Behavior Research Methods (2023) 55:3078–3099
https://doi.org/10.3758/s13428-022-01951-z

The Sabancı University Dynamic Face Database (SUDFace): Development and validation of an audiovisual stimulus set of recited and free speeches with neutral facial expressions

Yağmur Damla Şentürk[1] · Ebru Ecem Tavacioglu[1] · İlker Duymaz[1] · Bilge Sayim[2,3] · Nihan Alp[1]

Accepted: 6 August 2022 / Published online: 26 August 2022

Abstract

Faces convey a wide range of information, including one's identity, and emotional and mental states. Face perception is a major research topic in many research fields, such as cognitive science, social psychology, and neuroscience. Frequently, stimuli are selected from a range of available face databases. However, even though faces are highly dynamic, most databases consist of static face stimuli. Here, we introduce the Sabancı University Dynamic Face (SUDFace) database. The SUDFace database consists of 150 high-resolution audiovisual videos acquired in a controlled lab environment and stored with a resolution of 1920 × 1080 pixels at a frame rate of 60 Hz. The multimodal database consists of three videos of each human model in frontal view in three different conditions: vocalizing two scripted texts (conditions 1 and 2) and one Free Speech (condition 3). The main focus of the SUDFace database is to provide a large set of dynamic faces with neutral facial expressions and natural speech articulation. Variables such as face orientation, illumination, and accessories (piercings, earrings, facial hair, etc.) were kept constant across all stimuli. We provide detailed stimulus information, including facial features (pixel-wise calculations of face length, eye width, etc.) and speeches (e.g., duration of speech and repetitions). In two validation experiments, a total number of 227 participants rated each video on several psychological dimensions (e.g., neutralness and naturalness of expressions, valence, and the perceived mental states of the models) using Likert scales. The database is freely accessible for research purposes.

Keywords Face database · Dynamic face · Neutral face · Natural face · Face recognition · Speech recognition

Faces are highly important visual stimuli that provide a broad range of signals. Facial information enables us to recognize the identity, sex, emotion, and other mental states of people we are interacting with (Palermo & Rhodes, 2007), often in an automatic and rapid fashion (Batty & Taylor, 2003; Öhman, 1997). It facilitates the formation and maintenance of social relationships and enables successful communication in social environments. Hence, facial expressions, including neutral expressions, have emotional and social significance (Carrera-Levillain & Fernandez-Dols, 1994). Not least due to this prominent role of faces and facial expressions in human interactions, they have become a major research topic in various disciplines, such as psychology, cognitive neuroscience, and computer science. In studies in developmental psychology, for instance, it was found that preferential looking at faces starts early in infancy (Mondloch et al., 1999; Otsuka, 2014). For example, the general preference—already present in newborns—to look at faces and face-like stimuli (Frank et al., 2009) has been shown to depend on face orientation, with a clear preference for the upright orientation (Mondloch et al., 1999). Furthermore, many neuroimaging studies have revealed distinct brain regions that respond to faces, such as the fusiform face area (Kanwisher et al., 1997), lateral fusiform gyrus, and superior temporal sulcus (Haxby et al., 2000), which have been shown to develop with age (Golarai et al., 2007). In these studies, the face stimuli are usually obtained from the internet or freely available face databases. If available

✉ Nihan Alp
nihan.alp@sabanciuniv.edu

[1] Psychology, Sabancı University, Orta Mahalle, Tuzla, İstanbul, 34956, Turkey

[2] SCALab - Sciences Cognitives et Sciences Affectives, Université de Lille, CNRS, Lille, France

[3] Institute of Psychology, University of Bern, Fabrikstrasse 8, 3012 Bern, Switzerland

sources are not satisfactory, researchers generate their own stimulus sets based on their needs. To address the needs of the research community on face perception, a wide variety of face databases should be available. Indeed, in the last few decades, a considerable number of face databases with different features (e.g., stimulus type; Livingstone & Russo, 2018, Ma et al., 2015, demographics; Tottenham et al., 2009; stimulus number, frame rate; Jobanputra et al. 2018; Yin et al., 2008), in different settings (e.g., natural; Grgic et al., 2011, or lab environment; Gur et al., 2002), and with technical characteristics (e.g., face angle variations; Moreno & Sanchez, 2004; for a review see Krumhuber et al., 2017) have been published.

Existing databases

An extensive number of face databases, with widely varying stimuli, exist in the literature. One of the main divisions of existing face databases is that the faces are either static (pictures) or dynamic (videos). The stimulus sets are generated by different methods, including taking images from online platforms (e.g., the VIP Attribute dataset: Dantcheva et al., 2018; Labeled Faces in the Wild: Huang et al., 2008; FaceScrub dataset: Ng & Winkler, 2014), capturing images in a laboratory (e.g., PUT face database: Kasinski et al., 2008; FERET Database: Phillips et al., 2000), or capturing them in uncontrolled settings (e.g., FIA [Face In Action] database: Goh et al., 2005; SCface: Grgic et al., 2011; for an extensive list of face databases see https://www.face-rec.org/databases/). These databases include a great number of variations in terms of stimulus properties, such as facial expressions (e.g., emotions or complex mental states), posing (e.g., frontal view or different angles of the face), illumination, image quality, model characteristics (e.g., age, ethnicity, language, human or computer-generated faces) and other factors.

Most of the published databases include a variety of facial expressions. The most common facial expressions portrayed in such databases are the six basic emotions: happiness, sadness, surprise, fear, anger, disgust (Cao et al., 2014; Ekman et al., 1987; Ekman & Friesen, 1971; Livingstone & Russo, 2018; Ma et al., 2015; Martin et al., 2006; Vaiman et al., 2017). Databases including complex mental states (e.g., shame: Beaupré et al., 2000; Kaulard et al., 2012; contemptuousness: Langner et al., 2010; playfulness: Schmidtmann et al., 2020; calmness: Tottenham et al., 2009) have also been developed to complement previous databases that represented only a small part of the entire spectrum of emotions.

Interestingly, if neutral facial expressions are included, they represent—despite their ubiquity in daily life and human interactions—only a small subset of the faces in (both static and dynamic) databases (Chen & Yen, 2007; Ebner et al., 2010; Garrido et al., 2017; Ma et al., 2015; O'Reilly et al., 2016; Tottenham et al., 2009).

Studies using dynamic faces

The majority of previous studies examined face perception using static stimuli. However, faces are highly dynamic visual stimuli, conveying important information through their dynamics. Therefore, the ecological validity of static stimuli has been questioned (Ferreira-Santos, 2015; Wehrle et al., 2000). One of the main criticisms is that meaningful information is conveyed by the face dynamics (Jack & Schyns, 2015). For instance, extracting cues about the emotional state (Bassili, 1978) and the intentions of a person (Nummenmaa & Calder, 2009), as well as understanding speech (Munhall et al., 2004), is often based on the dynamic information of the face. Many studies have shown that—compared to static stimuli—dynamic stimuli lead to more accurate expression recognition (Calvo et al., 2016; Cunningham & Wallraven, 2009; Trautmann et al., 2009; Wallraven et al., 2008) and improve speech comprehension (Rosenblum et al., 1996, 2002) as well as sex (Hill & Johnston, 2001) and identity recognition (Christie & Bruce, 1998). Previous studies have also suggested that dynamic information enhances the recognition of ambiguous facial expressions (Cunningham & Wallraven, 2009), the differentiation between posed and spontaneous expressions (Krumhuber et al., 2017), and how realistic an expression appears (Biele & Grabowska, 2006). Dynamic information seems to provide a strong set of visual cues, which enhance visual processing (Grainger et al., 2017). Moreover, neuroimaging studies revealed that brain activations differ across static and dynamic stimuli. For instance, Pitcher et al. (2011) showed that the right posterior superior temporal sulcus was more active during the perception of dynamic faces compared to static ones.

Because dynamic faces have many advantages compared to static faces, there has been an increased interest in generating dynamic face databases (Battocchi et al., 2005; Busso et al., 2008; Livingstone & Russo, 2018; McCool et al., 2012; Navas et al., 2004; O'Reilly et al., 2016; Pigeon & Vandendorpe, 1997; Zhalehpour et al., 2017). To obtain facial dynamics, often emotional expressions or speech articulation have been used. Table 1 shows commonly used audiovisual face databases published between 2000 and 2020 that meet the following criteria: databases which (a) are publicly accessible, (b) are digital recordings, (c) include multimodal (audiovisual, including speech articulation) stimuli, (d) use real human models, and (e) show individual portrayals (i.e., video shooting of single models). As shown in Table 1, in all of these databases, the dynamics are obtained by speech articulation by models in different emotional states (e.g., interested, worried, happy, joyful).

Table 1 Summary of audiovisual face databases in the literature

Name of the database	Number of models	Number of stimuli	Duration of stimuli	Type of speech	Expression elicited	Ethnicity	Language
SUDFace	*50*	*150*	*60 seconds*	*Two recited and one Free Speech*	*Neutral*	*Turkish*	*Turkish*
GEMEP (Bänziger et al., 2012)	10	1260	2.29 seconds (average)	Two standardized sentences	Joy, amusement, pride, pleasure, relief, interest, admiration, tenderness, surprise, rage, panic fear, despair, irritation, anxiety, sadness, disgust, contempt, and shame	Not stated	French
The EU-Emotion Stimulus Set (O'Reilly et al., 2016)	19	418	2–52 seconds	Scripted scenarios (independent of the video footage)	Anger, disgust, afraid, happiness, sadness, surprise, ashamed, bored, disappointed, excited, hurt, interested, joking, jealous, kind, proud, sneaky, frustrated, unfriendly, worried, and neutral	Mainly Caucasian, African American	English, Swedish, Hebrew
RAVDESS (Livingstone & Russo, 2018)	24	7356	1 sentence long	Lexically matched speech and song	Calm, angry, fearful, sad, happy, disgust, surprise, and neutral	Mostly Caucasian, East Asian and Black Canadian	English
IEMOCAP (Busso et al., 2008)	10	not stated	4.3–4.6 seconds	Scripted and spontaneous speech	Happy, sad, anger, frustration, and neutral	Not stated	English
DaFEx (Battocchi et al., 2005)	8	1008	4–27 seconds	Scripted sentences (utterance condition) Includes no-utterance condition too	Happiness, surprise, fear, sadness, anger, disgust, and neutral expression	Not stated	Italian
MPI Facial Expression (Kaulard et al., 2012)	19	20.000	4.31 seconds (average)	Scripted sentences (independent of the video footage)	56 different expressions	Not stated	German
Audiovisual Database of Emotional Speech in Basque (Navas et al., 2004)	1	665 items (not related to emotion), 450 items (related to emotion)	1 hour and 35 minutes in total	Acted speech	Sadness, happiness, anger, fear, surprise, and disgust	Not stated	Basque
BAUM-1 (Zhalehpour et al., 2017)	31	1222	4.07 seconds (average)	Acted and spontaneous speech	Happiness, sadness, fear, anger, disgust, confusion, boredom, interest	Turkish	Turkish

Similar to other existing databases mentioned above (for a comprehensive list of face databases see: https://www.face-rec.org/databases/), neutral expressions are included only as a small subset within the dynamic face databases. Moreover, most of the dynamic face databases which include neutral expressions as a subset either contain clips of very short durations (IEMOCAP [Interactive Emotional Dyadic Motion]: 4.2–4.6 seconds; Busso et al., 2008; GEMEP [Geneva Multimodal Emotion Portrayals]: mean duration of 2.29 seconds; Bänziger et al., 2012; MPI: mean duration of 4.31; Kaulard et al., 2012; Moving Faces and People Database: mean duration of 6.03 seconds; O'Toole et al., 2005) or contain very few models (e.g., DaFEx [Database of Facial Expressions]: 8 models, Battocchi et al., 2005; The EU-Emotion Stimulus Set: 19 models, O'Reilly et al., 2016; SAVE Database: 20 models, Garrido et al., 2017). Here, we complement the existing databases with a new database containing a large set of long clips (60 seconds) of dynamic, neutral expressions by 50 models, which will be useful for a wide range of applications.

Importance and applications of audiovisual dynamic neutral face databases

Developing a face database with neutral expressions is important, as *neutralness* (or non-emotive states) of faces is frequently encountered in daily life. "Neutral facial expressions," defined as expressions that do not show any emotion, have been suggested to occur when faces display no facial muscle contractions (Tian & Bolle, 2003). Importantly, neutralness has been considered a distinct category similar to the six distinct basic emotions (Etcoff & Magee, 1992; Matsumoto, 1983). The categorical distinction of neutral from other emotions is supported by studies, which showed that people perceive categorical boundaries between emotions and non-emotive (neutral) expressions (Etcoff & Magee, 1992; Matsumoto, 1983). Hence, it was suggested that neutral faces are perceived not as faces that contain low degrees of emotionality, but instead as categorically different from all other emotions (Etcoff & Magee, 1992). On the other hand, it has been reported that six basic emotions can be perceived in neutral faces (Albohn & Adams, 2021). One possible explanation might be related to individuals having different temperaments, defined as "the constellation of inborn traits that determine a child's unique behavioral style and the way he or she experiences and reacts to the world" (Kristal, 2005), which may affect their neutral facial expressions. For instance, an individual who has a positive temperament may have a "neutral" face, which might be perceived more positively compared to individuals with negative temperaments. In dynamic situations, this might be because the face muscles, even when individuals are in non-emotive states, yield subtle, unintentional movements (Albohn & Adams, 2021). Following Ekman and Friesen (1978), here, we define a (dynamic) neutral face as a face with limited muscle contraction which does not lead to intense facial expression (neither positive nor negative). In contrast to static faces, neutralness in dynamic faces cannot be defined as faces "without any muscle contractions" since the dynamics are due to muscle contractions.

Facial characteristics, such as facial maturity and attractiveness (Zebrowitz, 1997), both of which can be extracted from neutral faces, are crucial sources when forming impressions about others. In order to study such facial characteristics without the confounding effect of emotional expressions, faces with neutral expressions have frequently been used (e.g., Carré et al., 2009; Hess et al., 2000; Marsh et al., 2005; Said et al., 2009). For instance, presenting neutral facial expressions, it was shown that facial characteristics that resembled expressions of happiness were rated as more trustworthy than faces without these characteristics (Jaeger & Jones, 2021). The perception of neutral faces also depends on group membership (Dotsch et al., 2012; Todorov et al., 2015; Zebrowitz et al., 2010). Research on race differences showed that neutral facial expressions of Caucasians resembled angry faces more strongly compared to other races (Blacks and Koreans; Zebrowitz et al., 2010). Also, it has been shown that stereotypical sex differences influence the perception of neutral faces: neutral male faces were perceived as angrier than neutral female faces, and neutral female faces were perceived as more surprised (Becker et al., 2007; Zebrowitz et al., 2010), cooperative, joyful, and less angry than male faces (Adams et al., 2012; Hareli et al., 2009; MacNamara et al., 2009). Moreover, neutral facial expressions were also used as ambiguous social stimuli in normal and clinical samples when investigating social perception (Cooney et al., 2006; Yoon & Zinbarg, 2007). It has been shown that in some clinical samples, neutral faces were perceived as threatening (i.e., social anxiety; Yoon & Zinbarg, 2007). Taken together, many studies investigated face perception using neutral facial expressions; however, as outlined below, there is a lack of available stimuli with neutral facial expressions.

Neutral faces are usually included only as a subset of face databases. As a consequence, the aforementioned studies mostly used different databases for their neutral face stimuli. For instance, while MacNamara et al. (2009) used face images from a single database (i.e., International Affective Picture System; Lang et al., 2005), Zebrowitz et al. (2010) and Adams et al. (2012) combined face stimuli from various databases which included small subsets of neutral faces (e.g., Pictures of Facial Affect; Ekman & Friesen, 1976, the Montreal Set of Facial Displays; Beaupré et al., 2000, the AR Face Database; Martinez & Benavente, 1998). Another approach was to generate a new database consisting of

desired facial expressions (Hareli et al., 2009). All the stimuli in these studies consisted of static neutral faces.

To develop dynamic face databases, various methods and techniques have been used. Most dynamic face databases are developed in a single modality (i.e., visual). For instance, the Amsterdam Dynamic Facial Expression Set (van der Schalk et al., 2011), the SAVE database (Garrido et al., 2017), BU-4DFE (Yin et al., 2008), and the Extended Cohn-Kanade Dataset (Lucey et al., 2010) are some of the databases in which the dynamic information is introduced by emotional facial expressions (e.g., pulling the lips upward for a smile). These face databases are often used to investigate how different emotional facial expressions are perceived (Abdulsalam et al. 2019; Esins et al., 2016; Wu & Lin, 2018). To develop naturalistic face databases, another modality (i.e., auditory through speech articulation) was introduced to elicit typical face dynamics. Thus, these face databases mostly include a combination of emotional facial expressions and speech articulation (e.g., MOBIO [Mobile Biometrics]: McCool et al., 2012; RAVDESS [Ryerson Audio-Visual Database of Emotional Speech and Song]: Livingstone & Russo, 2018; see Table 1 for the detailed list), and were often used to generate computational models of facial movements of emotional expressions (e.g., Adams et al., 2015; Fridenson-Hayo et al., 2016; Issa et al., 2020; Kaulard et al., 2012; Sagha et al., 2016) and to investigate the neural correlates of distinct aspects of temporal sequences (such as increasing or decreasing and natural versus artificial emotions) during dynamic face perception (Reinl & Bartels, 2014).

The current study

In the current study, we developed and validated a neutral and natural multimodal dynamic face database. The Sabancı University Dynamic Face (SUDFace) database provides a standardized set of multimodal (audiovisual) stimuli of dynamic human faces with speech articulation. This database, freely accessible for research purposes, includes dynamic neutral (and natural) facial expressions captured from 50 models articulating three different speeches. Two of the speeches were scripted texts with sections from the Turkish National Anthem (Ersoy, 1921) and Atatürk's Address to the Turkish Youth (Atatürk, 1927). The final speech was on a topic of each model's choice (we will refer to it as Free Speech hereafter). We recorded videos with a duration of 60 seconds for each speech. Additionally, the current study provides detailed information of physical properties of the faces (e.g. pixel-wise calculations of the face and nose lengths, eye width) and speeches (e.g., duration of speech and repetitions), as well as the detailed script of the articulated speeches (for detailed information visit the link provided for the Open Science Framework [OSF] in the "Availability of data and materials" section, also see "Video Transcriptions" folder in OSF). The psychological dimensions of perceived neutralness, naturalness, valence, and the mental states of the models were quantified in a validation experiment with 227 participants who judged videos of the three speech types of the 50 models on these dimensions.

As mentioned earlier, only a small subset of the existing dynamic face databases contain neutral expressions. Considering the crucial role of neutral facial expressions in daily life, it is surprising that researchers only recently started to systematically investigate their perceptions (Jaeger & Jones, 2021). However, current dynamic face databases do not include systematically developed large-scale stimulus sets for neutral expressions. The SUDFace database addresses this lack of suitable stimuli, providing a large set of dynamic neutral faces. One of the key features of the SUDFace database is that all stimuli are neutral facial expressions with minimal variations in the stimulus set. To ensure minimal variation in regard to emotional cues, all recordings that included emotional cues such as emotional movements of the mouth, the eyes, or other facial areas were excluded from the database. Similarly, all facial accessories, including glasses, necklaces, earrings, makeup, and beards, were excluded from the database by instructing the models before the recordings to provide minimal physical variations. The background, illumination, and camera angle settings were identical for each recording. We report detailed stimulus information such as the objective measures of facial features (see "Facial Features and Measurements.xlsx" file in the OSF directory, under the "Dataset" folder) and speech articulations (see "Video Transcriptions" folder for detailed scripts of each model in the OSF directory).

Existing face databases usually consist of a wide range of static emotional expressions. Therefore, to validate facial expressions most studies mainly used emotion identification questions with multiple-choice options for stimuli (e.g., "happy," "sad," "neutral"; Chung et al., 2019; Dalrymple et al., 2013; Ebner et al., 2010; Livingstone & Russo, 2018; O'Reilly et al., 2016; Tottenham et al., 2009), as most databases contain many different emotional expressions. However, in the Chicago Face Database's validation experiment, only the stimuli with neutral faces were validated based on participants' ratings on several psychological dimensions (e.g., threatening, attractive, trustworthy) using a 1–7 Likert scale. Similarly, an increasing number of validation experiments obtained normative ratings of each face stimulus and asked participants to judge certain dimensions of expressions (e.g., valence, arousal, intensity) or other facial features (e.g., age, attractiveness, babyfaceness, ethnicity; Garrido et al., 2017; Langner et al., 2010; Ma et al., 2015; O'Reilly et al., 2016; van der Schalk et al., 2011). Here we employ a similar Likert scale.

In the current study, all the models were required to maintain a neutral facial expression throughout the recording while they were articulating the speeches. Hence, instead of broad categorization questions, we designed a validation procedure similar to the Chicago Face Database (Ma et al., 2015) using a Likert scale. The stimuli of SUDFace were evaluated in terms of neutralness, naturalness, and the valence of the facial expression, as well as the perceived mental state (proud, confused, bored, relaxed, concentrated, thinking, and stressed) of the models. First, we investigated whether the perceived neutralness and naturalness of models changed across different speeches (recited speeches: The Turkish National Anthem, Atatürk's Address to the Turkish Youth, and Free Speech). We expected that Free Speech would have increased naturalness compared to the recited speeches, as reciting a speech from memory may require higher levels of concentration which is expected to be associated with participants' perception (i.e., higher levels of concentration) that may decrease the perceived naturalness. Overall valence ratings were expected to be zero (i.e., neutral). To investigate the consistency of the (neutral) facial expressions throughout each video we evaluated whether the neutralness and naturalness of the models remained constant. In particular, we selected three 4-second clips from each 60-second-long video: beginning (7–11 seconds), middle (28–32 seconds), and end (56–60 seconds). Each segment was evaluated in regard to its neutralness, naturalness, the valence of the facial expression, and the perceived mental state by a different group of participants.

Development of the SUDFace database

Methods I

Models

A total of 70 adults were recruited for the database development. All models were Turkish and Sabancı University students. None of the models were professional performers. Recordings of 14 models were excluded due to technical problems and six models were excluded as they wore makeup. After the post-recording eliminations, the database includes recordings of 50 individuals (25 female, 25 male; age range: 19 to 25 years; *M*: 22.34; *SD:* 1.57). All models articulated three different speeches during the recording session. All models were asked before the scheduled recording session to memorize the aforementioned two texts (the Turkish National Anthem: Ersoy, 1921; Atatürk's Address to the Turkish Youth: Atatürk, 1927), and to be prepared to freely talk for 60 seconds. Models were asked to maintain their neutral facial expressions during all recordings. After each of the speeches, a short break was given.

The study was approved by Sabancı University's ethics committee (SUREC; No. FASS 2018-61). The majority of the models were recruited through the Sabancı University Recruiting System (SONA) and received course credit for their participation. One model was recruited through online announcements. Models were informed that they could quit the video shooting at any time. All models agreed that their videos could be used for non-commercial research purposes. Models gave written consent for participation and online distribution of the database.

Stimuli

The SUDFace database contains 150 audiovisual stimuli, consisting of different speech articulations with neutral facial expressions, recorded from 50 models. Speech types consist of two recited and one Free Speech. Specifically, all models articulated the three different speeches in three different videos as follows: (1) recited speech 1: scripted text which contains the first two verses of the Turkish National Anthem (Ersoy, 1921), (2) recited speech 2: scripted text which contains the first five sentences of Atatürk's Address to the Turkish Youth (Atatürk, 1927), and (3) Free Speech: models were asked to talk about a topic of their choice. Importantly, all speeches were articulated with a neutral face. This was accomplished by specifically instructing models to maintain a neutral facial expression throughout each speech. The National Anthem and Atatürk's Address to the Turkish Youth were chosen as scripted texts because most Turkish people are familiar with these texts, making it more convenient to memorize them. The two recited speeches included repetitions of certain parts of the texts (see below).

The language of all videos in the database is Turkish. There are 150 videos in total, with the three speech types in each of the 50 individual models. All videos are 60 seconds long and recorded with 60 frames per second in MP4 format and 1080p resolution. The illumination, background, outfit of the model (white T-shirt), and camera angle were kept constant (Fig. 1).

The scripted texts were the same for all models. Due to the long video duration (60 seconds), models were instructed to repeat the scripted text from the beginning if they had finished before the entire 60 seconds of recording time. Therefore, even though all videos of the Turkish National Anthem (and the Address to the Turkish Youth) started with the same opening sentence, they ended with different sentences depending on the model's articulation speed. To describe the speech properties of the scripted texts, we calculated the articulation duration per cycle separately for each text. In particular, one cycle contains the complete first two verses of the National Anthem and five complete sentences of Atatürk's Address to the Turkish Youth. The mean articulation durations per cycle can be found in Supplementary

Fig. 1 Example stimulus of the SUDFace database. *Note.* Single-frame examples of three different participants were taken from the videos.

Material 1. Free Speech was not included in this analysis as it did not require repetition. We also provide texts of all the speeches as supplementary documents through OSF, under the "Video Transcriptions" folder (please see the OSF link in the "Availability of data and materials" section).

Recording setup

For the video recordings, we used a Canon EOS 6D Mark II camera with a Canon 24-105 mm macro 0.45 m/1.5 ft lens. The focal length of the lens was kept at 95 mm for all recordings. The angle of view was kept constant at 13.2° × 9.6° × 16.3° in horizontal, vertical, and diagonal dimensions, respectively (for 35 mm sensor format for videos, also known as a full frame). The camera was placed on a Manfrotto 290 Xtra tripod. As a background, a green curtain was used. The camera was placed at a distance of 1.55 meters from the models and zoomed in according to the standardized grid requirements (see the "Recording process" section for further details). The head and upper shoulders of the models, and the green background were visible in the display (Fig. 2). Fluorescent light was used to illuminate the scene from above. The illumination level of the room was 260 lux, measured with a Sekonic L-608CINE Light Meter. All these properties were kept constant for each video recording.

Recording process

The SU students (models) were informed about the recordings and details about the procedure through email or face-to-face communication. Before coming to the recording session, all models were asked to memorize the scripted texts they received at least a day before the recordings. The order of the speech articulation was (1) National Anthem, (2) Speech to Youth (from now on this is how we refer to Atatürk's Address to the Turkish Youth), and (3) Free Speech, respectively. However, if a participant could not complete a particular speech for any reason, we moved to the next speech and repeated the missing one(s) at the end of the session. Crucially, to reduce contextual variance, all models were required to tie their hair, wear a basic white T-shirt and take off all face accessories (e.g., earrings, piercings, headband, makeup, beards).

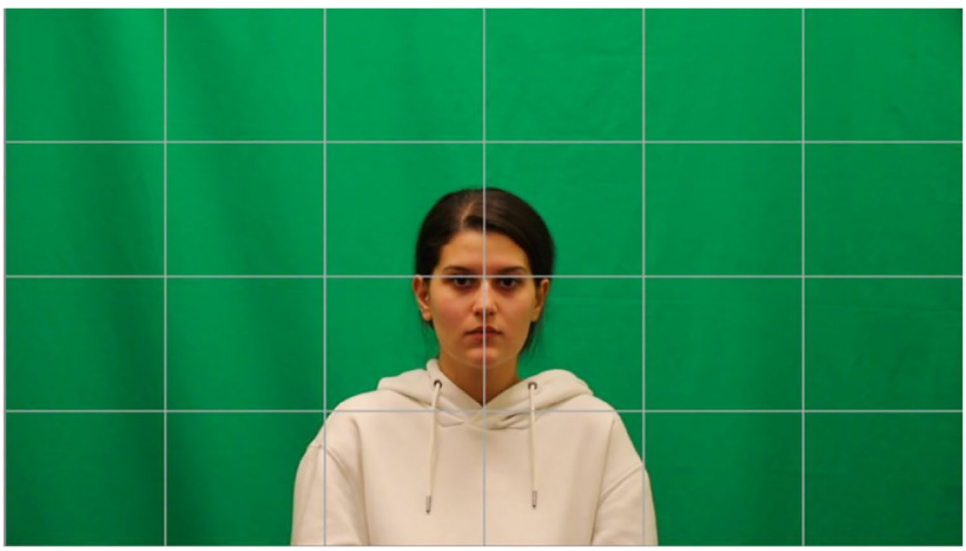

Fig. 2 Example of a model's position. *Note.* The head of the model was positioned in the central four squares so that the central horizontal line of the grid was at eye level, and the central vertical line was aligned with the midline of the face.

Models were asked to stand as still as possible during the recordings to minimize head movements. To control the head position, we located the model's faces in the center of a 4 × 6 camera grid (Fig. 2). Two research assistants were involved in the recording sessions to ensure the accurate localization of the head and monitor the neutral emotional expressions during the recordings. The EOS Utility software (by Canon) enabled us to monitor the models' movements during the recordings. When a research assistant detected that a model displayed an emotional expression or moved her/his head outside the central four squares while recording, the video was re-recorded. If a video still did not satisfy the criteria, it was excluded from the database during post-production. The aforementioned settings were identical for each recording session to eliminate any distinguishable emotional variations of the face or environment.

Objective measures of facial features

We measured the physical features of the faces in the database. This enabled us to analyze the variations in the physical characteristics of our stimuli. In particular, we measured the pixel-wise differences of the face length, nose length, nose width, nose shape, forehead length, chin length, chin size, eye height, eye width, eye shape, eye size, and face width at mouth distance and face roundness. We followed the formulations provided in the Chicago Face Database (Ma et al., 2015), making our results directly comparable to the Chicago Face Database. Table 2 describes a list of formulas for the computed measurements. To make the measurements, the two research assistants followed the illustration in Fig. 3 and the guidelines in Table 2, and made the measurements independent from each other. Later, the inter-rater reliability of the facial features was computed (see section: "Analyses of the objective measures of facial features").

Validation of the SUDFace database

All of the recordings were completed without significant head movements. However, there was some minor variation in facial expressions. To provide further detailed information about stimulus properties, and to quantitatively validate neutral facial expressions, two validation experiments were conducted. Participants (raters) evaluated all 150 videos of the SUDFace database (recordings of the 50 models for the three speeches; National Anthem, Speech to Youth, and Free Speech), in regard to facial neutralness, naturalness, and perceived valence using a Likert scale. Additionally, each stimulus was evaluated in regard to seven perceived mental states of the models: proud, confused, bored, relaxed, concentrated, thinking, and stressed, following the BAUM-1 (Zhalehpour et al., 2017). Importantly, we included "proud" as the scripted texts could be related to (national) pride, and speech recognition may affect the perceived mental state. This set of seven mental states did not include any extreme or complex mental states, such as aggression or panic, as they were not expected to describe our neutral (or very close to neutral) facial expressions in any sensible way. "Neutral" was not included among the mental states to choose from as this would have potentially led to a strong "neutral" bias, leading participants to choose almost exclusively "neutral" as any other emotions were barely visible. By using the seven mental states, we also increased the probability of participants detecting minor emotional variations. Previous validation studies made extensive use of Likert scales when

Table 2 Measurements of facial features

Facial feature	Measurement
Face length	Distance between bottom of chin to the edge of the top of forehead/hairline
Nose length	Distance between nose tip and the upper edge of eyes at nose tip center
Nose width	Distance between the outside edge of the nose at widest point
Nose shape	(Nose width) ÷ (Nose length)
Forehead length	Distance from center of the top of forehead/hairline to the center between the eyes at pupils
Chin length	Distance from bottom edge of lips to base of chin
Chin size	(Chin length) ÷ (Face length)
Eye height	Distance between upper and lower inner eyelid at pupil center (right and left measured separately and averaged)
Eye width	Distance between inner and outer corner of eye (right and left measured separately and averaged)
Eye shape	(Eye height) ÷ (Eye width)
Eye size	(Eye height) ÷ (Face length)
Face width at mouth distance	Distance between outer edges of cheeks at mid-mouth
Face roundness	(Face width at mouth) ÷ (Face length)

Note. All (definitions of) features are taken from the Chicago Face Database (Ma et al., 2015), and this table is adjusted accordingly.

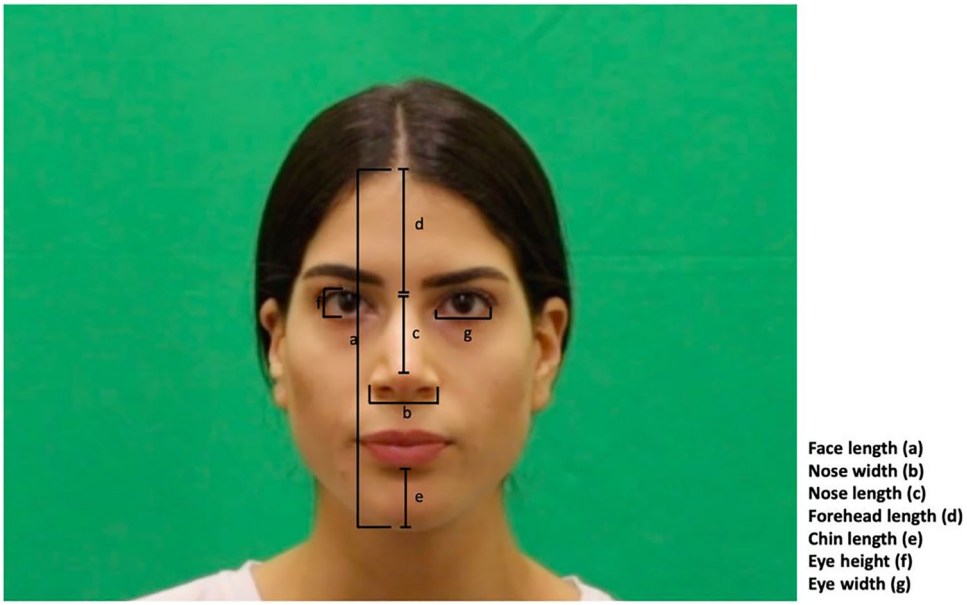

Fig. 3 Facial measurements guide. *Note.* The calculation of each measure can be found in Table 2.

evaluating the stimuli on different psychological dimensions (Garrido et al., 2017; Langner et al., 2010; Ma et al., 2015; McEwan et al., 2014; O'Reilly et al., 2016), and testing the recognition of mental states by forced-choice paradigms (Schmidtmann et al., 2020). We adopted similar quantitative methodologies when gathering norming data about the SUDFace.

We ran two different validation experiments. In the first experiment, the participants evaluated each video's 4-second segment between the 7th and 11th seconds. This particular range was chosen to minimize the possibility of participants understanding the speech content through lip reading. In the second validation experiment, we extracted three 4-second segments from each video, taken from the beginning (seconds 7–11), middle (seconds 28–32), and end of the videos (seconds 56–60). In a between-subjects design, we presented videos from one of the three time segments and asked participants to evaluate each video. The set of questions was kept the same as in the first experiment. This allowed us to test to what extent the expressions remained constant throughout the video recordings.

Methods II

Participants (raters)

A total number of 227 raters has been recruited online for the validation experiment (75 male, and 152 female; between 18 and 29 years of age, $M = 21.51$, $SD = 1.56$). 53 raters were excluded because they did not meet one of the inclusion criteria (see procedure section). Hence, neutralness and naturalness responses of 174 raters (62 male, and 112 female; between 18 and 29 years of age, $M = 21.61$, $SD = 1.72$) were used. The majority of the raters were Turkish speakers ($N = 150$; 86.2%), followed by Urdu ($N = 7$; 4.0%), Arabic ($N = 4$; 2.3%), and other languages ($N = 13$; 7.5%), including Russian, Azerbaijani, German, Spanish, Italian, and Moroccan. Among the 24 non-Turkish speakers, five (23.5%) did not speak or understand Turkish, and 19 (76.5%) spoke and understood Turkish at different levels. Their level of understanding was as follows: nine (47.4%) beginner, four (21.1%) elementary, two (10.5%) intermediate, two (10.5%) advanced, two (10.5%) proficient.

In Validation Experiment 1, 84 participants rated the beginning segments (7th to 11th second) of each video. In Validation Experiment 2, two groups of participants rated the middle (28th to 32nd second; $N = 42$) and end segments (56th to 60th second; $N = 48$). Due to the unequal sample sizes between time segments, we subsampled the larger groups by choosing random participants to equalize all groups when an analysis required comparing data from different time segments.

The validation experiment was approved by the Sabancı University ethics committee (SUREC; No. FASS 2020-58). The majority of the raters were Sabancı University students, recruited via the SONA system, and gained course credit in return. Other raters were recruited by online announcements. Raters gave online informed consent to proceed with the experiment.

Procedure

Each of the 150 audiovisual stimuli in the database was evaluated in terms of the perceived level of neutralness, naturalness, valence, and mental states. The experiment was conducted online using Qualtrics software (Qualtrics, Provo, UT).

All the necessary information about the experimental procedure was explained on the first page. To ensure that we recruited unique raters, they were required to generate a custom ID at the beginning of every experiment. This consists of (1) the first two letters of their name, (2) the first two letters of their mother's name, (3) the last four digits of their phone number, (4) the date (XX) and month (XX) of their birthday. For instance, "zaey10210311" would be an example of a custom ID (note that the given example is not taken from the experiment's ID list but randomly generated for clarification).

In the validation experiment, in each trial, a randomly chosen stimulus (a [muted] dynamic face video; high definition: 1920 × 1080) from the database was presented for four seconds, and followed by three questions. Raters were asked to answer the following two questions using a 1–7 Likert scale (1 = *least* and 7 = *most*): (1) how neutral is the expression (level of neutralness of the expression), and (2) how natural is the expression (level of naturalness of the facial expression). Moreover, raters were asked to indicate the valence of each video using a Likert scale (−3 = *very negative*, 0 = *neither negative nor positive*, +3 = *very positive*). Lastly, raters were also asked to evaluate each stimulus in terms of the most prominent mental state of the model, selecting from the following options: proud, bored, stressed, confused, relaxed, concentrated, and thinking. The order of the choices was randomized in every trial. All questions were presented on a single page. The experimental design can be seen in Fig. 4. The database evaluation was completed in two blocks, in which the stimulus set was randomized. Raters had the possibility to take a 10-minute break between the blocks if desired. The experiment was self-paced, and completed within approximately 90 minutes.

Before finalizing the experiment, raters were required to evaluate the overall level of confidence in their response on a 0–100 scale (0 = *not confident*, 100 = *highly confident*). This was an overall evaluation for all responses done at the end of the experiment. Additionally, they were asked whether they participated as a model in the video-recording process of the database, and whether they understood the content of the presented video. If raters had participated as models, they would not be allowed to participate in validation experiments. Finally, raters were also asked to indicate whether they understood any of the spoken content. Here, we provide analyses of the raters who did not understand the spoken content (please see analyses of the Validation Experiment 1 and 2 for further details. Analyses of raters who understood the spoken content can be found in Supplementary Material 3).

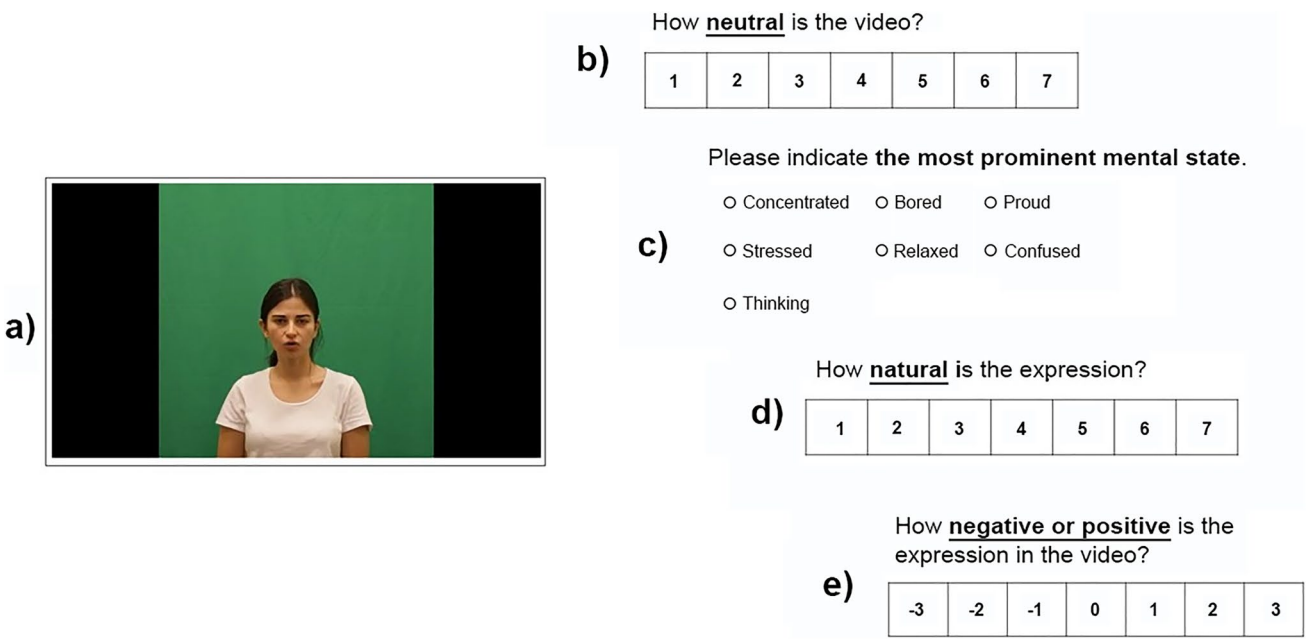

Fig. 4 Example of the experimental design**.** *Note.* In the experiment, all the parts (**a**–**e**) were displayed on a single page. In the first (**b**) and third (**d**) questions, participants rated the neutralness and naturalness of the stimulus on a 7-point scale (1 = *not neutral [natural]*, 7 = *very neutral [natural]*). In the last question (**e**), participants rated the valence of the stimulus on a bidirectional scale ranging from −3 (*very negative*) to 3 (*very positive*).

Results

We ran a series of analyses to characterize and quantify our database. All models memorized the same part of the Turkish National Anthem and Speech to Youth prior to the video recording and repeated these parts within the video duration (60 seconds). The duration of each cycle differed across recited speeches due to the different articulation speeds of the models. First, we measured the articulation durations across multiple repetitions and compared the differences in the articulation durations across multiple repetitions of the recited speeches (These results are provided in Supplementary Material 1). Second, we computed the inter-rater reliability scores (Table 3) for the objective measurements of facial features (see "Facial Features and Measurements.xlsx" file in the OSF directory, under the "Dataset" folder). Third, in the validation experiments, participants rated the neutralness, naturalness, valence, and the most prominent mental state of each video. We report the descriptive statistics for these ratings (Table 4). We ran three repeated-measures analyses of variance (ANOVAs) on the subjective ratings of participants who completed Validation Experiment 1 with more than 50% confidence. Speech types (National Anthem [NA], Speech to Youth [SY], and Free Speech [FS]) and models' sex (Female and Male) were within-subjects factors. Fourth, we investigated whether neutralness, naturalness, and valence ratings were stable across different segments of the videos by running three ANOVAs on the data from Validation Experiment 2 (beginning, middle, and end segments). These analyses were all conducted on the data from the subset of participants who did not understand the spoken content of the videos (discussed below). Additionally, we also provided the results of the participants who understood the spoken content (see Supplementary Material 3). Finally, to further investigate emotional consistency within the 60 seconds, we asked two research assistants to indicate deviations from neutral facial expressions, and computed agreement between two research assistants.

Table 3 Inter-rater reliability of the objective measures of facial features

Face features	Correlation coefficients
Nose width	0.91***
Nose length	0.89***
Nose shape	0.74***
Forehead length	0.82***
Chin length	0.88***
Chin size	0.81***
Eye height	0.93***
Eye width	0.90***
Eye shape	0.81***
Eye size	0.87***
Face width	0.91***
Face length	0.96***
Face roundness	0.74***
Overall measurements	0.95***

Notes. *P*-values correspond to *$p < .05$, **$p < .01$, ***$p < .001$. Correlations are calculated in terms of Spearman's rho coefficient.

Analyses of the objective measures of facial features

The physical measurements of the facial features have been taken independently by two research assistants following the descriptions in Table 2, and can be found in the "Facial Features and Measurements.xlsx" file provided in the OSF directory, under the "Dataset" folder (see "Availability of data and materials section" for the link). Two research assistants calculated pixel-wise measures of nose width, nose length, nose shape, forehead length, chin length, chin size, eye height, eye width, eye shape, eye size, face length, and face roundness. The inter-rater reliability of the measurements was obtained through Spearman's rho correlation coefficients. The inter-rater reliability of the overall measurements of the physical features was very high (.95; the inter-rater reliability for each facial feature can be found in Table 3).

As shown in Table 3, although all of the correlation coefficients are very high, there is some variance in the

Table 4 Descriptive statistics of the evaluations of neutralness, naturalness, and valence

	N	Neutralness		Naturalness		Valence	
		Mean	*SD*	Mean	*SD*	Mean	*SD*
Database (overall)	8400	4.54	1.56	4.3	1.62	−0.01	1.25
National Anthem	2800	4.58	1.56	4.23	1.65	−0.06	1.23
Speech to Youth	2800	4.56	1.53	4.28	1.61	−0.01	1.21
Free Speech	2800	4.47	1.59	4.38	1.62	0.05	1.29

Notes. Neutralness and naturalness were rated on a Likert scale (1–7). Valence was rated on a bidirectional scale ranging from −3 to 3.

inter-rater reliability across the different facial features. For instance, while the inter-rater reliability is very high for face width (0.91) and length (0.96), it decreases for the nose shape (0.74) and face roundness (0.74). Different levels of difficulty to measure these facial features likely underlie the variance of inter-rater reliabilities.

Analyses of the subjective evaluations of raters

In the two validation experiments, participants rated the neutralness, naturalness, valence of the face videos, and the most prominent mental states of each model (proud, confused, bored, relaxed, concentrated, thinking, and stressed).

Analyses of Validation Experiment 1

One of the critical aspects that might influence raters' judgments is understanding the content of the speech by reading the models' lips. Hence, at the end of the experiment we asked all raters whether they understood any part of the speech in any of the videos. Out of the 84 raters who participated in the initial validation experiment, 66.7% of the raters ($N = 56$) indicated that they had not understood any spoken content, while the remaining 33.3% ($N = 28$) reported understanding at least one of the speeches. To not confound purely visual aspects of the videos by extracted semantic information, and to make our results generalizable to a broader population, we analyzed the data separately for the two groups. Specifically, we focused on the findings from the 56 raters who did not understand any of the spoken content. Figure 6 shows the percentages of the perceived mental states for each speech type. Table 4 shows the descriptive statistics for the neutralness, naturalness, and valence ratings. Supplementary Material 2 contains detailed descriptive statistics separately for each speech type/model. (Additionally, we provide analyses for participants who understood the spoken content in Supplementary Materials 2 and 3).

Analyses of Validation Experiment 1 (neutralness, naturalness, valance) To test whether the different speech types (NA, SY, and FS) had an effect on the ratings of neutralness, naturalness, and valence, we ran three repeated-measures ANOVAs on raters' subjective evaluations from Validation Experiment 1 (Beginning; $N = 56$). In all three analyses, speech type was used as a within-subjects factor. We also included the models' sex as a second within-subjects factor to explore any effects of sex.

The assumption of sphericity was violated for speech type in the neutralness ($\chi^2(2) = 38.74$, $p < 0.001$, $\varepsilon = 0.97$) and valence ($\chi^2(2) = 14.09$, $p < 0.001$, $\varepsilon = 0.99$) analyses, and for the speech type*models' sex interaction in the naturalness analysis ($\chi^2(2) = 15.12$, $p < 0.001$, $\varepsilon = 0.99$). Since ε values were greater than 0.75 for all three cases, we report Huyn-Feldt-corrected results where sphericity was violated.

There were main effects of speech type on neutralness ($F(1.95, 2727.31) = 7.41$, $p < 0.001$, $\eta^2_p = 0.005$), naturalness ($F(2, 2798) = 13.41$, $p < 0.001$, $\eta^2_p = 0.009$), and valence ($F(1.98, 2774.13) = 9.86$, $p < 0.001$, $\eta^2_p = 0.007$). Post hoc tests revealed that FS was significantly different than NA and SY for all three dependent variables. FS videos ($M = 4.47$, 95% CI [4.4, 4.54]) were rated as less neutral than NA ($M = 4.58$, 95% CI [4.51, 4.64], $p_{bonf} = 0.001$) and SY videos ($M = 4.57$, 95% CI [4.5, 4.63], $p_{bonf} = 0.005$; Fig. 5a). In contrast, FS videos ($M = 4.39$, 95% CI [4.31, 4.46]) were rated as more natural than both NA ($M = 4.23$, 95% CI [4.16, 4.3], $p_{bonf} < 0.001$) and SY videos ($M = 4.29$, 95% CI [4.22, 4.36], $p_{bonf} = 0.005$; Fig. 5b). In terms of valence, FS videos ($M = 0.05$, 95% CI [0.003, 0.1]) were rated as more positive than NA ($M = -0.06$, 95% CI [−0.11, −0.01], $p_{bonf} < 0.001$) and SY videos ($M = -0.009$, 95% CI [−0.06, 0.04], $p_{bonf} = 0.045$; Fig. 5c). NA and SY ratings did not significantly differ in neutralness ($p_{bonf} = 0.99$), naturalness ($p_{bonf} = 0.15$), or valence ($p_{bonf} = 0.14$).

Interestingly, we also found main effects of models' sex on neutralness ($F(1, 1399) = 23.25$, $p < 0.001$, $\eta^2_p = 0.016$), naturalness ($F(1, 1399) = 27.21$, $p < 0.001$, $\eta^2_p = 0.019$), and valence ($F(1, 1399) = 4.23$, $p = 0.04$, $\eta^2_p = 0.003$). Female models were rated less neutral ($M_{diff} = -0.17$, $p_{bonf} < 0.001$) and more natural ($M_{diff} = 0.18$, $p_{bonf} < 0.001$) than male models. In terms of valence, female models were perceived as more positive than male models ($M_{diff} = 0.07$, $p_{bonf} = 0.04$). There were no significant speech type*models' sex interactions in any of the dependent variables (for all tests $p > 0.05$).

Descriptives of the perceived mental state Figure 6 shows the percentages of the perceived mental states separated by speech types (NA, SY, and FS). The perceived mental states for all three speech types had highly similar distributions. Overall, the most prominent perceived mental state in the database was concentrated (28%), followed by bored (17%), relaxed (15%), stressed (14%), thinking (10%), confused (9%), and proud (7%). More detailed descriptive statistics and the most prominent mental states individually for each model are provided in Supplementary Material 2.

Analyses of Validation Experiment 2

To examine whether neutralness, naturalness, and valence were stable throughout the videos, we compared three segments (beginning, middle, and end) from each video in Validation Experiment 2. As discussed in the Methods sections, the number of participants was unequal among the validation experiments (beginning: $N = 84$; middle: $N = 42$; end: $N = 48$), which required subsampling of the

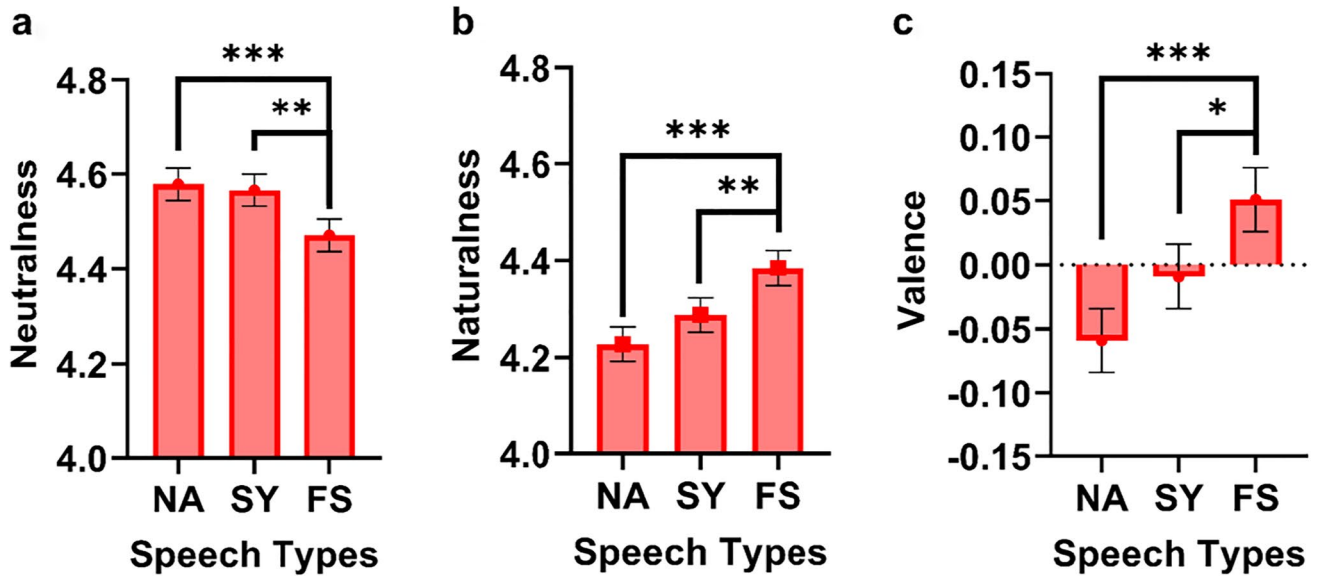

Fig. 5 Neutralness, naturalness, and valence ratings across speech types. *Note.* Average subjective ratings for speech types NA (National Anthem), SY (Speech to Youth), and FS (Free Speech). (**a**) Neutralness and (**b**) naturalness were rated on a Likert scale (1–7). (**c**) Valence was rated on a bidirectional scale ranging from −3 to 3. Asterisks indicate p-values of $*p < .05$; $**p \leq .005$; $***p \leq .001$.

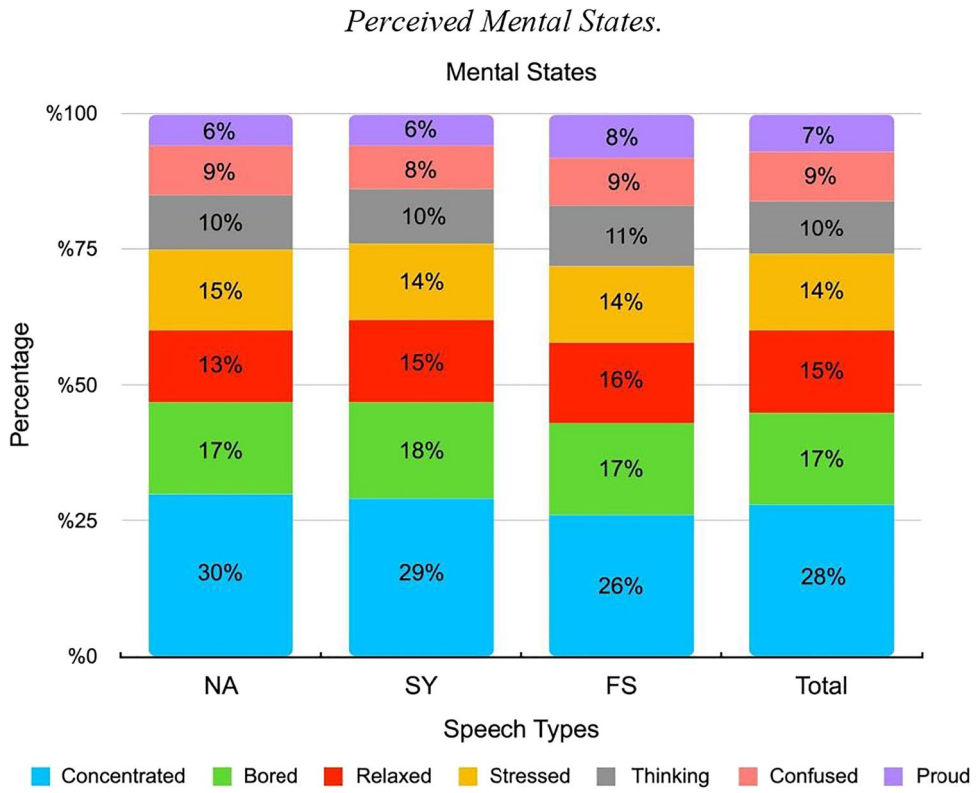

Fig. 6 Perceived mental states. *Note.* These distributions are calculated from 56 participants' ratings of the 150 videos. NA = National Anthem, SY = Speech to Youth, FS = Free Speech, Total = the overall distribution for the whole database.

data to have equal group sizes. As we were only interested in differences in time among the participants who did not understand any spoken content, we excluded participants who understood any type of speech (remaining *N*s = beginning: 56, middle: 27, end: 36) and subsampled them from the remaining participants to equalize all groups. This led to 27 participants for each time segment.

We ran three one-way ANOVAs with time (beginning, middle, and end) as a between-subjects factor on the subsampled data. Time had a significant effect on neutralness ($F(2, 4047) = 45.19$, $p < 0.001$, $\eta^2 = 0.01$), naturalness ($F(2, 4047) = 18.18$, $p < 0.001$, $\eta^2 = 0.01$), and valence ratings ($F(2, 4047) = 23.43$, $p < 0.001$, $\eta^2 = 0.01$). The follow-up post hoc tests revealed a similar results for each dependent variable: For neutralness, beginning segment ($M = 4.58$, 95% CI [4.51, 4.65]) was rated significantly higher than middle ($M = 4.15$, 95% CI [4.08 , 4.22]) and end segments ($M = 4.18$, 95% CI [4.11, 4.25]). Beginning segment ($M = 4.43$, 95% CI [4.36, 4.5]) was also rated the highest in naturalness, while middle ($M = 4.17$, 95% CI [4.09, 4.24]) and end segments ($M = 4.15$, 95% CI [4.08, 4.22]) did not significantly differ. For valence, only the middle segment ($M = -0.19$, 95% CI [−0.24, −0.14]) of the videos was rated significantly lower than the beginning ($M = 0.02$, 95% CI [−0.04, 0.07]) and the end segments ($M = 0.05$, 95% CI [−0.005, 0.1]).

To further investigate whether these differences could also stem from larger changes in a few outliers and to better quantify the changes in expression for all the videos in our database, we evaluated the entire 60 seconds of each video clip regarding emotional changes. Two research assistants assessed all videos and separately rated each second of each video in terms of valence. Thereby, it was possible to extract and compare the onsets, offsets, and direction of expression changes reported by the two assistants. Out of 9000 one-second time frames from all videos, the two assistants agreed on 8686 (96.5%) to be neutral, and 30 (0.34%) to be positive. Ratings from both assistants agreed that 122 out of 150 videos (81.3%) did not have any expression changes. Expression changes that occurred in six (4%) of the videos were agreed upon by both assistants, while only Assistant 2 reported expression changes in the remaining 22 (14.67%). There was no agreement on expression changes in the negative direction (Assistant 1: 0; Assistant 2: 137). Figure 7 shows still images from before and after an expression change occurred in one of the videos in which both assistants reported an expression change (top), and the same for one of the videos for which only one of the assistants reported an expression change (bottom). We provide the onset and offset times of all noted deviations for each clip (in which deviations occurred; see "Expression Changes.xlsx" file in the OSF link, under the "Dataset" folder).

Discussion

In this paper, we introduce a new dynamic face database, the Sabancı University Dynamic Face (SUDFace) database, which consists of three different speeches articulated with neutral facial expressions. This study aimed to develop a highly controlled dynamic (natural and neutral) face database with a large sample of neutral facial expressions and long video durations. The database consists of 1-minute-long videos with three different speeches. Videos were highly controlled by ensuring the head location and orientation (front view of the faces), neutral facial expressions, and controlled illumination and background. The physical features of all faces were measured and are reported in Table 3. In two validation experiments, the videos were evaluated regarding the psychological states of the models (perceived neutralness, naturalness, valence, and mental states). In the first validation experiment, the beginning (seconds 7–11) of each video was rated. In the second validation experiment, different segments (beginning: seconds 7–11; middle: seconds 28–32; end: seconds 56–60) of each video were evaluated to validate the consistency of the models' expressions throughout the videos. As there was a main effect of time, we further investigated emotional variations across the entire 60 seconds by quantifying emotional deviations.

Objective measures of facial features

All reported measurements of facial features are provided as a supplementary document (see "Facial Features and Measurements.xlsx" in the OSF directory, under the "Dataset" folder) and can be used to select subsets of stimuli from the database as needed. The measurements were independently performed by two raters. Inter-rater reliability was high for each feature (above 0.74 for all features) and very high for all features pooled (0.95), indicating that the reported measurements are a good representation of the facial features in our database (see Table 3 for all inter-rater reliability coefficients). Relatively low inter-rater reliabilities were obtained for nose shape (0.74) and face roundness (0.74), presumably because measurements of these features are challenging compared to features such as face length (0.96) and face width (0.90), which contain precisely defined boundaries.

Analysis of the subjective evaluations of raters

Our goal is to provide a dynamic face database with neutral and natural facial expressions, recorded during different types of speeches—two recited: National Anthem and Speech to Youth, one Free Speech. Already during the recordings, we ensured that no major emotions occurred,

Fig. 7 Examples of expression changes. *Note.* Still images from before (left) and after (right) an expression change as reported by the two research assistants that rated all expression changes in the videos. The top is from one of the videos in which both assistants reported an expression change. The bottom is from one of the videos in which only one of the assistants reported an expression change.

and either repeated a recording when non-neutral emotions were detected by the research assistants who evaluated the emotions during the recordings or discarded video clips if deviations from neutral were determined during postproduction. However, subtle emotional changes are highly likely in complex and dynamic situations, as in the speeches given by our models. Hence, to address the possibility of slight variations of emotions, we conducted two validation experiments to quantify perceived deviations from neutral and natural. Additionally, we asked raters to indicate the perceived emotional valence of each video clip.

Analyses of Validation Experiment 1

In the first validation experiment, we evaluated the overall neutralness, naturalness, valence, and potential differences between the speech types (National Anthem, Speech to Youth, and Free Speech).

Analyses of Validation Experiment 1 (neutralnesss, naturalness, valance) The ratings of neutralness, naturalness, and valence were made on Likert scales from 1 to 7 for neutralness and naturalness, and −3 (negative) to +3 (positive)

for valence. Importantly, the entire stimulus set consisted of highly neutral and natural facial expressions. Ratings on these scales do not represent evaluations of the entire (or large parts of the) spectrum of emotions but small (positive or negative) deviations from neutral. Hence, the obtained values should be interpreted accordingly. For example, ratings of maximum positive deviations from neutral (valence = +3) do not represent highly positive emotions, but only "highly" positive relative to "perfectly" neutral stimuli. In the speech types, neutralness ratings ranged between 4.47 and 4.58 (on average 4.54; *SD* = 1.56), naturalness ratings between 4.23 and 4.39 (on average 4.3, *SD* = 1.62), and valence ratings between −0.01 and −0.06 (on average −0.01; *SD* = 1.25). These values seem to slightly favor neutrality and naturalness on our scale (midpoint of 4, maximally neutral/natural: 7), supporting the central characteristic of our dataset. Moreover, these results capture that there were some variations (e.g., although rated as neutral on the neutralness scale, some models' perceived valence (see Supplementary Material 2) was negative (e.g., Mov subj12, Mov subj26, and Mov subj30) or positive (e.g., Mov subj11, Mov subj33, and Mov subj45). Overall, these results indicate that participants successfully made judgments relative to the provided spectrum and not judgments considering the entire spectrum of human emotions. Within a different set of stimuli, including highly positive and negative (as well as highly "unnatural" stimuli), the clips of our database would be expected to yield ratings very close to high neutralness/naturalness.

Notably, while the perceived neutralness, naturalness, and valence levels were highly similar for the two recited speeches (National Anthem and Speech to Youth), they differed from the Free Speech. Surprisingly, the clips of the National Anthem and the Speech to Youth were perceived as more neutral than the Free Speech. The Free Speech, by contrast, was perceived as more natural and more positive (valence) compared to the National Anthem and the Speech to Youth videos. Assuming that the vocalization of scripted texts required more concentration than freely talking about a chosen subject, these results can possibly be explained by facial expressions associated with concentration. Concentration has been proposed to show similar features as confusion and particularly worry (Ekman, 1979; see also Pope and Smith (1994), both rather negative emotions. Hence, the Free Speech expressions may have appeared more positive than the two other speeches.

Descriptives of the perceived mental state In the first validation experiment, participants also rated the perceived mental states of the models. We chose seven mental states (proud, confused, bored, relaxed, concentrated, thinking, and stressed). We did not include (1) any extreme or complex mental states or (2) neutralness as an option, as the latter would have been expected to be chosen highly frequently, not capturing any deviations from neutralness. The most attributed mental state was "concentrated." This might be due to the challenging task and situation of giving a speech while being filmed and being asked to follow several instructions, such as keeping one's posture and neutral facial expression. Similarly, the raters' awareness of the difficulty of the situation could have yielded a bias to choose "concentrated" more often than the other mental states. Additionally, the recited speeches are related to national ideas (e.g., National Anthem and Speech to Youth), which may lead models to feel required to concentrate. However, this does not account for the similar "relaxed" ratings for the National Anthem and Speech to Youth as for the Free Speech.

Emotional consistency across the entire 60-second clips

The second validation experiment (beginning, middle, end) showed an effect of time on neutralness, naturalness, and valence levels. Overall, the results suggested that the beginning was perceived more positively than other segments, while the middle and end ratings were mainly similar. To go beyond the evaluations of the three segments in Validation Experiment 2, and evaluate emotional changes across the entire length of all video clips, two research assistants rated the valence of any changes in expression they detected throughout the entire videos. Taken together, ratings from both assistants showed that changes of expressions are rare and isolated in time, as only 0.34% of the 1-second time frames from all videos in the database contained expression changes agreed upon by both assistants. Only a small number of the videos (6 out of 150) were reported as having expression changes by both assistants. Assistant 2 reported expression changes in 22 videos that were not agreed on by Assistant 1. As illustrated in Fig. 7, these disagreements were caused by very subtle changes in expressions.

In general, the results of the validation experiments were in line with our expectations. Even though there were some variations (see Supplementary Material 4 for further examples), all the videos were rated as highly neutral and natural. In addition to that, the perceived neutralness, naturalness, and valence levels of the SUDFace database were relatively stable across time.

Neutral faces in the literature and advantages of the SUDFace database

In previous studies, neutral face stimuli were commonly evaluated with an identification task in which participants were asked to identify the emotion of stimuli from the provided choices (Chung et al., 2019; Langner et al., 2010; Yang et al., 2020). Importantly, neutral stimuli were usually presented within the same blocks as (highly) emotional

stimuli (Garrido et al., 2017; Ma et al., 2015; O'Reilly et al., 2016; Tottenham et al., 2009). However, previously shown emotional stimuli influence participants' judgments of neutral stimuli. For example, Wyczesany et al. (2018) showed the influence of negative and positive mood (elicited by positive or negative emotional pictures) on the perception of nonemotional, neutral stimuli. Likewise, Anderson et al. (2012) showed the influence of affective information on the perception of neutral faces even in a situation where the cues were incidental to participants. This indicates that presenting participants with affective information can elicit perception of emotions in neutral faces; hence, neutral faces should be rated in isolation from highly emotional faces. In this way, one can objectively measure neutralness of facial expressions without strong influences of previously shown stimuli. Given that one of the main features of SUDFace is to provide highly neutral facial expressions, we eliminated perceptual influences that can be elicited from previously shown emotional stimuli by presenting only neutral stimuli in the validation experiment. Furthermore, by including valence ratings on the faces, we quantified any deviations from neutralness of each model within our database. Deviations from neutralness were very small, and faces were perceived as neither positive nor negative, hence neutral.

Although face perception is highly dynamic, most studies that investigated the underlying neural responses of face perception used static face images. One reason for using static images (despite their disadvantages) is that it is difficult to exclude all factors that are related to face dynamics, such as language and emotion, which modulate neural responses. For instance, when face dynamics are introduced by a moving mouth that results in a happy face over time, the neural responses to the dynamics of the face will be intermingled with the neural responses to its emotional expression. When a dynamic speaking face is shown, it is difficult to separate the underlying neural responses of expression and language. The language of this database is Turkish, which is not a commonly spoken language worldwide. Therefore, in addition to the advantage of having minimized emotional expressions, factors related to language processing, in particular by lip reading, are strongly reduced or abolished in experiments with participants that do not speak/understand Turkish. Hence, our database will facilitate studies on the underlying neural correlates of dynamic face perception by excluding (or minimizing) other processes (i.e., language and emotional processes).

The application areas of the SUDFace

The SUDFace database can be applied in several research areas. The importance and applications of databases with dynamic neutral faces have already been outlined in the introduction (see "Importance and applications of audiovisual dynamic neutral face databases"). Some of the main applications are in the fields of face perception, emotional processing, and psycholinguistics. Importantly, there is a strong need for databases with neutral facial expressions to investigate clinical populations where the perception of neutral faces is essential (Bochet et al., 2021; Cooney et al., 2006; Leppänen et al., 2004; Tottenham et al., 2014). For example, Leppänen et al. (2004) showed that patients with depression perceive neutral faces—but not faces with emotional expressions—differently than healthy people. There are also differences in processing neutral faces between clinical populations, such as those with social anxiety disorder and healthy individuals (Cooney et al., 2006). Moreover, Tottenham et al. (2014) showed that people with autism spectrum disorder (ASD) perceive neutral faces differently than people without ASD, and that they confuse neutral facial expressions with negative expressions (see also, Bochet et al., 2021). The SUDFace database will also be useful in research investigating basic cognitive-perceptual processes and their underlying neural mechanisms. For example, the database was recently used to investigate the underlying neural correlates of temporal integration processes in dynamic face perception (Alp & Ozkan, 2022). It was found that the temporal integration (binding successive frames in time) was more enhanced when the face was displayed in the correct temporal order compared to shuffled or reversed orders. The SUDFace database is an ideal instrument to investigate other questions regarding integration processes, for example, to what extent language processing and language comprehension are influenced by temporal order manipulations. Hence, the SUDFace database will be highly useful for the growing number of studies that investigate neutral facial expressions with or without auditory components, both in healthy and clinical populations.

Beyond studies investigating perception of faces and emotional expressions, the SUDFace database will also be useful in psycholinguistics. As the SUDFace database includes both visual and audio content, it can be used to investigate a range of questions concerning the interplay and integration of audio-visual information, such as effects of visual distraction (Cohen & Gordon-Salant, 2017) and background noise (Leibold et al., 2016) on speech perception. Moreover, previous research has shown that certain speech properties convey information about the personality, emotion, or mood of a person (Boomer & Dittman, 1964; Fay & Middleton, 1941; Ray, 1986; Guidi et al., 2019). One of the speech properties that influences personality perception is pitch variation. Addington (1968) showed that males with higher pitch variations were rated as more energetic and aesthetically inclined, while females were rated as more energetic and extroverted. Similarly, many studies showed the correlation between prosodic features (e.g., mean pitch, pitch variations, or speaking rate) and perception of extraversion, competence, or

dominance (Scherer & Scherer, 1981). Again, the SUDFace database will be highly valuable in these fields as it provides an extensive set of dynamic faces and speech.

The limitations of the current study

There are a few limitations of our database that should be mentioned. First, the SUDFace database only includes Turkish models and the Turkish language. However, our database thereby adds a large stimulus set of Turkish people and the Turkish language to the already existing databases that focus on a single ethnicity (Chinese [The CAS-PEAL-R1]: Gao et al., 2008; Taiwanese [TFEID]: Chen & Yen, 2007; Korean [KUFEC]: Kim et al., 2011; Argentinian [Argentine Set of Facial Expressions]: Vaiman et al., 2017; American [Chicago database: Ma et al., 2015; EU-Emotion Stimulus set, O'Reilly et al., 2016]) and a single language (French [GEMEP]: Bänziger et al., 2012; Italian [DaFEx]: Battocchi et al., 2005; German [MPI Facial Expression]: Kaulard et al., 2012). Second, all the models recruited to develop the database were young and nonprofessional models. Thus, there is a possibility that they did not regulate their emotions as effectively as older adults (Sims et al., 2015) or professional actors. Third, achieving the articulation of a 60-second speech without emotional expressions in such a controlled environment is difficult. Therefore, two trained researchers monitored the whole recording session and the models' facial muscle movements to ensure that a neutral expression was maintained (recordings were repeated until the desired neutral expression was achieved). Fourth, the validation experiment did not include any information regarding the attractiveness or warmth of the models that might also affect the perception of the participants. All the abovementioned points should be taken into account when using the stimulus set.

Conclusion

Most face databases consist of static stimuli. The SUDFace database provides a large set of long dynamic videos (N = 150) with a long duration (60 seconds), and from a large number of models (N = 50). Three speeches per model are provided in the database, all with neutral facial expressions and natural speech articulation. All stimuli were validated in two experiments. Detailed information about all stimuli is provided in the database. The database is freely accessible upon request.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13428-022-01951-z.

Acknowledgment We would like to thank Beril Timuçin, Mahnoor Nadeem, Güneş Şirvancı for their help in video recordings and Mert Yılmaz, Hazal Nil Kırelli, Sahcan Ozdemir, Ceren Saglam, and Selin Yilmaz for their help in data collection and analysis.

Funding This research was supported by the Starting Grant from Sabancı University (B.A. CG-19-01966) to NA.

Availability of data and materials The SUDFace database generated in the development study can be accessed in the "Sabancı University Dynamic Face Database" folder from drive.google.com/drive/folders/1xzxLbza4qiI3XkkHIPycAKyypFspu_vb?usp=sharing. Also, datasets generated during and analyzed during the validation study are available in the "Dataset" folder, https://osf.io/b4vju/?view_only=1dd006d0d4504a2982d31968d9c360f6.

Code availability The video editing code used in the database stimuli can be accessed in the "Video editing code" folder from https://osf.io/b4vju/?view_only=1dd006d0d4504a2982d31968d9c360f6.

Declarations

Ethics approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Sabancı University's ethics committee (SUREC) (for Validation Experiment: Date September 2020 /No. FASS 2020-58; for video recordings Date September 2018 /No. FASS 2018-61).

Consent to participate Informed consent was obtained from all models that participated in the video recordings of the database. Raters in the validation experiment gave their online informed consent.

Consent for publication All models provided informed consent regarding publishing their data and recordings.

Conflicts of interest/Competing interests Yağmur Damla Şentürk, Ebru Ecem Tavacioglu, Ilker Duymaz, Bilge Sayim, and Nihan Alp have no relevant financial or non-financial interests to disclose.

Note The experimental output for this paper was generated using Qualtrics software, Version December 2020 of Qualtrics. Copyright © 2020 Qualtrics. Qualtrics and all other Qualtrics product or service names are registered trademarks or trademarks of Qualtrics, Provo, UT, USA. https://www.qualtrics.com.

References

Abdulsalam, W. H., Alhamdani, R. S., & Abdullah, M. N. (2019). Facial emotion recognition from videos using deep convolutional neural networks. *International Journal of Machine Learning and Computing, 9*(1), 14–19.

Adams, R. B., Nelson, A. J., Soto, J. A., Hess, U., & Kleck, R. E. (2012). Emotion in the neutral face: A mechanism for impression formation? *Cognition & Emotion, 26*(3), 431–441. https://doi.org/10.1080/02699931.2012.666502

Adams, A., Mahmoud, M., Baltrušaitis, T., & Robinson, P. (2015). Decoupling facial expressions and head motions in complex emotions. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 274–280). IEEE.

Addington, D. W. (1968). The relationship of selected vocal characteristics to personality perception. *Speech Monographs, 35*(4), 492–503. https://doi.org/10.1080/03637756809375599

Albohn, D. N., & Adams, R. B. (2021). Emotion residue in neutral faces: Implications for impression formation. *Social Psychological and Personality Science, 12*(4), 479–486. https://doi.org/10.1177/1948550620923229

Alp, N., & Ozkan, H. (2022). Neural correlates of integration processes during dynamic face perception. *Scientific Reports, 12*(1), 1–12.

Anderson, E., Siegel, E., White, D., & Barrett, L. F. (2012). Out of sight but not out of mind: Unseen affective faces influence evaluations and social impressions. *Emotion, 12*(6), 1210–1221. https://doi.org/10.1037/a0027514

Atatürk, M. K. (1927). Nutuk. *Kaynak Yayınları*.

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva multimodal expression corpus for experimental research on emotion perception. *Emotion (Washington, D.C.), 12*(5), 1161–1179. https://doi.org/10.1037/a0025827

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance, 4*(3), 373.

Battocchi, A., Pianesi, F., & Goren-Bar, D. (2005). The properties of DaFEx, a database of kinetic facial expressions. In *International conference on affective computing and intelligent interaction* (pp. 558–565). Springer.

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research, 17*(3), 613–620.

Beaupré, M. G., Cheung, N., & Hess, U. (2000). The Montreal Set of Facial Displays of Emotion [Slides]. (Available from Ursula Hess, Department of Psychology, University of Quebec at Montreal, P.O. Box 8888, Station "Centre-ville", Montreal, Quebec H3C 3P8.).

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology, 92*(2), 179–190. https://doi.org/10.1037/0022-3514.92.2.179

Biele, C., & Grabowska, A. (2006). Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Experimental Brain Research, 171*(1), 1–6.

Bochet, A., Sperdin, H. F., Rihs, T. A., Kojovic, N., Franchini, M., Jan, R. K., Michel, C. M., & Schaer, M. (2021). Early alterations of large-scale brain networks temporal dynamics in young children with autism. *Communications Biology, 4*(1), 1–10. https://doi.org/10.1038/s42003-021-02494-3

Boomer, D. S., & Dittman, A. P. (1964). Speech rate, filled pause, and body movement in interviews. *Journal of Nervous and Mental Disease, 139*(4), 324–327. https://doi.org/10.1097/00005053-196410000-00003

Busso, C., Bulut, M., Lee, C. C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J. N., Lee, S., & Narayanan, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation, 42*(4), 335–359.

Calvo, M. G., Avero, P., Fernández-Martín, A., & Recio, G. (2016). Recognition thresholds for static and dynamic emotional faces. *Emotion (Washington, D.C.), 16*(8), 1186–1200. https://doi.org/10.1037/emo0000192

Cao, H., Cooper, D. G., Keutmann, M. K., Gur, R. C., Nenkova, A., & Verma, R. (2014). Crema-d: Crowd-sourced emotional multimodal actors dataset. *IEEE Transactions on Affective Computing, 5*(4), 377–390.

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science, 20*(10), 1194–1198. https://doi.org/10.1111/j.1467-9280.2009.02423.x

Carrera-Levillain, P., & Fernandez-Dols, J. M. (1994). Neutral faces in context: Their emotional meaning and their function. *Journal of Nonverbal Behavior, 18*(4), 281–299. https://doi.org/10.1007/BF02172290

Chen, L. F., & Yen, Y. S. (2007). *Taiwanese facial expression image database*. Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University.

Christie, F., & Bruce, V. (1998). The role of dynamic information in the recognition of unfamiliar faces. *Memory & Cognition, 26*(4), 780–790. https://doi.org/10.3758/bf03211397

Chung, K. M., Kim, S., Jung, W. H., & Kim, Y. (2019). Development and Validation of the Yonsei Face Database (YFace DB). *Frontiers in Psychology, 10*, 2626. https://doi.org/10.3389/fpsyg.2019.02626

Cohen, J. I., & Gordon-Salant, S. (2017). The effect of visual distraction on auditory-visual speech perception by younger and older listeners. *The Journal of the Acoustical Society of America, 141*(5), EL470. https://doi.org/10.1121/1.4983399

Cooney, R. E., Atlas, L. Y., Joormann, J., Eugène, F., & Gotlib, I. H. (2006). Amygdala activation in the processing of neutral faces in social anxiety disorder: is neutral really neutral? *Psychiatry Research, 148*(1), 55–59. https://doi.org/10.1016/j.pscychresns.2006.05.003

Cunningham, D. W., & Wallraven, C. (2009). Dynamic information for the recognition of conversational expressions. *Journal of Vision, 9*(13), 1–17. https://doi.org/10.1167/9.13.7

Dalrymple, K. A., Gomez, J., & Duchaine, B. (2013). The Dartmouth Database of Children's Faces: acquisition and validation of a new face stimulus set. *PLoS One, 8*(11), e79131. https://doi.org/10.1371/journal.pone.0079131

Dantcheva, A., Bremond, F., & Bilinski, P. (2018). Show me your face and I will tell you your height, weight and body mass index. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 3555–3560). IEEE.

Dotsch, R., Wigboldus, D. H. J., & Van Knippenberg, A. (2012). Behavioral information biases the expected facial appearance of members of novel groups. *European Journal of Social Psychology, 43*(1), 116–125. https://doi.org/10.1002/ejsp.1928

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods, 42*(1), 351–362.

Ekman, P. (1979). About brows: Emotional and conversational signals. In M. von Cranach, K. Foppa, W. Lepenies, & D. Ploog (Eds.), *Human ethology* (pp. 169–248). Cambridge University Press.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124–129. https://doi.org/10.1037/h0030377

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: Investigator's guide*. Consulting Psychologists Press.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*(4), 712–717. https://doi.org/10.1037/0022-3514.53.4.712

Ersoy, M. A. (1921). *İstiklal Marşı*. Retrieved from https://www.tdk.gov.tr/genel/istiklal-marsi-ve-genclige-hitabe/

Esins, J., Schultz, J., Stemper, C., Kennerknecht, I., & Bülthoff, I. (2016). Face perception and test reliabilities in congenital prosopagnosia in seven tests. *i-Perception, 7*(1), 2041669515625797. https://doi.org/10.1177/2041669515625797

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44*(3), 227–240. https://doi.org/10.1016/0010-0277(92)90002-y

Fay, P. J., & Middleton, W. C. (1941). The ability to judge truth-telling, or lying, from the voice as transmitted over a public

address system. *Journal of General Psychology, 24*, 211–215. https://doi.org/10.1080/00221309.1941.10544369

Ferreira-Santos, F. (2015). Facial emotion processing in the laboratory (and elsewhere): Tradeoffs between stimulus control and ecological validity. *AIMS Neuroscience, 2*(4), 236–239. https://doi.org/10.3934/Neuroscience.2015.4.236

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition, 110*(2), 160–170. https://doi.org/10.1016/j.cognition.2008.11.010

Fridenson-Hayo, S., Berggren, S., Lassalle, A., Tal, S., Pigat, D., Bölte, S., Baron-Cohen, S., & Golan, O. (2016). Basic and complex emotion recognition in children with autism: Cross-cultural findings. *Molecular Autism, 7*, 52. https://doi.org/10.1186/s13229-016-0113-9

Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., & Zhao, D. (2008). The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 38*(1), 149–161.

Garrido, M. V., Lopes, D., Prada, M., Rodrigues, D., Jerónimo, R., & Mourão, R. P. (2017). The many faces of a face: Comparing stills and videos of facial expressions in eight dimensions (SAVE database). *Behavior Research Methods, 49*(4), 1343–1360.

Goh, R., Liu, L., Liu, X., & Chen, T. (2005). The CMU face in action (FIA) database. In *International workshop on analysis and modeling of faces and gestures* (pp. 255–263). Springer.

Golarai, G., Ghahremani, D. G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J. L., Gabrieli, J. D., & Grill-Spector, K. (2007). Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nature Neuroscience, 10*(4), 512–522. https://doi.org/10.1038/nn1865

Grainger, S. A., Henry, J. D., Phillips, L. H., Vanman, E. J., & Allen, R. (2017). Age deficits in facial affect recognition: The influence of dynamic cues. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 72*(4), 622–632. https://doi.org/10.1093/geronb/gbv100

Grgic, M., Delac, K., & Grgic, S. (2011). SCface–surveillance cameras face database. *Multimedia Tools and Applications, 51*(3), 863–879.

Guidi, A., Gentili, C., Scilingo, E. P., & Vanello, N. (2019). Analysis of speech features and personality traits. *Biomedical Signal Processing and Control, 51*, 1–7.

Gur, R. C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., … Gur, R. E. (2002). A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. *Journal of Neuroscience Methods, 115*(2), 137–143.

Hareli, S., Shomrat, N., & Hess, U. (2009). Emotional versus neutral expressions and perceptions of social dominance and submissiveness. *Emotion, 9*(3), 378–384. https://doi.org/10.1037/a0015958

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223–233. https://doi.org/10.1016/s1364-6613(00)01482-0

Hess, U., Blairy, S., & Kleck, R. E. (2000). The influence of expression intensity, gender, and ethnicity on judgments of dominance and affiliation. *Journal of Nonverbal Behavior, 24*, 265–283.

Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology: CB, 11*(11), 880–885. https://doi.org/10.1016/s0960-9822(01)00243-3

Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition*.

Issa, D., Demirci, M. F., & Yazici, A. (2020). Speech emotion recognition with deep convolutional neural networks. *Biomedical Signal Processing and Control, 59*, 101894.

Jack, R. E., & Schyns, P. G. (2015). The human face as a dynamic tool for social communication. *Current Biology, 25*(14), R621–R634. https://doi.org/10.1016/j.cub.2015.05.052

Jaeger, B. & Jones, A. L. (2021). Which facial features are central in impression formation? *Social Psychological and Personality Science, 13*(2), 553–561.

Jobanputra, M., Chaudhary, A., Shah, S., & Gandhi, R. (2018). Real-time face recognition in hd videos: algorithms and framework. In *2018 Annual IEEE international systems conference (SysCon)* (pp. 1–8). IEEE.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Kasinski, A., Florek, A., & Schmidt, A. (2008). The PUT face database. *Image Processing and Communications, 13*(3-4), 59–64.

Kaulard, K., Cunningham, D. W., Bülthoff, H. H., & Wallraven, C. (2012). The MPI facial expression database—a validated database of emotional and conversational facial expressions. *PLoS One, 7*(3), e32321.

Kim, M. W., Choi, J. S., & Cho, Y. S. (2011). The Korea University Facial Expression Collection (KUFEC) and semantic differential ratings of emotion. *Korean Journal of Psychology Genetics, 30*, 1189–2111.

Kristal, J. (2005). *The temperament perspective: Working with children's behavior styles*. Brookes Publishing Co.

Krumhuber, E. G., Skora, L., Küster, D., & Fou, L. (2017). A review of dynamic datasets for facial expression research. *Emotion Review, 9*(3), 280–292. https://doi.org/10.1177/1754073916670022

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual (Tech. Rep. No. A-6)*. University of Florida.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. D. (2010). Presentation and validation of the Radboud faces database. *Cognition and Emotion, 24*(8), 1377–1388.

Leibold, L. J., Yarnell Bonino, A., & Buss, E. (2016). Masked speech perception thresholds in infants, children, and adults. *Ear and Hearing, 37*(3), 345–353. https://doi.org/10.1097/AUD.0000000000000270

Leppänen, J. M., Milders, M., Bell, J. S., Terriere, E., & Hietanen, J. K. (2004). Depression biases the recognition of emotionally neutral faces. *Psychiatry Research, 128*(2), 123–133. https://doi.org/10.1016/j.psychres.2004.05.020

Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual database of emotional speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLoS ONE, 13*(5), e0196391. https://doi.org/10.1371/journal.pone.0196391

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended Cohn-Kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (pp. 94–101). IEEE.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

MacNamara, A., Foti, D., & Hajcak, G. (2009). Tell me about it: neural activity elicited by emotional pictures and preceding descriptions. *Emotion, 9*(4), 531.

Marsh, A. A., Adams, R. B., & Kleck, R. E. (2005). Why do fear and anger look the way they do? Form and social function in facial expressions. *Personality and Social Psychology Bulletin, 31*(1), 73–86. https://doi.org/10.1177/0146167204271306

Martin, O., Kotsia, I., Macq, B., & Pitas, I. (2006). The eNTERFACE'05 audio-visual emotion database. In *22nd International conference on data engineering workshops (ICDEW'06)* (pp. 8–8). IEEE.

Martinez, A., & Benavente, R. (1998). The AR face database: CVC technical report, 24.

Matsumoto, D. (1983). Behavioral predictions based on perceptions of facial expressions of emotion. *Social Behavior and Personality, 11*, 97–104.

McCool, C., Marcel, S., Hadid, A., Pietikäinen, M., Matejka, P., Cernocký, J., Poh, N., Kittler, J., Larcher, A., Lévy, C., Matrouf, D., Bonastre, J.-F., Tresadern, P., & Cootes, T. (2012). Bi-modal person recognition on a mobile phone: using mobile phone data. In *2012 IEEE international conference on multimedia and expo workshops* (pp. 635–640). IEEE.

McEwan, K., Gilbert, P., Dandeneau, S., Lipka, S., Maratos, F., Paterson, K. B., & Baldwin, M. (2014). Facial expressions depicting compassionate and critical emotions: the development and validation of a new emotional face stimulus set. *PLoS One, 9*(2), e88783. https://doi.org/10.1371/journal.pone.0088783

Mondloch, C. J., Lewis, T. L., Budreau, D. R., Maurer, D., Dannemiller, J. L., Stephens, B. R., & Kleiner-Gathercoal, K. A. (1999). Face perception during early infancy. *Psychological Science, 10*(5), 419–422. https://doi.org/10.1111/1467-9280.00179

Moreno, A. B., & Sanchez, A. (2004). GavabDB: a 3D face database. Workshop on Biometrics on the Internet, pp. 77–85.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychological Science, 15*(2), 133–137. https://doi.org/10.1111/j.0963-7214.2004.01502010.x

Navas, E., Castelruiz, A., Luengo, I., Sánchez, J., & Hernáez, I. (2004). Designing and recording an audiovisual database of emotional speech in Basque. In *LREC* (pp. 1387–1390).

Ng, H. W., & Winkler, S. (2014). A data-driven approach to cleaning large face datasets. In *2014 IEEE international conference on image processing (ICIP)* (pp. 343–347). IEEE.

Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences, 13*(3), 135–143. https://doi.org/10.1016/j.tics.2008.12.006

O'Reilly, H., Pigat, D., Fridenson, S., Berggren, S., Tal, S., Golan, O., Bölte, S., Baron-Cohen, S., & Lundqvist, D. (2016). The EU-Emotion Stimulus Set: A validation study. *Behavior Research Methods, 48*(2), 567–576. https://doi.org/10.3758/s13428-015-0601-4

O'Toole, A. J., Harms, J., Snow, S. L., Hurst, D. R., Pappas, M. R., Ayyad, J. H., & Abdi, H. (2005). A video database of moving faces and people. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(5), 812–816. https://doi.org/10.1109/TPAMI.2005.90

Öhman, A. (1997). As fast as the blink of an eye: Evolutionary preparedness for preattentive processing of threat. *Attention and orienting: Sensory and motivational processes* (pp. 165–184).

Otsuka, Y. (2014). Face recognition in infants: A review of behavioral and near-infrared spectroscopic studies. *Japanese Psychological Research, 56*(1), 76–90.

Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia, 45*(1), 75–92. https://doi.org/10.1016/j.neuropsychologia.2006.04.025

Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(10), 1090–1104.

Pigeon, S., & Vandendorpe, L. (1997). The M2VTS multimodal face database (release 1.00). In *International conference on audio-and video-based biometric person authentication* (pp. 403–409). Springer.

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage, 56*(4), 2356–2363. https://doi.org/10.1016/j.neuroimage.2011.03.067

Pope, L. K., & Smith, C. A. (1994). On the distinct meanings of smiles and frowns. *Cognition and Emotion, 8*, 65–72.

Ray, G. B. (1986). Vocally cued personality prototypes: An implicit personality theory approach. *Communication Monographs, 53*(3), 266–276.

Reinl, M., & Bartels, A. (2014). Face processing regions are sensitive to distinct aspects of temporal sequence in facial dynamics. *NeuroImage, 102*(Part 2), 407–415. https://doi.org/10.1016/j.neuroimage.2014.08.011

Rosenblum, L. D., Johnson, J. A., & Saldaña, H. M. (1996). Point-light facial displays enhance comprehension of speech in noise. *Journal of Speech and Hearing Research, 39*(6), 1159–1170. https://doi.org/10.1044/jshr.3906.1159

Rosenblum, L. D., Yakel, D. A., Baseer, N., Panchal, A., Nodarse, B. C., & Niehus, R. P. (2002). Visual speech information for face recognition. *Perception & Psychophysics, 64*(2), 220–229. https://doi.org/10.3758/BF03195788

Sagha, H., Matejka, P., Gavryukova, M., Povolný, F., Marchi, E., & Schuller, B. W. (2016). Enhancing multilingual recognition of emotion in speech by language identification. In *Interspeech* (pp. 2949–2953).

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion, 9*(2), 260–264. https://doi.org/10.1037/a0014681

Scherer, K. R., & Scherer, U. (1981). Speech behavior and personality. *Speech Evaluation in Psychiatry, 1*, 460.

Schmidtmann, G., Jennings, B. J., Sandra, D. A., Pollock, J., & Gold, I. (2020). The McGill face database: Validation and insights into the recognition of facial expressions of complex mental states. *Perception.* https://doi.org/10.1177/0301006620901671

Sims, T., Hogan, C., & Carstensen, L. (2015). Selectivity as an emotion regulation strategy: Lessons from older adults. *Current Opinion in Psychology, 3*, 80–84. https://doi.org/10.1016/j.copsyc.2015.02.012

Tian, Y. L., & Bolle, R. M. (2003). Automatic detecting neutral face for face authentication and facial expression analysis. In *AAAI-03 spring symposium on intelligent multimedia knowledge management* (Vol. 3, pp. 24–26).

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66*, 519–545.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., … Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research, 168*(3), 242–249.

Tottenham, N., Hertzig, M. E., Gillespie-Lynch, K., Gilhooly, T., Millner, A. J., & Casey, B. J. (2014). Elevated amygdala response to faces and gaze aversion in autism spectrum disorder. *Social Cognitive and Affective Neuroscience, 9*(1), 106–117. https://doi.org/10.1093/scan/nst050

Trautmann, S. A., Fehr, T., & Herrmann, M. (2009). Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Research, 1284*, 100–115. https://doi.org/10.1016/j.brainres.2009.05.075

Vaiman, M., Wagner, M. A., Caicedo, E., & Pereno, G. L. (2017). Development and validation of an Argentine set of facial expressions of emotion. *Cognition and Emotion, 31*(2), 249–260.

van der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion, 11*(4), 907–920. https://doi.org/10.1037/a0023853

Wallraven, C., Breidt, M., Cunningham, D. W., & Bülthoff, H. H. (2008). Evaluating the perceptual realism of animated facial expressions. *ACM Transactions on Applied Perception (TAP), 4*(4), 1–20.

Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology, 78*(1), 105–119. https://doi.org/10.1037//0022-3514.78.1.105

Wu, B. F., & Lin, C. H. (2018). Adaptive feature mapping for customizing deep learning based facial expression recognition model. *IEEE Access, 6*, 12451–12461.

Wyczesany, M., Ligeza, T. S., Tymorek, A., & Adamczyk, A. (2018). The influence of mood on visual perception of neutral material. *Acta Neurobiologiae Experimentalis, 78*(2), 163–172.

Yang, T., Yang, Z., Xu, G., Gao, D., Zhang, Z., Wang, H., Liu, S., Han, L., Zhu, Z., Tian, Y., Huang, Y., Zhao, L., Zhong, K., Shi, B., Li, J., Fu, S., Liang, P., Banissy, M. J., & Sun, P. (2020). Tsinghua facial expression database - A database of facial expressions in Chinese young and older women and men: Development and validation. *PLoS One, 15*(4), e0231304. https://doi.org/10.1371/journal.pone.0231304

Yin, L., Chen, X., Sun, Y., Worm, T., & Reale, M. (2008). A high-resolution 3D dynamic facial expression database. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition* (pp. 1–6).

Yoon, K. L., & Zinbarg, R. E. (2007). Threat is in the eye of the beholder: Social anxiety and the interpretation of ambiguous facial expressions. *Behaviour Research and Therapy, 45*(4), 839–847. https://doi.org/10.1016/j.brat.2006.05.004

Zebrowitz, L. (1997). *Reading faces: Window to the soul?* Westview Press.

Zebrowitz, L. A., Kikuchi, M., & Fellous, J. M. (2010). Facial resemblance to emotions: Group Differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology, 98*(2), 175–189. https://doi.org/10.1037/a0017990

Zhalehpour, S., Onder, O., Akhtar, Z., & Erdem, C. E. (2017). BAUM-1: A spontaneous audio-visual face database of affective and mental states. *IEEE Transactions on Affective Computing, 8*(3), 300–313.