

Designing A Corpus Of Affective Communication*

Orphée Religieux Computer Sciences, BeNeFri, Master Programme
orphee.religieux@unine.ch

Keywords

Emotion Recognition, Corpus, Design, Theory

ABSTRACT

Computer sciences will soon be centenarian. People are increasingly having direct unsupervised interaction with computer. This can be seen through the amount of new automated services that are openly propose. But these interaction can quickly become wheezy when the put in place, use cases are to rigid. This lack of addaptability, is still to be improved and a way to do this is to render computers able to detect the mood of their users.

To perform this, one need to have the possibility to recognize emotions and therefore, since the end of last century, computer sciences count a new field in its range: Automated emotion recognition. Albeit it is still young, emotion recognition field already has enough stand back to be able to see where improvement are to be made and to determine its own limitation. One of the frequently highlighted need is to have a referential corpus of emotion. that could be used for differents purposes and that would be freely accessible. This article will mainly present important steps that are to be considered to build such a corpus, give example of such existing realization and while putting forward some limitations to the desirable improvement, will try to show where than can be made.

1. INTRODUCTION

The increasing number of articles published per year during the last decade [2] is relevant of a growing interest for automated emotion recognition. This field is found used in many of the computer sciences topics, like the human-machine interaction to allow programs to adapt their execution rou-

*The Paper is a delivrabled of the Msc Research Seminar Emotion Recognition performed in 2011 in the DIVA Group University of Fribourg, supervised by Dr. D. Lalanne, Dr. F. Ringeval.

tine to the mood of the user, for instance by changing the difficulty level of a game when detecting boredom on the user face while playing or in the matter of an automated phone service it could automatically redirect the caller to a human interlocutor when detecting stress in the callers voice. Glancing through emotion recognition litterature, one sees that there are lots of differents emotions corpuses that have already been build. Still a common comment in many articles, related to emotion recognition, is the lack of a free standard multimodal corpus of emotion that could be used as a reference for any purpose. So, groups willing to work in the field of emotion recognition must whether addapt their experiment to one of the many existing corpuses or when having budget, time and resources for it, to build their own. But, as this article will try to enhance, due to the work that has to be done, it is a great investment.

This article will pinpoint important concepts that people willing to build a corpus of emotions will have to face. So to do, it will go through the notions of the dimensionality of emotion. In other words how to modelize an emotion. Section 3 will present used methods to elicit emotions. Furthermore, section 4, will present the modalities on which a corpus can be build on. To continue, section 5 will cover lasting notion like the different type of profile that participant to the experiment should have and precises their role. Finally, will conclude by asking ourself what the perfect corpus should sound like. Briefly, this article will try to be not a HOWTO but more some kind of WHATTO related to emotion corpus building.

2. DIMENSIONALITY OF THE EMOTIONS

How to represent and discriminate emotions is much related to the psychology field state-of-the-art. And as there are few agreements on the right manner to modelize emotion or on the definition of units to qualify emotions. And more generally on the definition of emotin itself at this time. One can almost regard emotions following its own feelings. In the emotion recognition domain, one mainly sees two schools facing each other. On one hand emotions are regarded as categories on the other they are considered as positions in a multidimensional space. The rest of this section will describe those two points of view.

2.1 Categorical

The categoricals models, propose the usage of labels to describe emotional states. Understand that these models consider emotions as being discrete states. People being in one emotional state at a time. Usually considering *full blown*

emotion expression as units. When choosing such an approach, the solutions consist of giving a set of labels to cover all suitable emotional states. The problem here is centered on the cardinality of the set. There are hundreds of words semantically related to emotions. And it seems impossible to work with so many categories. Some people announce that using about 60 labels would be enough to cover the most common encountered emotion, which is a nice improvement but it still is too much. Nowadays, corpuses most of the existing corpuses only consider about ten different labels. That can be a source of trouble when confronting the builded corpus to real-life examples.

2.2 Continuous

As announced before, another reading of emotions, consider them as being part of some continuous vectorial space. Emotional states differing from one another by the strength of the attributes' expression represented as a continuous value weighting at the considered moment. Emotions are then seen as vector of continuous values along the axes of this multidimensional space. The problem here is to find the entries that will represent and dissociate the different emotional states.

And once again, there is no general agreement regarding the cardinality of the emotion space model. However, most models are at least composed of a *valence* axe, expressing the positivity or the negativity of the felt emotion (supposing that enjoyment = -anger), an *activation* axe, also called arousal axe, that could allow to distinguish anger from deception for instance and a *dominance* axe which permits to differentiate anger from fear. Some models reduce this 3D space to a 2D space omitting the dominance, when others increase it up to a 5D or 8D space as in [3] using:

- Valence (attracted-repulsed)
- Personal Strength (powerful-weak)
- Freedom (free-constrained)
- Probability (certain-uncertain)
- Intentional focus (generalize-focused)
- Temporal flow (future directed-current-past directed)
- Temporal Duration (enduring-sudden)
- Social Connection (connected-disconnected)

This second approach becomes more popular nowadays for it seems to be a reasonable mean to solve the problems raised by categorical models, such as the relevancy of the model that only considers full-blown emotion and the *in situ* emotion that could be kind of more complex by being in fact mixed basic emotions.

3. METHODS TO ELICIT AN EMOTION

One of the first idea to have material from which emotion marker could be extracted, was to ask people to simulate some wanted emotions and than to take picture of them. Another method make use a single (or a group of) subject(s) , to asks it/them to repeatedly read some given lines

and to put some emotional content in their acting performances. This method allow the people in charge of the data collection to have a close control on the procedure and on the collected emotion.

Furthermore, instead of acting emotion, one tries to induce emotions. Using picture that are supposed to be strongly emotionally related, to help the subject from which the data are collected to recall emotions. Another method to induce emotion is the well known *wizard-of-oz* (WOZ) experiment. It can be describe as follow: Two people interact with one another but one of them, the wizard, conducts the exchange in such a way that the other will express some emotions. For instance, while thinking to call some helpdesk or some support center, one will face an especially rude wizard that will try to induce hanger in the caller.

Finally, one can try to track down and record emotions from real-life situations. This type of collected data are referenced as *naturalistic*. Meaning that the exposed emotion will neither be acted nor induce from any manner. What one will be able to extract, will be the expressed (and felt) emotion at the moment. There are famous corpuses that were build based on such collected data like the airport lost luggage office one.

Beside the methods to use to elicit emotions, one could ask itself what emotion can be artificially elicited. Everyone has already laughed or felt fear while watching a movie that shows that some emotions are quite easy to elicit but concerning others like shame for instance. How to elicit this kind of emotions? Moreover, before building some experiment to collect emotion related data from subject(s), one should ask itself what kind of emotion can, ethically speaking, allow itself to elicit. for instance how to elicit pain?.

Examples of different emotions corpuses with used elicitation methods specification can be found on the web @: emotion-research.net/wiki/Databases

4. MODALITIES TO RECORD EMOTION AND MATERIAL TO USE

There are about five sorts of modalities that can betray emotions. Considering a subject from which one would like to extract emotion. Recordings of each of those modalities comprise technical constraints that will be exhibit. If ideally, all modalities would be recorded at once, it is not possible due to the material involved in the recording of some of them.

4.1 Speech

The speech is the most natural way to communicate with one another. When accepting the hypothesis that it can be influenced by the emotional state of the observed subject, it can easily be recorded using microphone for instance. Actually, due to this facility, emotion corpuses based on speech features are the most common ones. Often, a differentiation is made between verbal and non-verbal speech. The first level concentrate on the meaning of the talk, in other words it is the semantical content. The second, focuses on the way that words are pronounced, like the ton used or their duration, as well as the inter-words glups or sobs for instance. It is to be noticed that using high quality microphones, while reducing noise can greatly improve the extraction of relevant

features. Another limitation to speech, is that a corpus of emotion only based on the speech modality will be dedicated to a single language.

4.2 Face

The Face might also betray emotions. And in a way it is more universal than speech. Indeed, even if some facial expression might be related to some cultural or geographical surrounding, it has been shown that a large amount of them, like disgust or sadness for instance, are on the contrary, expressed the same way by everyone.

It can be recorded using a camera. Extracting emotional content from this kind of data can be complicated, because it is very rare for someone not to move, especially the head and when willing to analyze the collected data, one has to manage this kind of irrelevant noisy movements.

4.3 Gesture

The term gesture embraces lots of body parts motion, like hands, corporal or head position and movements. One can distinguish *semiotic* gestures that are "use to communicate meaningful information or indication", from *ergotic* that are "usually associated with a particular instrument or job", and *epitemic* "gesture that are concerned with providing tactile feedback that helps to confirm what or where an object is". [2]

As for faces, bodies motions can be recorded with cameras but its automated analysis is still a great challenge for now, Even if the actual state-of-the-art in automated image analysis has significant results notably in the security fields or in medical imagery, there still is a lot of work to do to get rid of the noise and to extract efficiently relevant emotional content. Therefore, gesture isn't often used in the existing corpus.

4.4 physiological

By physiological modalities, one understands as different features as blood pressure, heart rate, corporal temperature, respiration or galvanic skin response. Each of them asks for dedicated recording material and specific knowledge for their analysis. Collecting these features is more intrusive than the one describe above because the participant from which the data are collected might have to be plugged. These recordings material constraint could prevent some elicitation scenarios and may forbid the data collection to be done through a fully naturalistic process. These reasons make their election as modalities to build some corpus less popular.

4.5 Brain

The brain imagery needs the most intrusive and restrictive recording material. For instance *Electroencephalography* (EEG) or *functional magnetic resonance imaging* (fMRI).

5. CATEGORIES OF PARTICIPANTS & DATA TO BALANCE

Choosing a model to represent the emotional states, the modalities on which the corpus is to be build on and the methods to elicit emotions are just part of the job. Once these choices made, the supervisor of the experiment still have to focus on the participants he wants to take part to corpus building process. He also has to concentrate on the

type and the format of the informations he wants to transcribe. And finally organize some human perception test of the builded emotional corpus.

what kind of participant the supervisor of the experiment should select? The answer is that the group of participant must be as representative as possible to the population the corpus is dedicated to. If it does not, one take the risk to see, its collected data, to be biased and unusable for any following experiments. So the group should be compose of subject of both gender, because the values that some modalities could take can be gender dependant. It should also contain people of all ages. Because ton for instance changes with time. More generally, participant should be a sample statistically representative of the targeted population. Beside those consideration on the subjects sample profile.

there are mainly two types of participants intervene in the process of building an emotional corpus: The ones the data are collected from and the ones who transcribe the perceived emotional states.

Before going further it seems important to enhance that until now, notions related to the corpus realisation have been presented as if they were independant one from each others but it is to be noticed that all those choice are strongly interlaced and some choices could forbid others or on the contrary, force others. For instance, choosing to record brain signal through fMRI wont allow the recording of any other modalities at the same time and therefore the collection to be a multimodal one. In the vein, to produce a fully naturalistic corpus, one will have to insure that the records material won't be visible or find a way to make the subject forget its presence.

Now, to continue, Regarding the first class of participants. Using professional actors or any forthcoming subjects embodies the arise cleavage between acted (also called posted) example based collections and databases build upon naturalistic, non-acted examples. Albeit, they have been utterly used and popular because of their convenience. Posted databases are now being criticize for the reason that they only produce simulated emotions. And that some emotion features cannot be faked. Resulting that, when facing *real* emotions examples, not from some laboratory worlds examples the performances of so builded databases rapidly fail down. The alternative could be full naturalistic raw data but logistics and ethical matters come in the deal and often make such enterprise to complicated.

Once or while the data are collected, its emotional content still needs to be transcribed. To perform this operation, one can always use its own intuition, but there are some standard that are proposed. For instance to analyze faces, one can use the *moving picture experts group* (MPEG)-4 format, which gives some units to code and decode [2] facial expressions. Like the *facial definition parameters* (FDPs) which define general shape of facial structure, the *Facial animation parameters* (FAPs), which define how the predefined shape do move, and for gesture analysis the *Body animation parameters* (BAPs).

The line above presents some standard that emphasize, still

is to answer how to perform these annotations of the database. There are some tools like the *FEELTRACE* program that allow its user while listening/watching data to evaluate the emotional state of the subject. Grossly, this tool presents a circle to the user and allow him to "move a cursor to a point on the screen whose co-ordinates reflects the speaker's current state" placing a cursor on a plane give a two dimensional judgment of the emotional states .[2]

Another example of tools, so to be called, developed to facilitate the transcription of emotional content is the *self assessment manikin* (SAM). It presents a same sheet schematic face representing sadness to happiness on one line. On a second line some human shape allowing to judge the arousal of the perceived emotion from quiet to active and on a third line figure to represent social implication from independent to dependant.

6. CONCLUSION

This article tried to enhance difficulties that people willing to build an emotional corpus have to face.

It has presented different notions that where of central interest to build such a corpus. First of all, ways to conceptualize emotion dimensionality and gave two antagonist point of view : the categorical and the continuous one. after that it has also formulated methods to conduct emotion collection through acting or by induction or just observation. Enhancing so doing elements like acted or naturalistic collected data. Continuing with the presentation of the most common modalities that could betray some emotional states. It has also presented the way that participant sample should be chosen to be representative of the global population and how data could be transcribed. Now, what would a perfect emotional corpus look like. It would [2]:

- be fully *naturalistic*. that would solve the problem of the representativeness of the collected data.
- consider all the modalities at once.
- cover the whole domain of emotion and emotion-related states.
- sample the whole range of cultural and individual differences in the expression of emotion.
- Recordings would be of high technical quality.
- comprehensively standardize labelled.
- be statistically tractable.
- be freely available.
- processes for its storing, distribution would be "ethically sound.

Now, this above is a dream, for now, if not for ever. There are lots of limitations that avoid that realization. For instance, collecting naturalistic data presuppose that the targeted subject does not now that he is being recorded and that brings some ethical issues. To continue, recording all modalities at once is impossible for now, because of the necessary material constraints. Added to this, the fact that

material can be hinderring could also impede naturelness of the recordings. Furthermore, covering all the emotional states would need some agreement on it which is not the case at this time. And one could ask itself whether all emotions can be artificially elicited. Even leaving out ethical matter. For the fourth attribute presented, building a representative sampling would mean having a full knowledge of the cultural space which seems kind of difficult just considering all the different languages that exist when willing to have some reasonable sample size. High quality recording seem to be the most affordable quality. Technology regularly brings better standards. Being freely available is just a matter a will from the supervisor but until now as building a efficient corpus is quite an investment, people that have managed to build such corpuses have kept them for their private usage.

All of these qualities that a perfect emotion related corpus should offer aren't affordable for now. But, there still are some of them that could easily be improved.

7. REFERENCES

- B. Schuller, A. Batliner, S. Steidl, D. Seppi, *Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge*, Volume 53, Issues 9-10, November-December 2011, Pages 1062-1087
- E. Douglas-Cowie and WP5 members, *D5c, Preliminary plans for exemplars: Databases, in humaine emotion-research.net, May 2004*
- Tom Cochrane, *8 dimensions for the emotions*, Social Science Information September 2009 vol.48, no. 3 379-420.
- M. Grimm, K Kroschel, E. Mower, S. Narayanan, *Primitives-based evaluation and estimation of emotions in speech*, Vol. 49, Issues 10-11, October-November 2007, pp. 878-800.
- C. Busso and S. Narayanan, *Recording audio-visual emotional databases from actors: a closer look*, in proc. of Languages Ressources and Evaluation (LREC), Marrakech, Morocco, pp. 17-22, May 2008.
- E. Douglas-Cowie, R. Cowie, M. Schröder, *A new emotion database: Consideration, sources and scope* In SpeechEmotion-2000, 39-44.
- F. Ringeval, *Speech anchors and dynamic models of prosody: application to the acted and spontaneous emotions recognition*, thesis work, University Pierre and Marie Curie, pp. 16-22.