yielded total accuracy of 83% and recognised combinations for different nuances within the affective-state groups. In addition to the ability to recognise these affective-state groups, the inference system was applied to characterisation of a very large variety of affective state concepts (549 concepts) as combinations of the affective-state groups. The system was also applied to annotation of affective states that were naturally evoked during sustained human–computer interactions and multi-modal analysis of the interactions, to new speakers and to a different language, with no additional training. The system provides a powerful tool for recognition, characterisation, annotation (interpretation) and analysis of affective states. In addition, the results inferred from speech in both English and Hebrew, indicate that the vocal expressions of complex affective states such as thinking, certainty and interest transcend language boundaries.

Keywords: Affective computing; Affect recognition; Cognition; Emotions; Human perception; Intelligent systems; Machine learning; Multi-label inference; Multi-modal analysis; Multi-modal database; Speech analysis; Speech corpora

1. Introduction

Affective states and their non-verbal expressions are important aspects of human reasoning, decision making and communication. Recognition of affective states can be integrated into fields such as human–computer interfaces and interactions (HCI), human–robot interactions (HRI) and speech technologies. The recognition

* Present address: Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Marcus Campus, POB 653, Beer-Sheva 84105, Israel. Tel.: +972 52 4042706.
E-mail address: tal.shikler@gmail.com

0885-2308/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.csl.2009.12.005

can enhance such systems and user performance and has many potential applications (Reeves and Nass, 1996; Picard, 1997; Becker et al., 2007). Recognition results can be used for analysis of user reactions in order to predict intentions and to generate appropriate response. It can also be used for annotation of speech corpora for synthesis of affective speech. In order to achieve that, the systems designed should be able to infer affective states occurring in real scenarios. The development of such systems entails collection and labelling of speech corpora (Douglas-Cowie et al., 2003), development of signal processing and analysis techniques, as well as consolidation of psychological and linguistic analyses of affective states (Cowie et al., 2001).

In this paper the term *affective states* refers to emotions, attitudes, beliefs, intents, desires, pretending, knowledge and moods. Their expression reveals additional information regarding the identity, personality and physiological state of the speaker, in addition to context-related cues and cultural display rules. This wide definition of the term *affective states* draws on a comprehensive approach to the role and origin of emotions (Cowie et al., 2001; Cornelius, 2000): affective states and their expressions are part of social behaviour (Whiten, 1991; Baron-Cohen, 1999), with relation to physiological and brain processes (James, 1884; Bechara et al., 1997). They comprise both conscious (Scherer, 1993) and unconscious reactions (Zajonc, 1980; den Noort et al., 2005; Bechara et al., 1997), and have cause and effect relations with cognitive processes such as decision making (Kahneman and Tversky, 1979; Bechara et al., 1997). A number of affective states can co-occur simultaneously (Scherer, 1995; Haynes and Rees, 2006; Slors, 2001), and change dynamically over time. A similar definition of the concept affective states is given by Höök (2004) who describes affect as human, rich, complex and ill-defined *experience*.

The affective states are inferred from their non-verbal expressions. The term *expression* refers here to the outward representation of the affective states. This is the observable behaviour (conscious or unconscious) that people perceive and interpret. It can be affected by factors such as context and cultural display rules.

There are three main approaches to the design of affect recognition systems. These approaches are used for inference from expressions in speech and in other behavioural cues such as facial expressions. The most commonly used approach (Petrushin, 1999; Oudeyer, 2003; Fernandez and Picard, 2005) is to infer a small set of *basic emotions* (Ekman, 1999; Darwin, 1898), such as *happy*, *sad*, *angry*, *afraid*, *disgusted* and *surprised*. The term refers to qualitatively distinct states that are held to be universal at least in essence, i.e. recognisable by most people from most backgrounds, and associated with brain systems that evolved to cope with various situations. These affective states have relatively clear definitions, although even within this set the need for finer definitions has been addressed, for example distinguishing between *cold anger* and *warm anger* (Cowie et al., 2001; Cornelius and Cowie, 2003; Cornelius, 2000; Vidrascu, 2007). Stereotypical expressions of these affective states are perceived as easier to act and to recognise, and therefore useful for both quick acquisitions of data-sets, and as a starting point for an emerging research field. However, these affective states do not encompass the entire range of human affective states and (in most applications) do not relate to nuances of expression. If the small set is used only as a starting point, it is an open question whether the same behavioural cues are used for both extreme emotions and subtle expressions of complex mental states. The inference is of a single emotion for each analysed sentence so the systems encompass a small set of affective states. This approach is a limited version of a broader perspective called the *categorical* representation method, in which additional affective states are defined, either as a blend of basic emotions, or in conjunction with other cognitive processes (Plutchik, 2001).

The second approach is to detect the existence of a selected affective state in real situations, such as drivers' stress, attempts at insurance fraud or post-natal depression (Fernandez and Picard, 2003; Nemesysco Ltd., 2006; Moore et al., 1994). This method is not used in order to recognise which affective states are expressed in the speech, but rather to detect whether a certain affective state exists or not. It does not refer to other co-occurring affective states or to different levels of the expression.

The third approach, which has recently become more widespread, is the *dimensional* approach, in which several expressions are identified each on a one-dimensional (1-D), two (2-D) or three dimensional (3-D) space, with dimensions such as passive–active, positive–negative and low–high arousal levels (Cowie et al., 2001; Schröder, 2004; Scherer, 1984; Whissell, 1989; Kim, 2007; Grimm and Kroschel, 2007; Hoque and Louwerse, 2006). The dimensional approach provides in theory a more continuous scale for interpretation but the research usually refers to recognition of the edges or areas, for example: positive and low arousal or negative and high arousal level. These descriptions are often correlated to physiological processes such