Study of Emotion Alternation using Deep Learning and computer vision

Pragyanshu Verma
Department of Electronics and Communication
Indian Institute of Information Technology, Guwahati
Guwahati 781001, Assam, India.

Contents

Abstract

This work presents a novel technique to emotional alternation functioning based on facial expression. Emotion transition primarily denotes the individual differences in emotional processing of different stimuli. Several mechanisms for perceiving and measurement of emotions have been proposed in recent years. However such studies have been based on the assumption that emotions are static and have ignored the dynamic paradigm of emotions. By attending to this altered nature of emotions, an accurate model of emotion transition can be established. To test this hypothesis it was assumed that the overall emotional arousal of a subject depends not only on the present stimuli, but the previous stimuli also has a significant effect. Consequently, a prognostic technique has been designed wherein subjective analysis was performed with 10 participants (8 male and 2 female). The participants were asked to view selected stimuli from the IAPS database. The subjective ratings were collected and the emotion after each stimulus was recorded. The database thus obtained was analyzed and the emotion at the end of the experiment was evaluated for each subject. The results obtained suggest that the emotion felt by participants' alter along with the changing stimuli and depends generally on the recently viewed stimuli. This study can thus represent the dynamics of emotional states and aid in the detection of changing emotions as felt by the subjects.

1. Introduction

1.1. Major contributions of the work

1.1 what is emotion recognition

In general, emotions are a relatively brief conscious experience characterized by intense mental activity and a high degree of pleasure or displeasure. Emotion recognition is the process of identifying human emotion, most typically from facial expression as well as from verbal expressions. This is both something that humans do automatically but computational methodologies have also been developed.[2]

Emotion recognition is very important in todays life. Smart cars can detect facial emotions and alert the driver if he looks sleepy or drowsy. Affectiva uses it to help advertisers and content creators to sell their products more effectively. Brands get customer feedback through emails, chats and social media messages. After analyses such data and deduces whether the underlying emotion is happy, excited, angry, sad, bored, afraid. Emotient was a startup company which utilized AI to predict "attitudes and actions based on facial expressions".[1]

1.2 Types of emotions

There are two way to define emotion — discrete emotions and emotions in dimensional space. Discrete emotion is considered six basic emotions including happy, sad, anger, disgust, surprise and fear which are defined by Ekman; whereas, emotions can be defined based on valance and arousal scores. In literature, two kinds of models are considered and several methods have been proposed. Generally, the emotion recognition is conducted by: first, users are excited by different kinds of stimuli, then the changes of the behaviour, physiological etc. of that subject is tracked; finally, based on those changes various comments on emotions are made. Certainly, such stimuli is very important in emotion recognition system (ERS), like how to select those stimuli, kind of stimuli, way of presentation etc. Next, important issues is what kinds of changes of the subject are taken into consideration. There are several kinds of stimuli applied so far in emotion recognition are audio, video, music, picture etc. On the other hand ERS are like voice, speech, facial expression, physiological signals, gesture etc based system.

1.3 Ways of emotion recognition

Emotion recognition can be done in four ways

1. facial expression based: Facial expression is one of the most cogent, naturally preeminent means for human beings to communicate emotions, to clarify and stress what is said, to signal comprehension, disagreement, and intentions, in brief, to regulate interactions with the environment and other persons in the vicinity.
2. Speech based: From speech we can predict the emotion of person. This speech based recognition mainly focussed on short sentences.The ability to automatically recognize emotions from speech is a highly desirable attribute to have in human machine interfaces and has potential applications in a number of scenarios ranging from call centers to smart devices. Consequently speech based emotion recognition has been an active area of research and has been so over the past decade. The majority of this research has focused on identifying suitable features, feature selection methods and classification techniques.[5]
3. Gesture Based: Body language includes different types of nonverbal indicators such as body posture, gestures and eye movements. These are important markers of the emotional and cognitive inner state of a person. In this work, we review the literature on automatic recognition of body expressions of emotion, a subset of body language that focuses on gestures and posture of the human body.[6]
4. Physiological based: Recently, there have been various attempts to classify the emotions by measuring bodily responses through sympathetic and parasympathetic nerve antagonism and then classifying emotions. To illustrate emotion classification using bodily responses, when a person is excited, his muscles become tense, his palms get sweaty, and both his heart rate and body temperature increase. In the situation, there are three general indicators that can be used for emotion recognition. First, Electromyography (EMG) measures the electrical impulses of muscles during contraction. Second, the skin conductance response (SCR), also known as galvanic skin response (GSR), is the phenomenon that the skin momentarily becomes a better

conductor of electricity when a person is in tension. Third, an increase in the electrocardiogram (ECG) signal indicates a state of stress or frustration. In addition to three indicators, changes in pupil size and reactivity are closely related to emotional changes. In particular, pupils dilate slightly in response to any exciting or interesting stimulus while pupil size tends to decrease in response to unpleasant stimulus.

1.4 Why we used facial expression based

In this project we used facial expression based emotion recognition. Because the implement of this is easy in real time based system. In speech based and gesture based emotion recognition data acquisition is a very big problem but in facial expression based data is easily available. In this project physiological based emotion recognition is not used because cost of set-up was very high.

In literature, most of the emotion recognition works are focused on to discriminate different kinds of basic emotions by employing several kinds of stimuli. However, human emotions cannot consistent for a long time; rather it fluctuates. Such alternation of emotions are very important to study the human emotion with details. Therefore, we intend to study the alternation of human emotion based on facial expression. In this purpose, we have designed and developed a tool to track the changes in facial expression during the subject is watching a certain image.

2. Motivation

In most of the emotion recognition work is done by considering your current facial expression. The cause of these facial expression may be an image or video stimuli. The current emotion recognition softwares tells you emotion of current state. But emotion can be depend on your previous stimuli also. If you see continuously happy, happy images so your final emotion will be happy. But if you are seeing sad, happy, happy images so your final emotion may be happy but not same as previous case (when you see only happy, happy). So via this project we are trying to find on which factors our emotion is dependent.

3. Problem definition

Previously most of the works found on only emotion recognition by showing different set of stimuli. This work aims to study alternation of human emotion when different set of stimuli are being shown. Further, how one emotion influences other set of emotion whenever, they are arising in sequence manner.

Objectives: The objectives of the project are:
a) Design a stimuli presentation tool to study emotion alternation based on facial expression.
b)Define a mathematical model of emotion alternation study.
c) Investigate how different set of emotional states impacts on overall emotion.

4. Workdone

4.1. Experiment

4.1.1 Experimental set-up

For this experiment there is a monitor on table with web camera, keyboard and mouse. Subject should be in comfortable position and ideal distance ‘d’ with camera (as showing in fig-1). For finding the value of ‘d’ there is a test image before any experiment to start. In this test image there are nine circles 4 at corners, 1 in center, in middle of the corner circles. For testing subject have to click on every image in any order at the same time input video from camera is also taken. In this video we checked for camera is detecting subject’s face or not. If if camera not detects with the 90% accuracy then the test experiment repeat until it got above 90% accuracy. If subject is not clicking inside the circles then also text experiment will restart.
Image:

4.1.2 Stimuli Design

Experiments related to emotion, the design of efficient and reliable stimuli is very important. In literature, several kind of emotional stimuli have been used so far including image, music, mental imagery, and films. In our experiment the images as stimuli has been taken into account. The images (from IAPS) which contains two types of emotion happy and sad. which give the subjects to feel real life scenario. The images has been chosen such a way that it can elicit strong and physiological changes.[8]

During showing stimuli we also collect the information of mouse. Person should click on every image once. We store information like how fast he is clicking during experiment and where he is clicking.

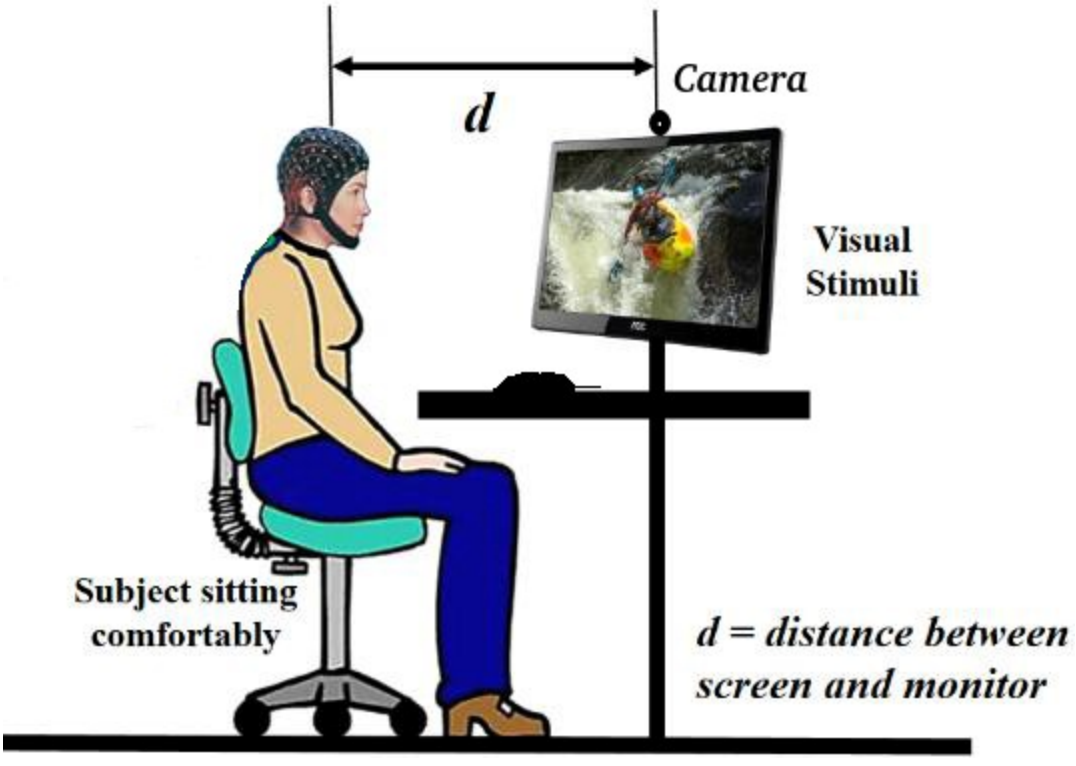

Figure-1: Experimental set-up

4.1.3 Participants

We intended to choose healthy subjects who are under graduate students from the department of Computer Science Engineering (CSE) and Electronics and Communication Engineering (ECE) from our institute, IIIT Guwahati. So, there will be male and female students with the age of around 18 to 21 years. Before the participation of the experimental process, their consents will be taken in written form along with some basic medical history. Such, report will not be disclosed publicly in any means and the PI will maintain the secrecy. We have done one session in which there are 2 trials.

4.1.4 Tasks

In our experiment, There are three type of tasks first for the finding basic conditional probabilities that we have used in in our mathematical formula, second task for the checking out our formula's result is correct or not, in third task we just shown images to the subject and predict the emotion alternation using that formula.
Here we define some term:

Pr(H|HH): probability of happy if they have seen images happy, happy

Pr(H|HS): probability of happy if they have seen images happy, sad

Pr(S|SS): probability of sad if they have seen images sad, sad

Pr(S|SH): probability of sad if they have seen images sad, happy

Pr(U|HS): this represents the error

(U: uncertain/error)

First Task: first task is for predicting different conditional probabilities as we mentioned above (Pr(H|HH),Pr(S|HH),Pr(H|HS),Pr(S|SH) etc). In this task we showed three images where first two images may be in HH or HS or SH or SS sequence and in third image we took final emotion of person by clicking on different levels of happiness and sadness. The time duration to seeing the image is depend on subject. For the first task we took data from 10 person(8 male, 2 female) for 2 trials. The following data we stored from task1.
(a) Valence value of images: This value have taken from the IAPS table.
(b) Elapsed time on images: During showing stimuli we also recording the time parameter. At which time he started or ended, how much time subject gives for a single image (elapsed time of image).

(c) Output: This value subject gives by clicking on third image. In which there are six options to click eg. These are the level of happiness and sadness. Subject have to click any one of them.

Second task: We did second task to verify the result of our mathematical formula. Result has been calculated using the data of the first task. The second task was to show the six random happy and sad images and after every image we have taken the output emotion from the subject in six levels eg H1, H2, H3, S1, S2, S3. Here we stored the same values as first task eg valence value of image, elapsed time on image and output emotion (given by subject). For second task we took data from 6 person(4 male, 2 female) in 2 trials.

Third task: In third task we just shown five to ten images to the subject and not any output from subject. Here we stored valence value of image, Elapsed time on image, average emotion.

Average Emotion score: Assume no. of frame per sec = N_f

if image seen by user from t_1 time to t_2 Let $|t_2 - t_1| = t_n$

total no of frame = $t_n \times N_f = N$

avg emotional score= $\Sigma_{i=1 \text{ to } N}\ j$ { j= +1 : if emotion is positive

j= -1 : if emotion is negative

j= 0 : if frame not detected

4.2. Emotion recognition

For this complete project we have 3 main task first, to detect the facial expressions of a person, second, showing random images to the person then record the emotion, mouse clicking, elapsed time on image, third, predict the final emotion of person after showing all images. For the first task, detect the facial expression of person, we built a CNN architecture. This architecture used Global Average Pooling to completely remove any fully connected layers. This was achieved by having in the last convolutional layer the same number of feature maps as number of neural class, and applying a softmax activation function to each reduced feature map. This architecture is a standard fully-connected neural network composed of 9 convolutional layers, ReLUs, Batch Normalization and Global Average Pooling (fig-4) [4].This architecture is trained with ADAM optimizer. This model contains approximately 600,000 parameters. It was trained on the FER-2013 dataset. This dataset contains 35,887 grayscale images.[3]

where each image belongs to one of the following classes “angry”, “disgust”, “fear”,“happy”,“sad”, “surprise” and “neutral”. But in our project we only take two

Fig-2: Samples from fer-2013 datase

types of images (positive and negative) to decreasing complexity. In positive images we include “happy”, “surprise” and “neutral” and in negative images we include “angry”, “disgust”, “fear” and “sad” type of images.

The emotion recognition process is showing in fig-3. First takes input as a video of subject's face from camera. Now this video we treat as a frame by frame. Now this frame/image goes to the next step of preprocessing. After preprocessing image works as a input in our CNN classifier (CNN description is given in fig-4). Now classifier classifies the emotion in that perticular frame is positive or negative.

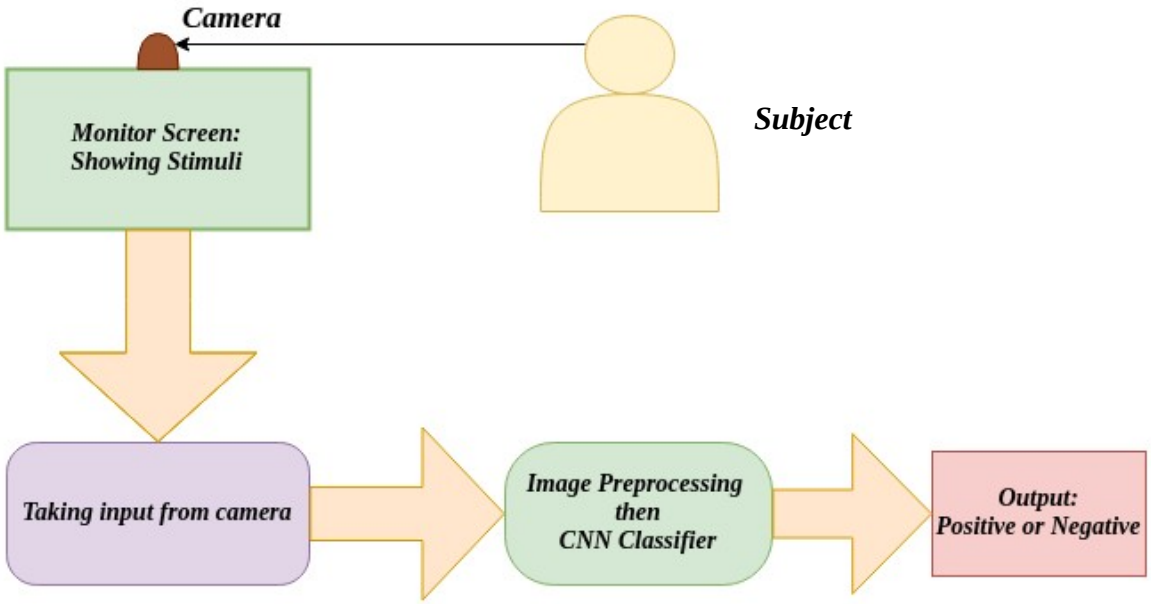

Fig-3: Process of Emotion Recognition

4.3. Primary emotion alternation scores

For predicting the final emotion we have to first predict what is the final emotion of a person if we are showing the images in the HH{happy-H,sad-S}, HS, SH and SS manner. For this we implement a set-up of camera, desktop and mouse. We took 10 people for our experiment and done 2 sessions to finding out the final emotion after showing the images in HH manner, final emotion after showing the images in HS manner, final emotion after showing the images in SH manner and final emotion after showing the images in SS manner.

These probability we have found from our experiment. In experiment we have shown images in different groups like HHB (H-happy, B-blank), HSB, SSB, SSB and SHB. Here in blank images we have taken the output of emotion after showing first two images. The person can give six type of outputs like H1, H2, H3, S1, S2 and S3. Here H represents happy and 1 represents level of happiness. So after showing any type of two images subject can give their emotion on clicking six levels of emotions. The results of this experiments is describe in the table (Fig-5) [9,10]. By seeing these values it is clear that the effect of just previous image is higher than other previous images.

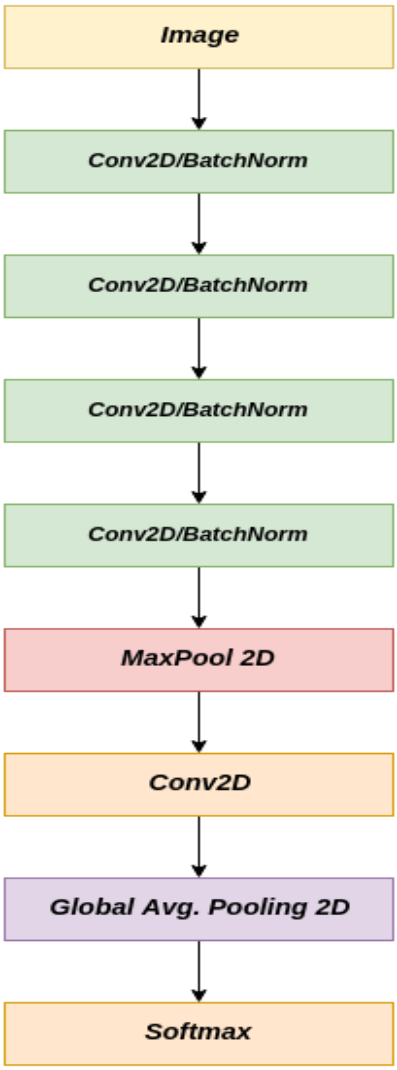

Fig-4: CNN model

Condition	Conditional Probability
Pr(H|HS)	0.7
Pr(S|HS)	0.2
Pr(U|HS)	0.1
Pr(H|SH)	0.9
Pr(S|SH)	0
Pr(U|SH)	0.1
Pr(H|HH)	1
Pr(S|HH)	0
Pr(U|HH)	0
Pr(H|SS)	0.2
Pr(S|SS)	0.7
Pr(U|HS)	0.1

Fig-5/ Table-1: Probability table

The images that were used in this experiment are from IAPS (International Affective Picture System). The IAPS is being developed to provide ratings of affect for a large set of emotionally-evocative, internationally-accessible, color photographs that includes contents across a wide range of semantic categories. The IAPS (pronounced EYE-APS), along with the International Affective Digitized Sound system (IADS), the Affective Norms for English Words (ANEW), as well as other collections of affective stimuli, are being developed and distributed by the NIMH Center for Emotion and Attention (CSEA) at the University of Florida in order to provide standardized materials that are available to researchers in the study of emotion and attention.

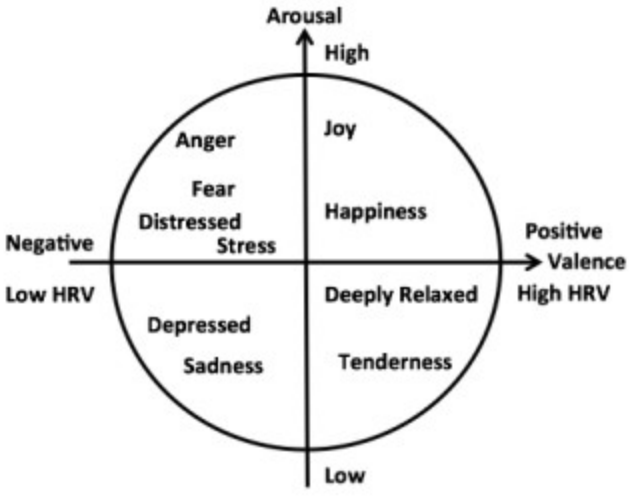

Figure 6 : Emotions classified by arousal/valence levels

We began by relying on a relatively simple dimensional view, which assumes emotion can be defined by a coincidence of values on a number of different strategic dimensions. This view is founded in Osgood's (Osgood, Suci, & Tanenbaum, 1957) seminal work with the semantic differential, in which factor analyses conducted on a wide variety of verbal judgments indicated that the variance in emotional assessments were accounted for by three major dimensions: The two primary dimensions were one of affective valence (ranging from pleasant to unpleasant) and one of arousal (ranging from calm to excited). A third, less strongly-related dimension was variously called 'dominance' or 'control' . Dimensional views of emotion have been advocated by a large number of theorists through the years, including Wundt (1898), Mehrabian and Russell (1974) and Tellegen (1985) To assess the three dimensions of pleasure, arousal, and dominance, the Self-Assessment Manikin (SAM), an affective rating system devised by Lang (1980) was used. In this system, a graphic figure depicting values along each of the 3 dimensions on a continuously varying scale is used to indicate emotional reactions. For the arousal dimension, SAM ranges from an excited, wide-eyed figure to a relaxed, sleepy figure. For the dominance dimension, SAM ranges from a large figure (in control) to a small figure (dominated). It is a 9-point rating scale for each dimension. Ratings are scored such that 9 represents a high rating on each dimension (i.e.,high pleasure, high arousal, high dominance), and 1 represents a low rating on each dimension (i.e., low pleasure, low arousal, low dominance). So here for simplicity we only take “valence” value for calculations. For a happy/positive image value of valence is high in compare to sad/negative image.

4.4. Results

Now we applied above primary conditional probabilities in mathematical formula to get the emotion result.[7] Here H = happy = positive S = sad = negative

current emotion state = S_i {S_i S_{i-1} can be H or S}
previous emotion state = S_{i-1}
Valence value of current image = V_i
i = number of image showed to subject

$$\mathbf{S_i = V_i + S_{i-1} * Pr(A|AC)} \quad \{A, B, C \text{ can be H or S}\}$$

A= previous emotion state = S_{i-1}
C= depends upon current valence value, if valence value is belongs to Happy image then C = H
if valence value is belongs to Sad image then C = S

For the first case S_{i-1} * Pr(A|AC) = { 1 : when showed image is Positive (happy)
-1 : when showed image is Negative (sad)}

These conditional probabilities we have already foud out in table-1. So if we showed N images to the subject so value S_1 is the emotion state of subject after seen only first image, S_2 is the emotion state of subject after seen first and second image. Like this S_N is the final emotion state after seen all images by subject.

4.5. Conclusion and future work

In this project we focused on how emotion is changing if subject seen random positive and negative images. By finding the primary conditional probabilities it is clear that the effect of just previous image on current image is higher than other previous images. We applied above (dicussed in result section) mathematical formula on 6 subject but we have not got the wanted result. In our future work we will increase the number of subject for taining purpose. Here we have done experiment in 1 session and 2 trials so we have to increase session and trials also. We will try to implement more modified formulas to increase the test accuracy.

List of Publication (in IEEE format)

[1]Cowie, Roddy, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G. Taylor. "Emotion recognition in human-computer interaction." *IEEE Signal processing magazine* 18, no. 1 (2001): 32-80.

[2] Loveland, Katherine A., B. E. L. G. I. N. Tunali–Kotoski, Y. Richard Chen, Juliana Ortegon, Deborah A. Pearson, Kristin A. Brelsford, and M. Cullen Gibbs. "Emotion recognition in autism: Verbal and nonverbal information." *Development and psychopathology* 9, no. 3 (1997): 579-593.

[3] Arriaga, Octavio, Matias Valdenegro-Toro, and Paul Plöger. "Real-time convolutional neural networks for emotion and gender classification." *arXiv preprint arXiv:1710.07557* (2017).

[4] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *arXiv preprint arXiv:1502.03167* (2015).

[5] Sethu, Vidhyasaharan, Eliathamby Ambikairajah, and Julien Epps. "Speaker normalisation for speech-based emotion detection." In *2007 15th international conference on digital signal processing*, pp. 611-614. IEEE, 2007

[6] Kessous, Loic, Ginevra Castellano, and George Caridakis. "Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis." *Journal on Multimodal User Interfaces* 3, no. 1-2 (2010): 33-48.

[7] Yamato, Junji, Jun Ohya, and Kenichiro Ishii. "Recognizing human action in time-sequential images using hidden markov model." In *Proceedings 1992 IEEE Computer Society conference on computer vision and pattern recognition*, pp. 379-385. IEEE, 1992.

[8] Constantinescu, Alexandra C., Maria Wolters, Adam Moore, and Sarah E. MacPherson. "A cluster-based approach to selecting representative stimuli from the International Affective Picture System (IAPS) database." *Behavior research methods*49, no. 3 (2017): 896-912.

[9] Evans, Jonathan St BT, Simon J. Handley, and David E. Over. "Conditionals and conditional probability." *Journal of Experimental Psychology: Learning, Memory, and Cognition*29, no. 2 (2003): 321.

[10] Hájek, Alan. "What conditional probability could not be." *Synthese* 137, no. 3 (2003): 273-323