Chapter 7
Chatterbox Challenge as a Test-Bed for Synthetic Emotions

Jordi Vallverdú
Universitat Autònoma de Barcelona, Spain

Huma Shah
Universitat Autònoma de Barcelona, Spain

David Casacuberta
Universitat Autònoma de Barcelona, Spain

ABSTRACT

Chatterbox Challenge is an annual web-based contest for artificial conversational systems, ACE. The 2010 instantiation was the tenth consecutive contest held between March and June in the 60th year following the publication of Alan Turing’s influential disquisition ‘computing machinery and intelligence’. Loosely based on Turing’s viva voca interrogator-hidden witness imitation game, a thought experiment to ascertain a machine’s capacity to respond satisfactorily to unrestricted questions, the contest provides a platform for technology comparison and evaluation. This paper provides an insight into emotion content in the entries since the 2005 Chatterbox Challenge. The authors find that synthetic textual systems, none of which are backed by academic or industry funding, are, on the whole and more than half a century since Weizenbaum’s natural language understanding experiment, little further than Eliza in terms of expressing emotion in dialogue. This may be a failure on the part of the academic AI community for ignoring the Turing test as an engineering challenge.

INTRODUCTION

In his anticipation of objections to the idea of machines thinking, and testing for it through an imitation game, Alan Turing reminded of a real-life scenario the viva voca in which an interrogator seeks answers to questions from a ‘witness’ (Turing, 1950). Pre-empting the argument from consciousness and quoting from Jefferson’s 1949 Lister Oration, “not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt ... not only write it but know that it had written it” (1950, section 6, p. 445), Turing
countered showing this stance was a solipsistic one. To say that “no mechanism ... could feel pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants” (p. 446), was, according to Turing, an extreme position: “the only way by which one could be sure that a machine thinks is to be the machine and to feel oneself thinking” - Turing’s emphasis (ibid). Turing replied, as Stins and Laureys put it “in a succinct British fashion” (2009, p. 265) that rather than labouring over the point “A is liable to believe A thinks but B does not while B believes B thinks but A does not” it is “usual to have the polite convention that everyone thinks” (1950, p. 446).

There are a variety of objections to the “Turing test” and whether it can really be a way to assert whether a computer is thinking or not, it is beyond the scope of this paper to review them, readers are directed to Shah and Warwick (2010b) nevertheless the concept of being able to pass the Turing test is a complex endeavour indeed, and analysing such a variant Turing test contest can be of great philosophical value.

Turing put forward possible questions and answers to show that if the responses were “satisfactory and sustained” then one might not describe the answers as “an easy contrivance” (1950, p. 447). It is this method to assess whether machine responses to questions are satisfactory and sustained that is evaluated in an annual contest – the Chatterbox Challenge. Artificial conversation systems (ACE), commonly known as ‘chatbots’ compete against each other across a number of categories. What is considered a satisfactory response can be subjective; one interrogator may find a response inappropriate while another may accept it as humorous. An answer to a question may seem ersatz-like, but may also be a satisfactory emotive response under interrogation or conveying disinterest in topic by the ‘witness’. Whether machines are capable of expressing emotion through their responses or whether they are still Eliza-like (Weizenbaum, 1966) can be found by analysing the contest’s transcripts.

**Emotion: Mood, Feeling or Expression?**

According to Broekens “Emotion is a complex topic, and agreement on one solid definition does not really exist” (2010). What does stand as an ‘emotion’ is also not clear cut (Barrett et al., 2007), and more if we take into account the qualia aspects of emotions, their feeling, we need to paraphrase Augustine of Hippo’s ideas on his subjective theory of time telling that “What then are emotions? If no one asks me, I know: if I wish to explain it to one that asked, I know not”.

In previous studies by two of the current authors (Vallverdú & Casacuberta, 2008, 2009), some neurological, philosophical and computational approaches to the analysis of natural or synthetic emotions were described. It is beyond the scope of the present paper to explain again all this information, but adding some basic different ideas such as Llinás (2001) and Damasio (1999), we provide a more sophisticated approach to the meaning of emotion from an evolutionary perspective. From our new and original point of view, emotions are:

1. **Embodied-intentionality generators**: with a genotypic punishment/reward management system pre-encoded in our bodies, emotions transforms bodies into goal-oriented complex systems. Basically, they are: feeding, self-preservation and reproduction (all these activities involve movement). Beings are not just things in the world, but things with an evolutionary force embedded in them. Under this embodied frame, we are emotionally oriented toward certain events of the world. In a certain way, emotions emerge from an embodied neurological a priori focus for the logic of survival. Interpreting in a different way the existentialist motto “existence pre-