**QUEEN'S
UNIVERSITY
BELFAST**

The challenges of dealing with distributed signs of emotion: theory and empirical evidence

Cowie, R., Douglas-Cowie, N. V., McKeown, G., & Gibney, C. (2009). *The challenges of dealing with distributed signs of emotion: theory and empirical evidence*. 351-356. Paper presented at ACII (IEEE publication), Amsterdam, Netherlands.

Document Version:

Early version, also known as pre-print

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

The challenges of dealing with distributed signs of emotion: theory and empirical evidence

Roddy Cowie
Queen's University, Belfast
Belfast BT7 1NN
r.cowie@qub.ac.uk

Gary McKeown
Ceire Gibney
Queen's University, Belfast
<http://www.psych.qub.ac.uk/>

Abstract

There clearly are important issues associated with the distinction between acted and naturalistic data, but focusing on acting may not be the best way to articulate them. An alternative is to focus on differences of structure which are often (but not always) associated with the distinction. Several such differences relate to the way signs are distributed: simultaneous or sequential appearance; in one modality or many; stable or unstable; suggesting one emotion or many; briefly or continually available. These distinctions clearly relate to different demands on analysis.

1. Introduction

The past decade has seen growing recognition that when it comes to emotion, there are deep differences between the kind of data that can be generated to order in a laboratory, and the kind of data that is likely to be encountered in an application. However, it is not at all clear that the core issue has been satisfactorily identified.

The problem is usually stated in terms of a dichotomy between acted and naturalistic data. However, that raises a multitude of problems, not least because a degree of acting forms part of perfectly normal interactions; and some expressions of emotion, while undoubtedly authentic, have very little resemblance to the expressions of emotions bearing the same name that one might expect to encounter in an application.

A natural alternative to the acted/naturalistic contrast is to consider what are the important structural differences that may distinguish material from different sources, either in the cues that appear, or in the way that similar cues are organized. Very little has been written about those structural differences, though there is evidence on issues that are related to it.

The aim of this paper is to suggest terminology that allows the relevant distinctions to be drawn, to identify some of the structural issues that seem to be relevant, and to present empirical research which gives preliminary pointers to issues that are technologically significant, and can be pursued experimentally.

2. Articulating the relevant distinctions

Everyday experience tells us that there are relatively

gross differences between the expressions of emotion that we expect to meet around us and the expressions that appear in TV dramas. An experiment in our laboratory confirmed that the intuition is well-founded. The study took clips from TV channels, some showing spontaneous expressions of emotion, others showing extracts from dramas in which actors simulated comparable emotions. Naïve raters watched them, and as they watched each clip, they used a slider to indicate which type they believed the clip was, and how confident they were in their judgment. Figure 1 shows the overall picture that emerged. The distinction was not perceived instantly, but ratings of the two types were significantly different within a few seconds, and had stabilized in about 15.

Data like these confirm that there is a core of truth in the intuition that there are major differences between acted and natural expressions of emotion. However, closer inspection indicates that the most interesting contrasts are probably not to do with acting *per se*.

One of the problems is revealed in the study that has just been described. One clip was classified contrary to expectation. It showed an actress being interviewed, and expressing happiness. The experimenters knew that the events she described were real, and that she had every reason to be happy about them; and in that context, they were in no doubt that the feelings were real. Raters, without the context, recognized the heightened expression, and judged that it was acted.

Discussion at a recent conference offered an inverse example. The presenter showed an extract from a database, and observed that it was obviously acted. A member of the audience replied that he saw very similar behaviour on the football terraces every week.

The point of these examples is that the key distinction is not whether material is acted. Deciding whether a sample is appropriate or inappropriate for training does not depend on specialized background knowledge, either about the life of an actress or about football fans. It depends on structural relationships between the sample and material that an application has to deal with.

Similar issues arise in different contexts. The examples above involve relatively expressive material. At the other extreme, Hansen [6] has described a database showing emotion evoked by appearing before a military tribunal. It is hard to doubt that the emotion was genuine. However, it seems very unlikely that its

Figure 1: Time course of judgments on the genuineness of displayed emotion.

expression would be structurally similar to, say, anxiety evoked by seeing that the mood in a section of a revelling crowd is becoming ugly. Structural similarity, not genuineness, determines whether it is suitable for training in a crowd surveillance application.

Genuineness is also contrasted with deception. There is a long tradition of research on the distinction between true and deceitful smiles [1,5]. The issue there is that given a relatively isolated smile, naïve observers are not good at telling whether it is genuine or posed. Children as young as 10 can simulate a smile that naïve observers fail to distinguish from a real one [3]. In this case, what is not clear is why research should take the trouble to obtain genuine smiles if they are similar enough to posed material that humans struggle to tell the difference – unless, of course, the application is detecting lies that humans fail to catch.

Although it is rarely stated, there is a premise that often seems to lie behind discussions of genuineness. It is that there is a core of signs that establish definitively whether a given emotion is present. If those quintessential signs can be detected, surface variability is irrelevant.

There may indeed be quintessential signs, but the obvious interpretation of research on the expression of emotion is that it can be signalled in a virtually limitless

variety of ways. It is also clear that if there are quintessential signs, they are not the basis of everyday human interaction – otherwise simulation and misreading would not play such a prominent part in interaction. Hence for research that aims to simulate everyday human interaction, the hypothesis that there are quintessential signs is of minor interest. For research that aims to better human perception, it is a hypothesis that goes against the weight of evidence, but whose simplicity makes it hard to relinquish completely.

The point of this discussion is not to belittle distinctions between acted and natural, genuine and deceitful, spontaneous and posed. It is to argue that research needs to look beyond them to the structural characteristics that distinguish different types of material, and make them suitable for different uses.

3. Structural issues in the literature

One of the things that different bodies of material do is to suggest very different conceptions of the challenges that face a system attempting to perceive emotion. There is material – typically, but not necessarily posed – that seems amenable to rather simple solutions. Other material – typically, but not necessarily spontaneous – poses much harder problems. The general strategy of what follows is to consider features of data that suggest

research has to look beyond the simple solutions that are often – but not always – invited by data created or selected to conform to idealized conceptions.

In particular, the discussion picks out a recurring theme. One of the key sources of complexity is distribution: instead of being presented in a convenient package, relevant information has to be assembled in non-trivial ways. The issue is a form of the classic problem of segmentation.

Some forms of the distribution problem are already documented in the literature. They are considered in this section. The next section summarises evidence from new studies that highlight other forms of distribution problem.

3.1. Diachronicity

In the archetypal displays of facial emotion, due to Ekman & Friesen [4], evidence of emotion is synchronic: it resides in a combination of local actions that occur at the same time. That has implications for processing: for instance, it means that templates representing holistic patterns can achieve respectable levels of recognition.

It now seems clear that facial expression is at least sometimes diachronic – if there are patterns involving multiple units, they are distributed through time. The point was made by Carroll and Russell some time ago [2], and recent studies by Scherer and Ellgring [8,9] reach similar conclusions. One particular emotion, happiness, does typically produce a pattern of synchronic activation in multiple units; but for others, the average number of action units present even at the apex of a recognizable emotion is about 3 rather than the six or more suggested by Scherer & Ellgring's table of theoretical predictions.

There are two obvious approaches to explaining why the data in the studies by Russell, Scherer and others does not show the familiar patterns of synchronic activation. The first is that the expressions were produced in the context of an activity, rather than for the sole purpose of expressing an internal state. It would not be surprising if expressions whose production was an end in itself were more elaborate than expressions that formed part of a larger action or interaction. The second option rests on the fact that the material is produced by actors. It is possible that the actors effectively underperformed, and displayed only reduced versions of the expressions that would naturally be given in corresponding situations.

A third, intermediate possibility has been proposed by Scherer [8]. He has argued that facial expressions naturally follow a temporal pattern reflecting appraisal checks that are carried out in sequence. Detecting actions in a fixed sequence is harder than detecting them in fixed, simultaneous patterns, but less hard than detecting them in variable, context-dependent sequences.

3.2. Instability

Research by McRorie & Sneddon [7] makes a point which is related to the previous one, but conceptually distinct. They contrasted acted material of the kind used in the Carroll and Scherer studies with naturalistic data from outdoor activities. They presented single frames from the two sources, and asked raters to identify the emotion expressed in each. They used the results to estimate how much expressed emotion changed between adjacent frames. The change was much greater in the naturalistic data. The implication is that in everyday life, perceivers may have to deal with rapid changes not only in the signs that are presented, but in the emotions that they express.

In the context of McRorie & Sneddon's study, it seems likely that the changes are driven by changes in attention; the situation is made up of events and people that carry different emotional charges, meaning that expression changes as attention lights on different aspects.

If so, signs are distributed according to attention; and the perceiver has the task of inferring the whole emotional landscape from the way signs change.

3.3. Multimodality

Most research in the area has considered a single modality at a time. However, one of the problems that is likely to occur with naturalistic data is that information is distributed across modalities. A good example is research by Cohn's group, which has shown that judgments traditionally taken to depend on the face alone in fact depend on other cues. For instance, Ambadar and colleagues showed that although timing does not affect interpretation for extreme expressions, it does for subtle ones [1]. Similarly, they have shown that natural smiles involve not only movements of the face, but also head & shoulder movements. Conversely, smiling without the relevant accompanying gestures is quite likely to mean something other than spontaneous happiness. [10] Later sections consider examples of the kind of situation where that point is practically important.

4. New studies

The material outlined above is the beginning of a case. Much more is needed even to establish that there is an agenda worth pursuing. This section describes new studies designed not to prove points, but to build the case for pursuing the agenda.

4.1. The distribution of concordance and divergence between modalities

The point has been made that evidence is distributed across modalities. That would be tractable if one could

Figure 2: Divergence between facial and vocal pointers to emotion in the Belfast Naturalistic Database. Circle size shows the number of cases with a given combination of features.

assume that the distributed signs pointed to a single state. However, it is not obvious how often different modalities do point in the same direction. This section tries to draw some estimates from a standard source, the Belfast naturalistic database.

The study has two phases. The first aimed for a broad brush overview. Five independent raters judged how concordant or divergent audio and visual information were in each clip where it was judged that there was noticeable emotion. The second looked more closely at clips selected on the basis of the first. Two broad sets were chosen – eight which seemed among the most concordant, eight which seemed among the most discordant. Each set contained two that were judged on the basis of full audio-visual information to lie in each quadrant of the familiar emotion space – two positive active, two positive passive, and so on.

Figure 2 summarises the results of the first phase. Raters used a four-point scale, from zero perceived divergence to complete divergence, and also rated how long the divergence lasted. The point it makes is that moderate degrees of divergence seem to be the norm – which means it cannot be assumed that the indications distributed across different channels will converge.

The second phase gave more detailed attention to clips selected to represent key points within the distribution. Figure 3 summarizes the key results.

The graph reinforces the first phase's conclusion that a degree of divergence is the norm. For clips whose overall tone was judged positive on the basis of full audio-visual information, moderate levels divergence were the norm, and the range was not particularly large.

Negative clips present a different picture. On one side, most of the negative material showed noticeably less divergence than any of the positive material. However, on the other hand, there was extreme divergence in the clips whose overall tone was judged negative active on the basis of full audio-visual information. Impressionistically, this corresponds to the fact that the

Figure 3: Ratings of clips selected to represent key subclasses within the Belfast Naturalistic database

speakers in question were expressing strong negative emotions – in one case about rape, in one case about terrorism – but smiling to offset the impact of the words. That is consistent with everyday experience.

The general point is that the distribution of signs is not uniform. There are cases where it is likely that signs will diverge; and that is part of the challenge facing interpretation.

4.2. The uneven distribution of evidence across time

Intuitively, it seems that evidence for emotion is not evenly distributed. Critical information can be compressed into brief episodes – and it may be specific to particular dimensions of emotion.

Preliminary studies of that issue have been carried out, again using the Belfast Naturalistic Database. The approach adopted is simply to delete sections of a clip that seem likely to be disproportionately informative, and to compare ratings of the resulting edited clip with ratings of the original.

As a preliminary, the database was examined to find cases where it seemed possible that information was disproportionately concentrated. From a large number of candidates, two relatively common types of case were identified. In one, there are relatively brief episodes involving tears, breaks in the voice, covering the face, and related signs of intense distress. In the other, there are relatively large scale body movements that convey active anger or contempt.

Six instances of each were selected. The clips were then edited to produce two variant forms. Original clips contained the full information. In edited clips, the episodes judged to contain concentrated information

Figure 4: Editing removed relatively brief sequences in the original (top panel) where strong signs of emotion seemed to be concentrated; they were replaced with blank sequences in the edited version (bottom panel).

Figure 5: Effects on rated intensity of editing out brief, intense passages (tearful or aggressive), in clips where they make up a lower (left hand side) and higher (right hand side) proportion of the original material. Versions with all the cues are on the left of each panel, versions with selected passages cut on the right.

were blacked out, or (in the case of hand movements) relevant material was deleted.

Viewers rated clips for the overall intensity of perceived emotion and on the standard dimensions of valence, activation and power. For the purpose of analysis, clips of each type (containing signs of distress and containing signs of aggression) were ordered according to the proportion of material that was judged to form brief, intense passages, and divided into two halves – the three with the lowest proportion, and the three with the highest proportion. Analysis of the overall intensity ratings showed that the two halves differed significantly ($F_{1,26}=59.0$, $p<0.001$) and the effect of cutting differed between the halves ($F_{1,26}=4.4$, $p=0.045$). Analysing the two halves separately showed that the cuts had a significant effect in the clips with larger proportions removed ($F_{1,26}=6.8$, $p=0.015$), but not those with lower proportions ($F_{1,26}=0.13$, $p=0.72$). Figure 5 shows the patterns. It can be seen that the clips with a higher density of brief, intense episodes were rated more intense overall; and editing them out reduces rated emotionality in those cases.

Beyond that level are more specific effects. Effects on intensity were particularly large in some individual clips; and cuts affected valence and power ratings for some clips, but not others. These indicate that as one might expect, the nature as well as the intensity of the local episodes matters.

The data are not simple, but they confirm the general intuition that in some cases, information is very unevenly distributed: short sequences have a disproportionate effect both on perceived emotionality overall, and on specific aspects of the impression.

Clearly, for a system attempting to perceive emotion, it is important to know that evidence may not be uniformly spread. Some brief episodes may carry disproportionate informational loads. That, of course, is particularly relevant to cases where there might be social reasons not to show an underlying emotion – as in both cases considered in this section.

5. Conclusion

It is not in doubt that differences between acted and natural material are important. The difficulty is to separate the differences that matter from differences that are of no very general interest. The proposal in this paper is that one way to do that is to attend to differences in the structure of the information that is offered by carefully constructed laboratory examples on one hand, and by complex, fluctuating, multimodal, ambivalent sequences on the other. The evidence does not establish the case conclusively: however, it does indicate that the agenda is worth following up.

ACKNOWLEDGEMENT

The research leading to this paper has received funding from the European Community's Seventh Framework

Programme (FP7/2007-2013) under grant agreement n° 211486 (SEMAINE)

References

- [1] Ambadar, Z. Schooler, and Cohn, J. (2005) Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, 16(5):403–410.
- [2] Carroll, J.M. & Russell, J.A. (1997) Facial Expressions in Hollywood's Portrayal of Emotion. *Journal of Personality and Social Psychology*, 72 (1), 164-176A. Alpher. Frobnication. *Journal of Foo*, 12(1):234–778, 2002.
- [3] Castanho AP, Otta E. (1999) Decoding spontaneous and posed smiles of children who are visually impaired and sighted. *J Vis Impair Blind* 93, 659–662.
- [4] Ekman, P. Friesen WV (1976) *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press
- [5] Ekman, P., & Friesen, W.V. (1982). Felt, false and miserable smiles. *Journal of Nonverbal Behavior*, 6, 238–252.
- [6] Hansen, J., Varadarajan, V., and Ayako, I. (2006) UT-SCOPE – A corpus for Speech under Cognitive/Physical task Stress and Emotion. *Proceedings of Workshop on Emotional Corpora (LREC 2006)*.
- [7] McRorie, M. & Sneddon, I. (2007) Real Emotion Is Dynamic and Interactive. *Proc Affective Computing and Intelligent Interaction 2007*. Berlin: Springer pp. 759-760
- [8] Scherer, K. R. & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion* 7, 113–130
- [9] Scherer, K.R. & Ellgring, H. (2007b) Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion* 7(1), 158–171
- [10] Ambadar, Z., Cohn, J. F., & Reed, L. I. (2009). All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of Nonverbal Behavior*, 33, 17-34.