The Sabancı University Dynamic Face (SU DFace) Database
Mahnoor Nadeem¹, Yaşmur Daşla Şentürk¹, Nihan Alp¹
¹Sabancı University, Istanbul, Turkey

Introduction
• Faces are salient visual stimuli that convey a wide range of social, biological, and emotional information. Due to this, faces have become a source of interest across various research disciplines, which often use stimuli from face databases. However, although faces are highly dynamic (Munhall et al., 2004), most face databases consist of either static stimuli (Ng & Winkler, 2014), or stimuli in which the dynamism is derived from emotional expressions (Trautmann et al., 2009).

To enable a comprehensive understanding of the nuanced behavioral and neural patterns of dynamic face perception, databases in which the influence of emotional expressions is excluded are needed.

The Sabancı University Dynamic Face database (SU DFace databases) provides a stimulus set in which extraction of dynamicity without the influence of emotional processes is possible, as it contains stimuli with only neutral facial expression and natural speech articulation.

Name of the database Type of the speech Duration of stimuli
RAVIDESS Lexically matched speech and song Each stimulus is 1 sentence long duration
MOBO Short response questions, free speech questions, and to read a pre-defined text 61 hours of audio-video data in total
IEMOCAP Spoken scenarios 12 hours of audio-video data in total
DaFEx Spoken scenarios and playing out emotions without using sentences 4-27 seconds

Method
• 170 one-minute-long audio-visual clips stored in 1920 X 1080 pixels videos at 60 Hz frame per second.
• Each model (58 students; 18 males, 40 females; M_age = 21.63) recorded three videos vocalizing two structured texts, Turkish National Anthem (Ersoy, 1986) and Ataturk’s Speech to Youth (Atatürk, 1927), with repetition in cycles and one spontaneous speech. All three speeches were articulated in Turkish with a neutral expression.

Stimuli
• Videos were recorded with Canon 6D Mark II camera, placed above the Manfrotto 290 Xtra tripod 1.55 m away from the model.
• Illumination, background, facial occlusions, and facial angles were controlled across all videos.
• Videos were recorded in a 45-minute recording session by two researchers: One recorded video, one observed controls through the EOS utility application.

Recording set up & procedure
• Videos were recorded with Canon 6D Mark II camera, placed above the Manfrotto 290 Xtra tripod 1.55 m away from the model.
• Illumination, background, facial occlusions, and facial angles were controlled across all videos.
• Videos were recorded in a 45-minute recording session by two researchers: One recorded video, one observed controls through the EOS utility application.

Development
• Pixel-wise differences of various facial features were computed as described in the Chicago face Database (Ma et al., 2015).
• Inter-rater reliability of objective measurements of overall facial features was high (.91).

Objective measurements of facial features
• Overall, the database was perceived to be above average in neutralness (M=4.7) and naturalness (M=4.5). The most common perceived mental state for the overall database was “concentrated” (29.5%).
• A significant main effect was only found for ‘question type’ (naturalness/neutralness).

Analysis of subjective measures

CONCLUSIONS
• The SU DFace Database provides a large sample of stimuli with controlled and consistent parametric variations and different speech articulations with long video duration.
• The database has been reliably tested and verified through analysis of stimulus properties, subjective evaluations, and objective measurements.
• One limitation of this database might be the unequal number of male and female models. We also did not check the effect of participants’ perceptions of the model’s attractiveness on speech comprehension.

REFERENCES

Method
• 4-seconds 170 audio-visual clips were presented to the participants on mute in two different blocks.
• Participants (176 Sabancı University students (58 males, 117 females; 1 undeclared) were asked to rate the naturalness, neutralness (1-7 Likert Scales), and indicate the mental state (confused, concentrated, bored, thinking etc.) of the model.
• Participants were also asked to indicate their overall confidence level, and a 50% confidence level was set as an inclusion criterion for further analysis.

Analysis of articulation
• The mean articulation duration of the National Anthem and Speech to Youth are significantly different between cycles. However, the minimum and maximum articulation duration range within the Turkish National Anthem (16-34 seconds) and Speech to Youth videos (15-34 secs) are very similar. This difference in articulation duration in the cycles is due to differences in the model’s vocalization speed.

Analysis of objective measurements

Visuo-spatial Measurements

Objective Measurements

Analysis of subjective measures

Participants were also asked to indicate their overall confidence level, and a 50% confidence level was set as an inclusion criterion for further analysis.