Producing the Smiley encoding

Using a slide show presentation program the images would be presented one by one to a subject. The subject then is encouraged to classify the picture as "smiley" or "non-smiley" as fast as possible, so as to act by instinct, by pressing designed keys in the keyboard. For every answer the program register the time it took to classify the picture, the filename of the picture, and the class it was classified as (*smiley* or *non-smiley*).

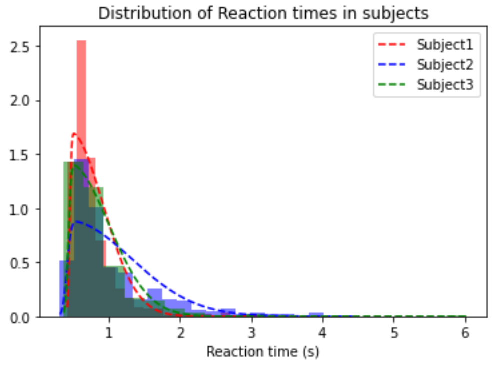
A single person was found to be prone to miss-classifying some faces (reportedly so) by accidentally pressing the erroneous key, thus the slide show was shown to 3 different subject so as to then use their collaborative classifications to filter out accidental miss-classifications.

The classification is binary, so in order to create a continuum of the smile-strength label the assumption is that a face would be strongly *non-smiley* or strongly *smiley* if the reaction time of the subject when classifying it was lower than its average reaction time, and a face would be weakly *non-smiley* or *smiley* if the reaction time was higher (more unsure about class). Upon looking at the distribution of reaction times two thresholds were chosen manually so as to throw away all reaction times above 2s arguing subject wanders around and looses focus, and below 0.1 s arguing accidental clicks. Then a final cleanup is done by removing all faces from the data-set where the 3 subjects do not agree on the class, since it would harm the accuracy of the assigned label to the face and so as to filter out accidental miss-classifications.

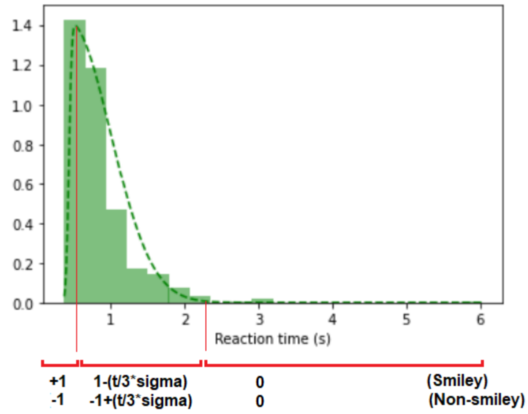
With the remaining faces, since the distribution of reaction times for the subjects had the shape of an asymmetric Gaussian a skewed Gaussian was fitted for every subject and using the parameters of the fit the *smile-strength* label was produced individually for each subject using the function:

$$Score\ Subj_i(t_j^{(i)}) = \begin{cases} 1 & if\ t_j^{(i)} \leq \mu^{(i)} \\ 0 & if\ t_j^{(i)} \geq \mu^{(i)} + 3 \cdot \sigma^{(i)} \\ 1 - t_j^{(i)}/3\sigma^{(i)} & \text{otherwise} \end{cases}$$

In figure 3 the distribution of reaction times for every subject is shown, together with an example annotated with the mapping function. Finally the encodings that each subject produce for a given face are combined by taking the mean to produce the final encoding of the *smile-degree*.



(a) Distribution of reaction times for the 3 subjects.



(b) Mapping function from reaction times to smile strength score.

Figure 3: Distributions of reaction times for the different subjects and example of the mapping function from reaction times to *smile-degree* encoding.

2.3 UTK-Face

Finally the UTK-Face data-set[2] is used with the aligned and cropped faces. The data-set has already been introduced, the difference now is that the data-set is no longer curated to filter out people with glasses, watermarks or other artifacts, making the data-set more challenging. The original UTK-Face data-set contains in the filename of the pictures information labelling the age and gender of the face, two labels for each picture. The age label is a integer value between 0-114 and the gender label is an integer 0:male, 1:female. Then as is present in the data-set the age can be thought as a continuous label and the gender a discrete binary label.

A sample of 32 pictures from this data-set is now shown in 45, notice now that people appear in a greater variety of postures, posing a greater challenge for the algorithms.

[2]https://susanqq.github.io/UTKFace/