

The Challenge of Recognizing Emotions Using Artificial Intelligence

Originally published June 26, 2017 at <https://www.aiia.net/artificial-intelligence/articles/the-challenge-of-recognizing-emotions-using>



'This call may be recorded for [machine] training purposes'

Analyzing our feelings is among the primary interests of Artificial Intelligence research, not least as it relates to the fields of security and customer research analysis, including the development of an individual's 'back-story' through their engagement with a company or corporate entity.

The two research areas currently receiving the most attention in the field of 'live' sentiment analysis are facial analysis and voice analysis. Both share a similar problem in terms of a priori information, as both are likely to work better with an established baseline — an understanding of when a person is deviating from repose and engaging in strong emotion.

In the case of voice analysis, AI-derived systems are progressing in their attempts to recognize aberrant states, for instance by understanding when a customer using an automated system is [becoming angry or upset](#) and may need to be transferred to a human operative.

Graven tones

Fujitsu Laboratories has undertaken extensive research into 'voice cheerfulness', claiming to have achieved high accuracy in understanding customer sentiment throughout a conversation, assuming the entire interaction is long enough to get that all important baseline reading. In a field trial using Machine Learning-monitored conversations with customers, Fujitsu was able to improve training times and outcomes for both the customer and respondent by 30%. The [2016 research](#) was due to be incorporated into the company's 'Human Centric' AI system [Zinrai](#).

Establishing the baseline is not necessarily enough. That year Australian researchers also [published](#) work relating to the growing field of context-aware 'mood mining', which emphasised the extent to which our apparent emotional responses are 'guided' by the person with whom we are talking:

'When the user takes part in a phone conversation the system will use the emotional construct of the speech of the person at the other end, the listener, as the "contextual information". If the listener is talking about an exciting event the user is expected to be excited or cheerful if he/she is in a positive mood, otherwise, it would be assumed that the user is in a negative mood.'

Hostile by default?

The context presents an additional possible challenge for AI-derived algorithms attempting to understand emotion (or actual neural networks processing archive data), since the user may actually enter an encounter 'in a bad mood' atypical of their regular manner; a situation that can be extremely typical of customers steeling themselves for a complaint procedure, for instance.

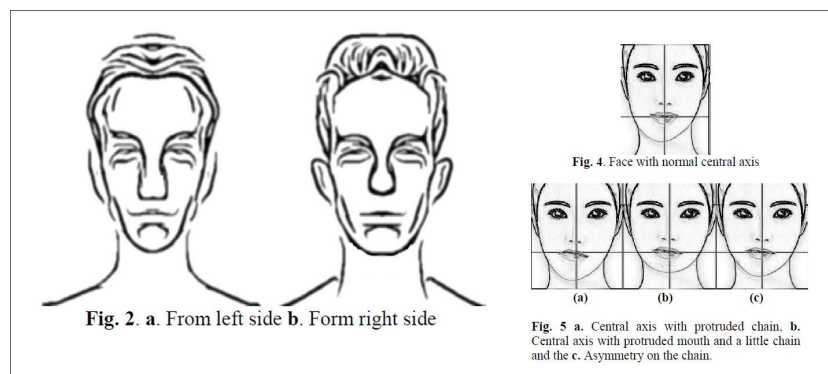
This raises various questions: should a mood analysis system assume a default 'mood deviation' from the first encounter with the customer because of the very nature of the interaction? If the system is also (or entirely) using facial sentiment analysis, is it capable of understanding that some people have [naturally sad or hostile faces in repose](#) — or that this particular individual might?

As if these were not adequate challenges, guile — or the 'poker face' play — also presents further obstacles to using either vocal tone or facial expression as a guide to sentiment. The ability to control visible emotion remains an [index of business acumen](#) whether at the card table or the board table.

Lopsided view

Once you enter the sphere of identifying mood via facial expression, an AI system is further challenged with the [natural asymmetry](#) of most people's faces. Asymmetrical facial features characterise facial expressions (most popularly bewilderment or pensiveness) as well as natural passive asymmetry.

This week researchers from Kent State University have added (perhaps literally) a new wrinkle to the general work around identifying facial expressions by [applying dihedral group theory](#) to the task.



The Machine Learning system involved in the research has to evaluate and distinguish any natural facial symmetry that the user possesses from the way that they use their facial geometry to signal mood or other conversational tokens.

Automatic symmetry detection is a venerable subset of research in this field, with Group Theory — the study of how sections of the face are aligned together — part of the well-established mathematical methods used to assess symmetry.

A piece of the action

In assigning emotion to various facets of the face, the various facial sections are considered in order of the priority of a points-based system, with eyebrows and lips rated at 3 points as indicators, and the middle of lips and eyes counting as four points (the assignment of importance of these facets of course represents a median distribution of expressive indicators across most people; certainly there are [notable exceptions](#)).

Flipping the image of the face can help to determine which asymmetries are congenital and which voluntary, and in assigning Dihedral groups for monitoring purposes. The work to date has used archival or non-live data, with further planned work intending to use live video data — a step which, common in Machine Learning projects which make use of neural networks, will need to address the ever-critical aspect of latency, with a view to the eventual development of real-time facial analysis systems.

IMAGES: [Wikimedia](#) / [Arxiv](#)