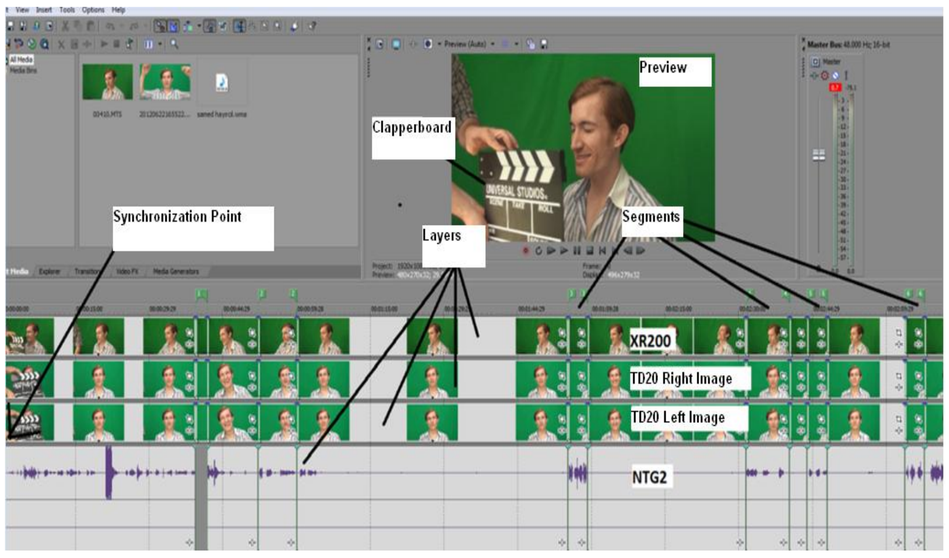
A REACTED AUDIO-VISUAL EMOTION AND MENTAL STATE DATABASE

Onur Önder[1], Çiğdem Eroğlu Erdem[1], Metehan Irak[2]
1Department of Electrical and Electronics Engineering
2 Departmentof Pyschology
Bahçeşehir University, Besiktas, İstanbul, Turkey
{onur.onder, cigdem.eroglu, metehan.irak}@bahcesehir.edu.tr

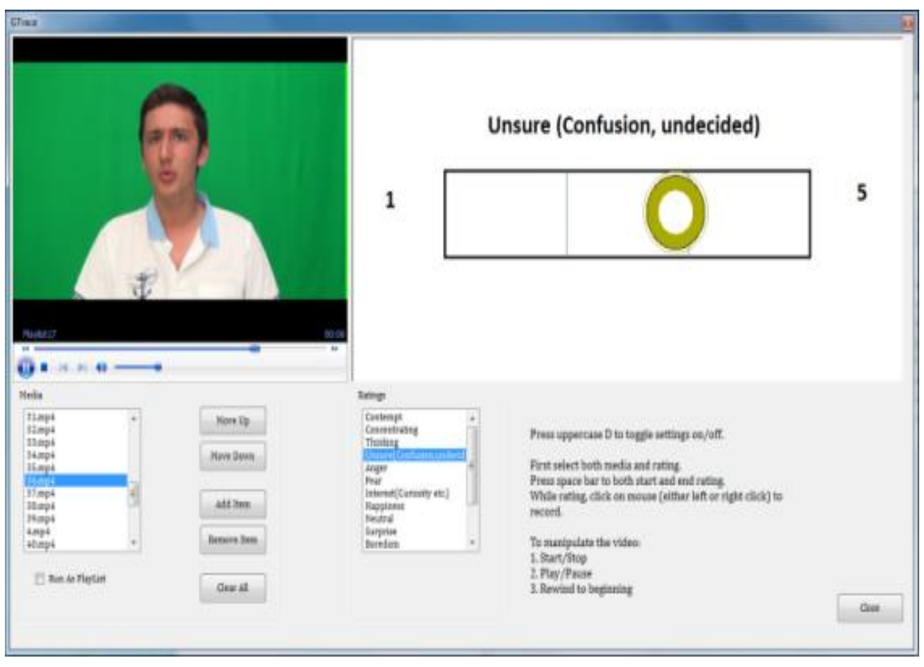
Goal: In this work, we aim to record a re-acted audio-visual emotion and mental state database in Turkish and in English.

- **FEATURES OF THE DATABASE:**
- The database contains synchronous recordings of a subject using a frontal stereo camera and a half profile mono camara, both of which are high definition.
- The subjects watch visual stimuli on a screen infront of them, which are designed and timed to elicit certain emotions and mental states.
- The subjects describe answer questions about the visual stimuli in an unscripted way.
- The target emotions are:
  - Six basic ones (*happiness, anger, sadness, disgust, fear, surprise*) and additionally *boredom*.
  - Complex mental states (contempt, concentrating, *unsure (including confused, undecided), thinking, concentrating, interested (including curious), and bothered*).
- The database also contains short acted recordings in Turkish.

Full HD recorded front view (stereo), 45 degree side view and audio example of a subject, who is talking in anger.

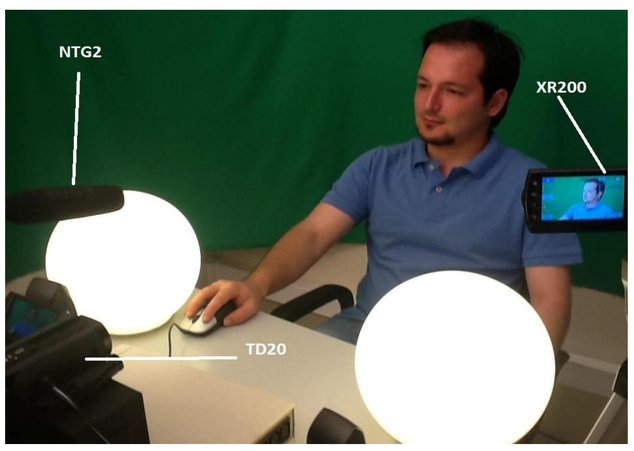


In postprocessing recorded video and audio are synchronized and segmented to small clips using Sony Vegas software and then rendered



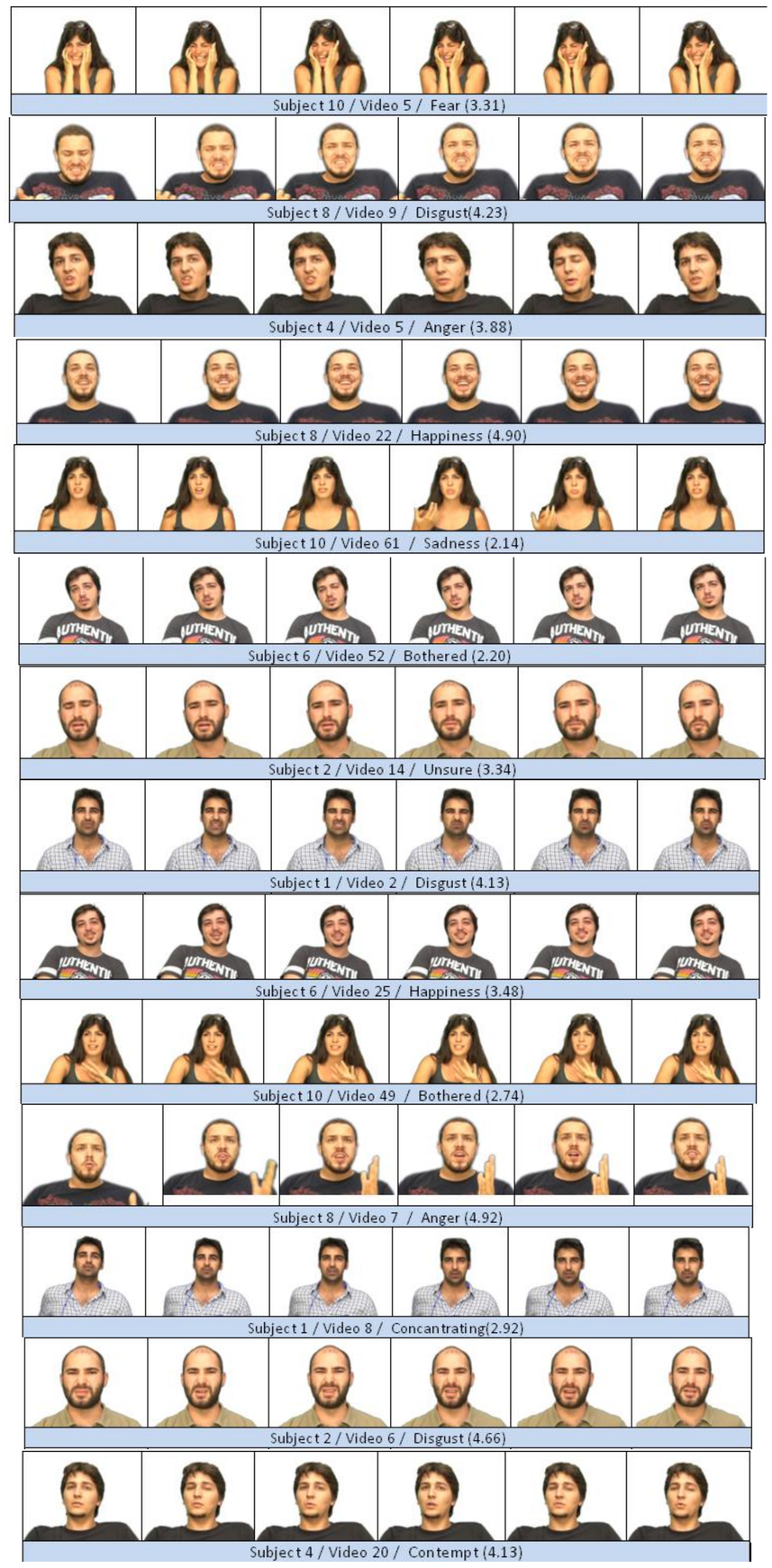
The GTrace tool, is used for annotation. An evaluator watchs the clip, selects the emotionor mental state the clip reflects and then also gives it a rate between 0 and 5.

LABELS
Neutral
Happiness
Sadness
Anger
Surprise
Disgust
Fear
Bored
Contempt
Concentrating
Thinking
Unsure (confused etc.)
Interested(Curious etc.)
Bothered



Recording Environment:

-- A Sony HDR-TD20 Stereo HD camera is used fro frontal view and a Sony HDR-XR200 Mono camera is used for half profile view with an angle of 45 degrees
-- Illumination is provided by using three 1000 Watt tunsten (Red Head) lights.
-- For recording the audio, a Rode NTG 2 shotgun (directional) microphone has been used.
-- A clapboard is used to assist in the synchronization of audio and video streams in a post-processing stage.



Conclusion and Future Work

We present our work towards a re-acted audio-visual emotion and mental state database in Turkish and in English. Right now, our database is consist of **24 subjects** and approximately **960 clips.**The number of subjects is increasing at a rate of 2 subjects per day. Soon we will make the database available to reseachers via a web site. We plan to eliminate outliers in the data by comparing the labels given by 5-7 annotators. We will also provide some baseline emotion and mental state recognition results using state-of-the-art classification methods.

This work has been supported by the Turkish Scientific and Technical research Council under Project 110E056