

Running head: DYNAMIC DATASETS

A Review of Dynamic Datasets for Facial Expression Research

Eva G. Krumhuber and Lina Skora

University College London

Dennis Küster

Jacobs University Bremen

Linyun Fou

University College London

Linyun Fou is now a research assistant at the Royal College of Obstetricians and Gynaecologists, London, United Kingdom.

Correspondence concerning this article should be addressed to Eva Krumhuber, Department of Experimental Psychology, University College London, 26 Bedford Way, WC1H 0AP London, United Kingdom. E-mail: e.krumhuber@ucl.ac.uk

Word count: 8413

Abstract

Temporal dynamics have been increasingly recognized as an important component of facial expressions. With the need for appropriate stimuli in research and application, a range of databases of dynamic facial stimuli has been developed. The present article reviews the existing corpora and describes the key dimensions and properties of the available sets. This includes a discussion of conceptual features in terms of thematic issues in dataset construction as well as practical features which are of applied interest to stimulus usage. To identify the most influential sets, we further examine their citation rates and usage frequencies in existing studies. General limitations and implications for emotion research are noted and future directions for stimulus generation are outlined.

KEYWORDS: facial expression, emotion, dynamic, dataset

A Review of Dynamic Datasets for Facial Expression Research

Existing research points towards the benefits of facial motion in emotion perception and recognition. By providing unique information about the direction, quality and speed of motion, dynamic stimuli enhance coherence in the identification of affect, lead to stronger emotion judgments, and facilitate the differentiation between posed and spontaneous expressions (for a review see Krumhuber, Kappas, & Manstead, 2013). In the last two decades, this advantage - paired with the stimuli's greater realism and ecological validity - has led to increased questioning and criticism regarding the use of static images (e.g., Tcherkassof, Bollon, Dubois, Pansu, & Adam, 2007; Wehrle, Kaiser, Schmidt, & Scherer, 2000), with a gradual shift in interest towards dynamic expressions.

The trend is reflected in the literature with exponential increases of relevant entries over the past thirty-five years. For example, a Google Scholar search for the word "dynamic face" and related phrasesⁱ returned a mere 13 articles in 1980-1989 and 87 articles in 1990-1999. This figure rose to 889 results in 2000-2009 and has more than doubled to 2,184 results in the past five years, from 2010 to 2015. In order to meet new demands in research on both human communication and progressively in machine recognition or human-computer interaction, several databases of dynamic facial stimuli have been developed.

This paper aims to provide a systematic review of the existing corpora and draw out the key dimensions and properties of the available dynamic sets. It should be noted that this review is not exhaustive with respect to the stimuli developed within the field of computer science (for an extensive overview of such see Cowie, Douglas-Cowie, & Cox, 2005; Sandbach, Zafeiriou, Pantic, & Yin, 2012; Zeng, Pantic, Roisman, & Huang, 2009). In order to account for the diversity of dynamic facial expression databases, the following selection criteria were applied for inclusion in the present review: (a) public accessibility of the database, (b) database paper accessible and published between 2000 and 2015, (c) a minimum of five emotions, (d) digital

format of recordings, (e) visual or audio-visual modality of stimuli, (f) real human encoders, and (g) individual portrayals (as opposed to emotive interactions; note that some might contain both types).

In an attempt to provide useful guidance for the readers of this paper, we classified databases in terms of three fundamental issues that are relevant to decisions about stimulus sets. These include a) *conceptual features*, which reflect thematic approaches in database construction and validation (Table 1), b) *practical features*, which concern applied aspects related to stimulus usage (Table 2), and c) *citation and usage frequencies* of dynamic datasets in the literature (Table 3ⁱⁱ), thereby elucidating their respective impact in the field. This latter issue can be categorized according to whether a dataset was used as stimulus material in research with human participants (social sciences) or for the training and testing of machine learning algorithms (computer sciences). With the tables designed to give specific information about each dataset, the accompanying text will focus on a general discussion, which is structured in terms of the key points listed in Table 1 and is intended to address both theoretical and technical issues, as well as possible directions for future stimulus development.

Emotion Content and Diversity

When choosing an appropriate database, selection criteria should be guided by the specific study question of the researcher (Wagner, 1997). Typically, these tap into two main areas: the expression of facial expressions (encoding) or their perception (decoding). While encoding studies target the expressive features associated with an underlying emotional state, decoding studies investigate how those features are perceived and interpreted by observers (Ekman, Friesen, & Ellsworth, 1982). Depending on the question pursued, available sets of dynamic facial expressions may differ in their suitability.

Table 1 lists key conceptual points which shed light on the scope and potential application of each dataset. A brief review of the number and types of emotion concepts demonstrates that many databases adopt a categorical approach. The categorical view suggests a division of emotions into basic, mutually exclusive categories, such that each belongs to one category, with more complex, compound emotions accounted for by a blending between basic ones (Ekman, 1994; Ekman & Cordaro, 2011). Mostly, these categories are the six basic emotions: anger, disgust, fear, happiness, sadness and surprise (and occasionally, contempt) (Ekman, 1992). Databases that are strictly categorically oriented, featuring between five and eight basic emotion concepts (i.e., BU-4DFE, DISFA, FG-NET, STOIC) are suitable for decoding studies. By allowing for the examination of the expressive cues used in perception, emotion attribution processes can be investigated using these sets. However, a greater variety of stimuli is needed for encoding studies to accurately represent the range of emotional states expressed in everyday life (Calvo & D’Mello, 2010; Zeng et al., 2009).

To account for this complexity, the hierarchical approach may be particularly valuable (Shaver, Schwartz, Kirson, & O’Connor, 1987). Whilst often retaining most (if not all) of the basic labels, databases arising from this framework differentiate to capture non-basic emotions, with their numbers varying between 11 (UT Dallas) and 55 (MPI). Of note is the CAM Face-Voice Battery which contains a hierarchical organisation of 412 emotion concepts in 24 overarching groups. Databases with subordinate differentiation within or in place of some of the basic emotion concepts (i.e., BNED, DynEmo, EU-Emotion, GEMEP) serve particularly well for representing different degrees of arousal, which would go unnoticed if generic labels alone were used (Russell, 1980). This approach increases the diversity within emotion types by offering subordinate exemplars of varying intensities (i.e., nervous, anxious and frightened under fear, or amusement, joy and excitement under happiness).

Databases that span a large range of emotion categories are also well suited for human-computer interaction research (Pantic & Bartlett, 2007). Increasing efforts are targeted towards computer systems that are able to recognize and respond to emotional signals. Such systems have an enormous potential for affective computing in terms of automatic human affect analysis, which can be applied in fields as diverse as security, medicine, education and telecommunication (Picard, 1997). This rising interest is also reflected in the citation and usage frequency of the available dynamic sets. As shown in Table 3, the most cited and frequently used databases are CK, CK+, FG-NET, and MMI, all of which were created by computer scientists. These databases typically tap into specific and applied research themes in the computer sciences (i.e., comparing and improving the recognition or detection accuracy of machine learning algorithms), whilst the datasets in the social sciences are generally used in a more diverse way. For affect-based recognition systems to process complex facial signals representative of numerous emotions, wide coverage of emotional phenomena including non-basic affective states may therefore be fruitful (Sandbach et al., 2012).

Attempts to extend the range of emotions represented in the databases would likewise pave the way for larger stimulus numbers in a practical sense. Databases sometimes portray fairly comprehensive sets of different types of expressions (see Tables 1 & 2) but only for a relatively small number of trained encoders (e.g., D3D-FACS, GEMEP), whereas others provide fewer videos per encoder but use a larger subject pool (e.g., DIFSA, DynEmo). A few sets (i.e., UT Dallas, BINED) include many videos for a medium number of encoders, but these databases typically do not provide behavioral coding for all stimuli which is disadvantageous in terms of facial action classification (e.g., Facial Action Coding System (FACS), Ekman, Friesen, & Hager, 2002). Given these trade-offs, the MMI database - with well over 800 FACS-coded stimuli and 78 encoders - is perhaps the closest to providing a large number of diverse and behaviorally-coded stimuli.

Elicitation Type and Control

A major issue for the selection of a stimulus set concerns the type of expression it contains. As can be seen in Table 1, the available dynamic databases tend to widely use deliberate expressions, with a majority employing a variant of posing over spontaneous emotion elicitation. Posed expressions can emerge from instructions to perform an expression/ facial actions (i.e., ADFES, CK, MMI, STOIC) or the enactment of emotional scenarios using the Stanislavski or other method acting techniques (i.e., DaFEx, EU-Emotion, GEMEP, MPI). Such datasets typically allow for good experimental control and yield standardized and prototypical displays that are similar across encoders (Scherer & Bänzinger, 2010). In addition to eliminating confounds prevalent in everyday emotion communication (e.g. display rules or emotion regulation strategies), posed expressions are often the preferred method of choice in decoding studies. Facial behavior of this type is more intense and unambiguous due to the clear intention to convey the desired emotion (Cohn, Ambadar, & Ekman, 2007). This can enhance recognition accuracy (Hess, Blairy, & Kleck, 1997) in studies that aim to test observers' judgments against a predefined label assigned to the expression (Sneddon, McRorie, McKeown, & Hanratty, 2007).

However, this advantage of comparability and reliability can be a disadvantage in terms of realism. Given that everyday emotional expressivity is relatively subtle and heterogeneous (Motley & Camden, 1988), posed expressions may have lower ecological validity, failing to occur in natural or pseudo-natural (e.g., films) emotion episodes (Cowie, 2009; Cowie et al., 2005; Scherer & Ellgring, 2007a). Indeed, evidence suggests that spontaneous expressions differ in appearance and timing from posed ones (Ekman & Rosenberg, 2005). Such differences are also reflected in the stimulus durations of the reviewed dynamic databases (see Table 2). Whilst posed sets (i.e., CK, DaFEx, MMI, STOIC) feature expressions of short (500ms) to medium length (180s), stimuli composed of spontaneous expressions (i.e., DISFA, DynEmo,

HUMAINE) can last up to several minutes. Approaches based on deliberate and often exaggerated portrayals may, therefore, potentially fail to generalize to real-world behavior (Zeng et al., 2009).

To study emotions that approximate more natural instances, spontaneous databases provide a valuable source of information, especially for encoding studies (Scherer & Bänzinger, 2010). Respective expressions are captured inconspicuously in either the lab or field (i.e., BNED, HUMAINE) or via emotion-specific eliciting techniques such as the presentation of emotionally-laden pictures/films (i.e., BINED, BP-4D, DISFA, DynEmo, UT Dallas; see Gross & Levenson, 1995). Besides allowing for more fine-grained and natural forms of expression, spontaneous displays can include context-specific information about the emotion-eliciting event. This makes them challenging to analyze as they are often blended rather than pure emotions, with significant variability in expression across encoders (Bänzinger & Scherer, 2007). Also, video backgrounds may vary (see Table 2), some having wavy curtains (BNED, HUMAINE) or naturalistic office-type environments that show additional objects such as cables and microphone holders (BINED, FG-NET).

For authentic emotion induction to become the method of first choice, researchers will likely need to aim for a compromise between spontaneity and experimental control (Sneddon et al., 2007; Zhang et al., 2014). At the moment, recording conditions are often not well technically controlled which affects the quality of the stimuli (see also Bänzinger et al., 2012). As a result, naturally-oriented databases lag behind in providing top-notch, technically sound, materials. From the available sets that include spontaneous expressions, best-buy recommendations are probably BP-4D, DynEmo, and UT Dallas, all of which (partially) standardize background and lighting and are of acceptable nominal resolution.

In the future, more work could be done to capture facial expressions at higher frames rates (60 fps and higher) using specialized recording equipment. A distinction could also be

made between what is visible to the encoder and to the camera/perceiver. For example, dataset authors might want to set up a comfortable and natural environment for the encoder (allowing for spontaneous behavior), while at the same time ensuring that what the camera captures is systematically controlled. In addition to existing and well-validated techniques for emotion induction (for an overview see Coan & Allen, 2007), novel social entities such as virtual agents, robots, and androids may constitute a viable option for eliciting spontaneous expressions. Since their appearance and behavior is fully controllable, human users' response patterns can be evoked and recorded in a consistent manner (MacDorman & Ishiguro, 2006).

Measurement and Validation

To validate the emotional content of expressions, judgment tasks (also referred to as recognition tests) serve as the primary validation measure in the context of the reviewed posed datasets (Scherer & Bänzinger, 2010). With the aim of assessing the accuracy of the conveyed relevant emotions, observers were asked to provide an emotion label that matched the viewed stimulus. Most often this occurred out of a closed set of categorical options (from 7 to 24). In some databases (i.e., BNED, BP-4D, DynEmo, HUMAINE) inter-rater agreement on emotion categories or segments is used as an extra measure of reliability, thus assessing recognition from a second perspective (and accounting for chance agreement if measured by the kappa statistic; Sayette, Cohn, Wertz, Perrott & Parrott, 2001). Although the forced-choice paradigm yields robust results, particularly in the case of basic emotions (Limbrecht-Ecklundt et al., 2013), it has been criticized for lacking ecological validity since it forces the use of labels that might not otherwise be selected (Russell, 1993; Wagner, 1997).

In order to allow for a more flexible selection of emotion terms, without restraining the observer to one response option, alternative methods include confidence and intensity judgments applied to all emotion labels (e.g., Hi4D-ADSIP, STOIC) or continuous emotion

ratings as expressions progress over time (e.g., DynEmo). A few databases use additional supportive measures that tap into the dimensions of valence, arousal and/or intensity (i.e., ADFES, BINED, BNED, EU-Emotion, GEMEP). These provide added value as they offer a more comprehensive framework for emotion assessment than mere categories or hierarchies (Russell, 1980) and can also increase the informative value of a given emotional episode.

For spontaneous datasets, introspective measures constitute an essential validation approach (e.g., BINED, BP-4D, DynEmo). Encoder self-reports of the emotion felt during the elicitation procedure provide insight into the elicitation effectiveness and accuracy of the resulting expression (Gray & Watson, 2007). This enables an evaluation of whether the target emotion was elicited. Nevertheless, reliance on self-report alone remains problematic due to potential discrepancies between what is experienced and what is reported (Nielsen & Kaszniak, 2007). In this context, additional information in the form of audiovisual cues (i.e., gesture, posture, speech) could be particularly useful to yield a coherent representation of the emotion in question (Cowie et al., 2005; Scherer & Ellgring, 2007b). Multi-modal stimuli have long been acknowledged to improve emotion classification (Russell, Bachorowski, & Fernandez-Dols, 2003; Van den Stock, Righart, & de Gelder, 2007). Encoding studies may therefore substantially benefit from the presence of multi-modal affective features in databases that allow examination across modalities (i.e., BINED, BNED, DynEmo).

Component measures such as the Facial Action Coding System (FACS) can be of considerable value in this regard by providing an objective classification of the observed behavior (Ekman & Friesen, 1982). Such measures permit a comparison between expressive features and emotion related variables (i.e., self-reports, physiological responses) in the encoder. In the reviewed datasets, FACS coding is available for both deliberate and spontaneous expressions for the dimensions of Action Unit (AU) occurrence, intensity, and/or

timing. The BP-4D set examines its stimuli using multiple methods (i.e., emotion self-reports, observer judgments, and FACS), thereby providing the most stringent validation of its content.

Some databases also submit their stimuli to machine recognition (i.e., DISFA, BP-4D, CK, D3D-FACS). FACS has been frequently used in studies of automatic expression classification, making it a prominent tool in affective computing (Lien, Kanade, Cohn, & Li, 2000; Cohn, Zlochower, Lien, & Kanade, 1999). Automatic AU recognition has been shown to achieve recognition rates comparable in accuracy to manual coding, indicating its potential to significantly facilitate the labor-intensive process (Cohn, Zlochower, Lien, & Kanade, 1999). However, most systems employing FACS for facial behavior measurement still have been using posed expressions to train the classifiers in recognition, thereby restricting their applicability in natural settings (Zeng et al., 2009).

To develop automatic systems that are robust to natural variations in appearance, behavior and context, future research should invest in more stimulus sets containing spontaneous expressions (see BP-4D, DISFA; Bartlett et al., 2006; Pantic, 2009). Such an endeavour would also be advantageous for the (automatic) analysis of the temporal dynamics of spontaneous expressions. Whilst there are a few such attempts (e.g., Cohn & Schmidt, 2004; Valstar, Pantic, Ambadar, & Cohn, 2006), the field is still in its infancy with respect to the extraction and modelling of the temporal structure of spontaneous facial actions, including their temporal relations. To fulfil this requirement, high frame rates and good resolution are necessary pre-conditions (see Sandbach et al., 2012). Whilst the nominal resolution has increased substantially for some of the most recent sets (i.e., BP-4D, BU-4DFE, D3D-FACS, Hi4D-ADSIP), the effectively-available visible area of the face in the video (i.e., Face-box, see Table 2) is still less than 300 square pixels for the majority of databases. Such a resolution could prove insufficient for exploring micro-expressions or subtle temporal features that require small parts of the face to be clearly visible.

In the future, cooperative efforts between psychology and computer science to work on a common dataset are indispensable (for a positive example see the ‘Facial Expression Recognition and Analysis’ (FERA) challenge, Valstar et al., 2015). At the moment, only a small number of dynamic stimulus sets tend to be commonly cited and employed (i.e., CK, CK+, FG-NET, MMI, GEMEP; see Table 3). When comparing dataset usage between disciplines over the past 15 years, the number of empirical papers in the computer sciences ($n = 1543$) vastly outnumbers those in the social sciences ($n = 124$). It therefore appears as if dataset usage in the social sciences is more restricted, with an almost exclusive focus on posed expressions. For knowledge transfer and dialogue to increase, researchers from both sides will have to embrace the wide variety of available stimulus sets. We hope that the present review helps to enable more work on the dynamic nature of emotions.

References

References with numbers in superscripts denote papers which describe a dynamic facial expression dataset as listed in Table 1.

- ¹⁵Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva Multimodal Expression corpus for experimental research on emotion perception. *Emotion, 12*, 1161-1179. doi:10.1037/a0025827
- ¹⁵Bänziger, T., Pirker, H., & Scherer, K. R. (2006). GEMEP – GENEVA Multimodal Emotion Portrayals: A corpus for the study of multimodal emotional expressions. In *Proceedings of the Fifth Conference on Language Resources and Evaluation* (Vol. 6, pp. 15-19). Genoa, Italy: ELRA.
- ¹⁵Bänziger, T., & Scherer, K. R. (2007). Using actor portrayals to systematically study multimodal emotion expression: The GEMEP corpus. In A. C. R. Paiva, R. Prada, & R. W. Picard (Eds.), *Lecture Notes in Computer Science: Vol 4738. ACI 2007 – Affective Computing and Intelligent Interaction, Second International Conference* (pp. 476-487). Berlin, Germany: Springer. doi: 10.1007/978-3-540-74889-2_42
- ¹⁵Bänziger, T., & Scherer, K. R. (2010). Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) corpus. In K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 271-294). Oxford, England: Oxford University Press.
- ⁶Baron-Cohen, S., Golan, O., Wheelwright, S., & Hill, J. J. (2004). *Mind Reading: The interactive guide to emotions*. London, England: Jessica Kingsley Limited.
- Bartlett, M. S., Littlewort, G. C., Frank, M. G., Lainscsek, C., Fasel, I. R., & Movellan, J. R. (2006). Fully automatic facial action recognition in spontaneous behavior. In *Proceedings of the Seventh IEEE International Conference on Automatic Face and*

Gesture Recognition (pp. 223-230). Los Alamitos, CA: IEEE Computer Society.

doi:0.1109/fgr.2006.55

¹⁰Battocchi, A., Pianesi, F., & Goren-Bar, D. (2005a). A first evaluation study of a Database of Kinetic Facial Expressions (DaFEx). In G. Lazzari, F. Pianesi, J. L. Crowley, K. Masey, & S. L. Oviatt (Eds.), *Proceedings of the Seventh International Conference on Multimodal Interfaces* (pp. 214-221). New York, NY: ACM Press.

doi:10.1145/1088463.1088501

¹⁰Battocchi, A., Pianesi, F., & Goren-Bar, D. (2005b). The properties of DaFEx, a Database of Kinetic Facial Expressions. In J. Tao, T. Tan, & R.W. Picard (Eds.), *Lecture Notes in Computer Science: Vol 3784. ACII 2005 – Affective Computing and Intelligent Interaction, First International Conference* (pp. 558-565). Berlin, Germany: Springer.

doi: 10.1007/11573548_72

¹⁰Battocchi, A., Pianesi, F., & Goren-Bar, D. (2005c). DaFEx: Database of Facial Expressions. In M. Maybury, O. Stock, & W. Wahlster (Eds.), *Lecture Notes in Computer Science: Vol 3814. INTETAIN 2005 – Intelligent Technologies for Interactive Entertainment, First International Conference* (pp. 303-306). Berlin,

Germany: Springer. doi:10.1007/11590323_39

Bradski, G. (2000). The OpenCV library. *Dr. Dobb's Journal of Software Tools*, 25, 120-126.

Calvo, R. A., & D'Mello, S. (2010). Affect detection: an interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1, 18-37.

doi:10.1109/t-afc.2010.1

Coan, J. A., & Allen, J. J. B. (2007). *Handbook of emotion elicitation and assessment*. New York, NY: Oxford University Press.

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. In J. A. Coan & J. J. B. Allen (Eds.),

Handbook of emotion elicitation and assessment (pp. 203-221). New York, NY: Oxford University Press.

Cohn, J. F., & Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, 2, 1-12. doi:10.1142/9789812704313_0005

Cohn, J. F., Zlochower, A. J., Lien, J. J., & Kanade, T. (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology*, 36, 35-43. doi:10.1017/s0048577299971184

⁹Cosker, D., Krumhuber, E., & Hilton, A. (2011). A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling. In D. Metaxas, L. Quan, A. Sanfeliu, & L. Van Gool (Eds.), *Proceedings of the 13th IEEE International Conference on Computer Vision (ICCV)* (pp. 2296–2303). Barcelona, Spain: IEEE. doi: 10.1109/iccv.2011.6126510

Cowie, R. (2009). Perceiving emotion: towards a realistic understanding of the task. *Philosophical Transactions: Biological Sciences*, 364, 3515-3525. doi:10.1098/rstb.2009.0139

Cowie, R., Douglas-Cowie, E., & Cox, C. (2005). Beyond emotion archetypes: Databases for emotion modelling using neural networks. *Neural Networks*, 18, 371-388. doi:10.1016/j.neunet.2005.03.002

¹⁹Cunningham, D. W., Kleiner, M., Wallraven, C., Bülthoff, H. H. (2005). Manipulating video sequences to determine the components of conversational facial expressions. *ACM Transactions on Applied Perception*, 2, 251-269. doi:10.1145/1077399.1077404

³Douglas-Cowie, E., Campbell, N., Cowie, R., & Roach, P. (2003). Emotional speech: Towards a new generation of databases. *Speech Communication*, 40, 33-60. doi:10.1016/s0167-6393(02)00070-5

- ³Douglas-Cowie, E., Cowie, R., & Schröder, M. (2000). A new emotion database: Considerations, sources and scope. In *Proceedings of the ISCA Workshop on Speech and Emotion* (pp. 39-44). Newcastle, Northern Ireland.
- ³Douglas-Cowie, E., Cowie, R., & Schröder, M. (2003). The description of naturally occurring emotional speech. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2877-2880). Barcelona, Spain.
- ¹⁷Douglas-Cowie, E., Cowie, R., Sneddon, I., Cox, C., Lowry, O., McRorie, M., ... , Karpouzis, K. (2007). The HUMAINE database: Addressing the collection and annotation of naturalistic and induced emotional data. In A. Paiva, R. Prada, & R. W. Picard (Eds.), *Proceedings of the Second International Conference on Affective Computing and Intelligent Interaction, ACII 2007* (pp. 488-500). Lisbon, Portugal: Springer. doi:10.1007/978-3-540-74889-2_43
- ¹⁷Douglas-Cowie, E., Cox, C., Martin, J. C., Devillers, L., Cowie, R., Sneddon, I., ... , Hönic, F. (2011). The HUMAINE database. In P. Petta, C. Pelachaud, & R. Cowie (Eds.), *Emotion-oriented systems. The HUMAINE handbook* (pp. 243-284). Berlin, Germany: Springer.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 169-200.
doi:10.1080/02699939208411068
- Ekman, P. (1994). All emotions are basic. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 15-19). New York: Oxford University Press.
- Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, 3, 364-370. doi:10.1177/1754073911410740

- Ekman, P., & Friesen, W. V. (1982). Measuring facial movement with the Facial Action Coding System. In P. Ekman (Ed.), *Emotion in the human face* (2nd ed., pp. 178-211). Cambridge, UK: Cambridge University Press.
- Ekman, P., Friesen, W. V., & Ellsworth, P. (1982). Methodological decisions. In P. Ekman (Ed.), *Emotion in the human face* (2nd ed., pp. 22-38). Cambridge, UK: Cambridge University Press.
- Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System: The manual on CD ROM*. Salt Lake City, UT: Research Nexus.
- Ekman, P., & Rosenberg, E. L. (2005). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System*. Oxford, UK: Oxford. doi:10.1093/acprof:oso/9780195179644.001.0001
- ⁶Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge Mindreading (CAM) face-voice battery: Testing complex emotion recognition in adults with and without Asperger Syndrome. *Journal of Autism and Developmental Disorders*, 36, 169 – 183. doi:10.1007/s10803-005-0057-y
- Gray, E. K., & Watson, D. (2007). Assessing positive and negative affect via self-report. In J. A. Coan, & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 171-183). New York, NY: Oxford University Press.
- Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion*, 9, 87-108. doi:10.1080/02699939508408966
- Hess, U., Blairy, S., & Kleck, R. E. (1997). The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior*, 21, 241-257. doi:10.1023/A:1024952730333
- ⁷Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Proceedings of the Fourth IEEE International Conference on Automatic*

Face and Gesture Recognition (pp. 46-53). Los Alamitos, CA: IEEE Computer Society.

doi:10.1109/afgr.2000.840611

¹⁹Kaulard, K., Cunningham, D. W., Bühlhoff, H. H., & Wallraven, C. (2012). The MPI facial expression database – a validated database of emotional and conversational facial expressions. *PLoS ONE*, 7(3), e32321. doi:10.1371/journal.pone.0032321

¹⁹Kaulard, K., Wallraven, C., Cunningham, D. W., & Bühlhoff, H. H. (2010). Laying the foundations for an in-depth investigation of the whole space of facial expressions [Abstract]. *Journal of Vision*, 10, 606. doi:10.1167/10.7.606

²⁰Kleiner, M., Wallraven, C., Breidt, M., Cunningham, D. W., & Bühlhoff, H. H. (2004).

Multi-viewpoint video capture for facial perception research. In N. Magnenat-Thalmann & D. Thalmann (Chairs), *Workshop on Modelling and Motion Capture Techniques for Virtual Environments* (CAPTECH 2004). Geneva, Switzerland.

Retrieved from

http://www.kyb.tuebingen.mpg.de/fileadmin/user_upload/files/publications/pdf3058.pdf

²⁰Kleiner, M., Wallraven, C., Bühlhoff, H. H. (2004). *The MPI VideoLab – A system for high quality synchronous recording of video and audio from multiple viewpoints* (Technical Report No. 123). Tübingen, Germany: Max Planck Institute for Biological Cybernetics.

Retrieved from

http://www.kyb.tue.mpg.de/fileadmin/user_upload/files/publications/pdfs/pdf2774.pdf

Krumhuber, E., Kappas, A., & Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review*, 5, 41-46.

doi:10.1177/1754073912451349

- Lien, J. J., Kanade, T., Cohn, J. F., & Li, C. (2000). Detection, tracking, and classification of action units in facial expression. *Robotics and Autonomous Systems*, *31*, 131-146. doi: 10.1016/s0921-8890(99)00103-7
- Limbrecht-Ecklundt, K., Scheck, A., Jerg-Bretzke, L., Walter, S., Hoffmann, H., & Traue, H. C. (2013). The effect of forced choice on facial emotion recognition: A comparison to open verbal classification of emotion labels. *GMS Psycho-Social-Medicine*, *10*, Doc04. doi:10.3205/psm000094
- ⁸Lucey, 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 *Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010)* (pp. 94-101). San Francisco, USA: SSPNET. doi:10.1109/cvprw.2010.5543262
- MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, *7*, 297-337. doi:10.1075/is.7.3.03mac
- ¹⁶Matuszewski, B. J., Quan, W., & Schar, L.K. (2011). High-resolution comprehensive 3-D dynamic database for facial articulation analysis. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops 2011)* (pp. 2128-2135). Barcelona, Spain: IEEE. doi:10.1109/iccvw.2011.6130511
- ¹⁶Matuszewski, B. J., Quan, W., Shark, L. K., McLoughlin, A. S., Lightbody, C. E., Emsley, H. C. A., & Watkins, C. L. (2012). Hi4D-ADSIP 3-D dynamic facial articulation database. *Image and Vision Computing*, *30*, 713 – 727. doi:10.1016/j.imavis.2012.02.002
- ¹¹Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P. (2012). Automatic detection of non-posed facial action units. In *Proceedings of the 19th IEEE International Conference on*

Image Processing (ICIP 2012) (pp. 1817-1820). Lake Buena Vista, FL: IEEE.

doi:10.1109/icip.2012.6467235

¹¹Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., & Cohn, J. F. (2013). DISFA: A spontaneous facial action intensity database. *IEEE Transactions on Affective Computing*, 4, 151-160. doi:10.1109/t-affc.2013.4

¹²Meillon, B., Tcherkassof, A., Mandran, N., Adam, J. M., Dubois, M., Dupré, D., ... , Caplier, A. (2010). DynEmo: A corpus of dynamic and spontaneous emotional facial expressions. In M. Kipp, J. C. Martin, P. Paggio, & D. Heylen (Eds.), *Proceedings of International Workshop Series on Multimodal Corpora, Tools and Resources. Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality* (pp. 31-36). Valetta, Malta: ELREC.

Motley, M., & Camden, C. (1988). Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Speech Communication*, 52, 1–22.

doi:10.1080/10570318809389622

Nielsen, L., & Kaszniak, A. W. (2007). Conceptual, theoretical and methodological issues in inferring subjective emotion experience. In J. A. Coan, & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 361-375). New York, NY: Oxford University Press.

¹³O'Reilly, H., Pigat, D., Fridenson, S., Berggren, S., Tal, S., Golan, O., Bölte, S., Baron-Cohen, S., Lundqvist, D. (in press). The EU-Emotion Stimulus Set: A validation study. *Behavior Research Methods*. Online publication. doi:10.3758/s13428-015-0601-4

²²O'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, 812 – 816. doi:10.1109/tpami.2005.90

- Pantic, M. (2009). Machine analysis of facial behaviour: Naturalistic and dynamic behaviour. *Philosophical Transactions of the Royal Society B*, 364, 3505-3513.
doi:10.1098/rstb.2009.0135
- Pantic, M., & Bartlett, M. S. (2007). Machine analysis of facial expressions. In K. Delac & M. Grgic (Eds.), *Face recognition* (pp. 377-416). Vienna, Austria: I-Tech Education and Publishing. doi:10.1098/rstb.2009.0135
- ¹⁸Pantic, M., Valstar, M., Rademaker, R., & Maat, L. (2005). Web-based database for facial expression analysis. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '05)* (pp. 317-321). Piscataway, NJ: IEEE.
doi:10.1109/icme.2005.1521424
- Picard, R. W. (1997). *Affective Computing*. Cambridge, USA: MIT Press.
- ²¹Roy, S., Roy, C., Éthier-Majcher, C., Belin, P., & Gosselin, F. (2007). *STOIC: A database of dynamic and static faces expressing highly recognizable emotions*. Montréal, Canada: Université De Montréal. Retrieved from:
https://www.researchgate.net/profile/Frederic_Gosselin2/publication/242092567_STOIC_A_database_of_dynamic_and_static_faces_expressing_highly_recognizable_emotions/links/552574530cf295bf160ea80b.pdf
- ²¹Roy, S., Roy, C., Fortin, I., Éthier-Majcher, C., Belin, P., & Gosselin, F. (2007). A dynamic facial expression database [Abstract]. *Journal of Vision*, 7, 944. doi:10.1167/7.9.944
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161-1178. doi:10.1037/h0077714
- Russell, J. A. (1993). Forced-choice response format in the study of facial expression. *Motivation and Emotion*, 17, 41-51. doi:10.1007/bf00995206

- Russell, J. A., Bachorowski, J. A., & Fernández-Dols, J. N. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54*, 329-349.
doi:10.1146/annurev.psych.54.101601.145102
- Russell, J. A., & Feldman Barrett, L. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology, 76*, 805-819. doi:10.1037/0022-3514.76.5.805
- Sandbach, G., Zafeiriou, S., Pantic, M., & Yin, J. (2012). Static and dynamic 3D facial expression recognition: A comprehensive survey. *Image and Vision Computing, 30*, 683-697. doi:10.1016/j.imavis.2012.06.005
- Sayette, M. A., Cohn, J. F., Wertz, J. M., Perrott, M. A., & Parrott, D. J. (2001). A psychometric evaluation of the Facial Action Coding System for assessing spontaneous expression. *Journal of Nonverbal Behavior, 25*, 167-185.
doi:10.1023/A:1010671109788
- Scherer, K. R., & Bänziger, T. (2010). On the use of actor portrayals in research on emotional expression. In K. R. Scherer, T. Bänziger, & E. Roesch (Eds.), *A blueprint for affective computing: A sourcebook* (pp. 166-178). Oxford, England: Oxford University Press.
- Scherer, K. R., & Ellgring, H. (2007a). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion, 7*, 113-130.
doi:10.1037/1528-3542.7.1.113
- Scherer, K. R., & Ellgring, H. (2007b). Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion, 7*, 158-171. doi:10.1037/1528-3542.7.1.158
- Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52*, 1061-1086. doi:10.1037/0022-3514.52.6.1061

- ²Sneddon, I., McRorie, M., McKeown, G., & Hanratty, J. (2007). The Belfast Induced Natural Emotion Database. *IEEE Transactions on Affective Computing*, 3, 32-41. doi:10.1109/t-affc.2011.26
- Tcherkassof, A., Bollon, T., Dubois, M., Pansu, P., & Adam, J. M. (2007). Facial expressions of emotions: A methodological contribution to the study of spontaneous and dynamic emotional faces. *European Journal of Social Psychology*, 37, 1325–1345. doi:10.1002/ejsp.427
- ¹²Tcherkassof, A., Dupré, D., Meillon, B., Mandran, N., Dubois, M., & Adam, J. M. (2013). DynEmo: A video database of natural facial expressions of emotions. *The International Journal of Multimedia and Its Applications*, 5(5), 61 – 80. doi:10.5121/ijma.2013.5505
- Valstar, M. F., Almaev, T., Girard, J. M., McKeown, G., Mehu, M., Yin, L., Pantic, M., & Cohn, J. F. (2015). FERA 2015 – Second facial expression recognition and analysis challenge. In *Proceedings of the Eleventh IEEE International Conference on Face and Gesture Recognition*. Ljubljana, Slovenia. doi:10.1109/fg.2015.7284874
- ¹⁸Valstar, M., & Pantic, M. (2010). Induced disgust, happiness and surprise: An addition to the MMI facial expression database. In *Proceedings of the International Conference on Language Resources and Evaluation, Workshop on EMOTION* (pp. 65-70). Valetta, Malta: ELRA.
- Valstar, M. F., Pantic, M., Ambadar, Z., & Cohn, J. F. (2006). Spontaneous vs. posed behavior: Automatic analysis of brow actions. In *Proceedings of the Eight International Conference on Multimedia Interfaces* (pp. 162-170). New York, USA: ACM Press. doi:10.1145/1180995.1181031
- Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion*, 7, 487-494. doi:10.1037/1528-3542.7.3.487

- ¹Van 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*, 907 – 920. doi:10.1037/a0023853
- Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 511-518). Los Alamitos, CA: IEEE Computer Society. doi:10.1109/cvpr.2001.990517
- Wagner, H. L. (1997). Methods for the study of facial behavior. In J. A. Russell, & J. M. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 31-54). Cambridge, England: Cambridge University Press. doi:10.1017/cbo9780511659911.004
- ¹⁴Wallhoff, F. (2004). FGnet – Facial expression and emotion database. [Online]. Retrieved from <http://www.mmk.ei.tum.de/~waf/fgnet/feedtum.pdf>
- Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K.R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality & Social Psychology, 78*, 105-119. doi:10.1037/0022-3514.78.1.105
- ⁵Yin, L., Chen, X., Sun, Y., Worm, T., & Reale, M. (2008). A high-resolution 3D dynamic facial expression database. In *Proceedings of the Eighth International Conference on Automatic Face and Gesture Recognition* (pp. 1-6). Los Alamitos, CA: IEEE Computer Society. doi:10.1109/afgr.2008.4813324
- Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of facial affect recognition methods: Audio, visual and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*, 39-58. doi:10.1109/tpami.2008.52
- ⁴Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., & Liu, P. (2013). A high-resolution spontaneous 3D dynamic facial expression database. In *Proceedings of*

the Tenth IEEE International Conference on Automatic Face and Gesture Recognition

(pp. 1-6). Los Alamitos, CA: IEEE Computer Society. doi:10.1109/fg.2013.6553788

⁴Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., Liu, P., & Girard, J.

M. (2014). BP4D-Spontaneous: A high-resolution spontaneous 3D dynamic facial expression database. *Image and Vision Computing*, 32, 692-706.

doi:10.1016/j.imavis.2014.06.002

Footnotes

ⁱ The search terms used were combinations of the phrases “dynamic face”, “dynamic facial”, “dynamic emotion” and “dynamic emotional” in conjunction with the keywords “expression”, “expressions”, “detection”, “recognition” and “perception”.

ⁱⁱ The overall inclusion criteria were journal articles or conference proceedings that cited at least one of the datasets, and were written in the English language. Since Google Scholar only provides details of the first 1000 search results for each dataset, it was not possible to check a large number of results (n=1922) for the CK and CK+ datasets (642 results from the search by citing references for CK; 1280 results from the combined CK and CK+ abstract/keyword/full title search). After omitting these results, and articles that were either duplicates (n=1145) or whose abstracts/full text were neither obtainable online nor through the British Library and university libraries (n=24) in the United Kingdom, a total of 11,380 articles remained for ‘dataset usage’ classification. The EU-Emotion Stimulus Set (O’Reilly et al., in press) was not included because it was not published at the time of this review.

Table 1. *Conceptual Features of 22 Dynamic Facial Expression Datasets*

Database	Expressions / Action Units (AUs)			Encoder Demographics			Measurement and Validation			
	Emotions (total)	Elicitation	Type	Ethnicities	Age	N	Emotion evaluation	Additional measures	FACS coding	Other features
ADFES ¹	Anger, contempt, disgust, embarrassment, fear, joy, neutral, pride, sadness, surprise (10)	Instruction to perform expression/AUs	Posed	North-European, Mediterranean	18-25	20	Emotion judgments by observers	Arousal and valence ratings	Yes	Head movements (45° towards/away)
BINED ²	Amusement, anger, disgust, fear, frustration, sadness, surprise (7)	Emotion-specific tasks/videos	Spontaneous	Caucasian, Peruvian	Adulthood	256	Emotion self-reports	Continuous intensity and valence ratings, Interrater agreement on valence by trained raters	No	Active and passive emotion elicitation, 3 datasets, Audio-visual recordings
BNED ³	Affectionate, afraid, amused, angry, bored, confident, content, disappointed, excited, happy, interested, loving, neutral, pleased, relaxed, sad, worried (17)	Conversation	Spontaneous/Natural	Caucasian	Adulthood	125	Emotion judgments by trained raters, Interrater agreement on emotion categories	Continuous arousal and valence ratings, Intensity ratings	No	Videos extracted from television or recorded in studio, Standalone but part of Humaine, Audio-visual recordings
BP-4D Spontaneous ⁴	Anger/upset, disgust, embarrassment, fear/nervous, happiness/amusement, pain, sadness, surprise/startle (8)	Emotion-specific tasks	Spontaneous	African-American, Asian, Euro-American, Hispanic	18-29	41	Emotion self-reports, Emotion judgments by observers, Interrater agreement on emotion categories	Machine recognition of expressions, FACS interrater agreement	Yes	2D + 3D stimuli
BU-4DFE ⁵	Anger, disgust, fear, happiness, sadness, surprise (6)	Instruction to perform expression	Posed	Asian, Black, Hispanic/Latino, White	18-45	101		Machine recognition of expressions	No	2D + 3D stimuli
CAM Face-Voice Battery ⁶	Emotion concepts (412) in emotion groups (24), incl. afraid, disgust, happy, sad, surprise	Instruction to perform expression	Posed	Multiple (unspecified)	Preschool-adulthood	6	Emotion judgments by observers		No	Audio-visual recordings, Professional actors
CK ⁷	AU sequences (23), incl. AUs for anger, disgust, fear, joy, surprise, sadness (6)	Instruction to perform AUs	Posed	African-American, Euro-American, other (6%)	18-50	97		Machine recognition of AUs, FACS interrater agreement	Yes	Head rotation up to 30°
CK+ ⁸	AU sequences (23), incl. AUs for anger, contempt, disgust, fear, happy, sadness, surprise (7)	Instruction to perform AUs, Conversation	Posed + Spontaneous (smiles)	African-American, Euro-American, other (6%)	18-50	123	Emotion judgments by observers	Machine recognition of AUs/expressions, FACS interrater agreement	Yes	Spontaneous smiles included, Extension to CK database

Table 1. (continued)

Database	Expressions / Action Units			Encoder Demographics			Measurement and Validation			
	Emotions (total)	Elicitation	Type	Ethnicities	Age	N	Emotion evaluation	Additional measures	FACS coding	Other features
D3D-FACS ⁹	AU sequences (97), incl. AUs for anger, disgust, fear, happiness, sadness, surprise (6)	Instruction to perform AUs	Posed	Caucasian	23-41	10		Machine recognition of AUs	Yes	3D stimuli
DaFEx ¹⁰	Anger, disgust, fear, happiness, neutral, sadness, surprise (7)	Scenarios enactment	Posed	Caucasian	M = 25.6	8	Emotion judgments by observers		No	3 intensity levels, Utterance and non-utterance, Audio-visual recordings, Professional actors
DISFA ¹¹	Disgust, fear, happy, sadness, surprise (5)	Emotion-specific videos	Spontaneous	African-American, Asian, Euro-American, Hispanic	18-50	27		Machine recognition of AUs, FACS interrater agreement	Yes	Frame-level FACS coding (gradation, intensity)
DynEmo ¹²	Amusement, annoyance, astonishment, boredom, curiosity, disappointment, disgust, fright, humiliation, moved, satisfaction, shame, surprise (13)	Emotion-specific tasks/videos	Spontaneous	Caucasian	25-65	358	Emotion self-reports, Continuous emotion judgments by observers, Interrater agreement on emotion expressive segments	Self-reports of arousal and action-readiness	No	Long clips with timelines allowing for excerpt selection, Covert emotion elicitation, Audio-visual recordings, 2 datasets
EU-Emotion Stimulus Set ¹³	Afraid, angry, ashamed, bored, disappointed, disgusted, excited, frustrated, happy, hurt, interested, jealous, joking, kind, neutral, proud, sad, sneaky, surprised, unfriendly, worried (21)	Scenarios enactment	Posed	Afro-Caribbean-Asian, Black, Mediterranean-Asian, White, White-Asian	10-70	19	Emotion judgments by observers	Valence, arousal, and intensity ratings	No	2 intensity levels (for 6 emotions), Body gesture and contextual social scenes, Audio-visual recordings, Professional actors
FG-NET FEEDtum ¹⁴	Anger, disgust, fear, happiness, neutral, sadness, surprise (7)	Emotion-specific videos	Spontaneous	Caucasian	Adulthood	19	None reported		No	No restrictions regarding head motion
GEMEP Core Set ¹⁵	Admiration, amusement, anxiety, cold anger (irritation) contempt, despair, disgust, joy (elation), hot anger (rage), interest, panic fear, pleasure, pride, relief, sadness, shame, surprise, tenderness (18)	Scenarios enactment	Posed	Caucasian	25-57	10	Emotion judgments by observers	Intensity, authenticity, and plausibility ratings	Yes	2 intensity levels + masked expressions, Frontal and side viewpoints, Audio-visual recordings, Pseudo-speech and nonverbal vocalizations, Professional actors

Table 1. (continued)

Database	Expressions / Action Units			Encoder Demographics			Measurement and Validation			
	Emotions (total)	Elicitation	Type	Ethnicities	Age	N	Emotion evaluation	Additional measures	FACS coding	Other features
Hi4D-ADSIP 3-D ¹⁶	Anger, disgust, fear, happiness, pain, sadness, surprise (7) Other facial articulations (7)	Instruction to perform expression	Posed	Multiple (unspecified)	18-60	80	Emotion judgments by observers	Machine recognition of expressions	No	3 intensity levels, 3D stimuli, Audio-visual recordings, Extension to ADSIP database
HUMAINE ¹⁷	Emotional and conversational expressions, incl. anger, contempt, disgust, fear, happiness, sadness, surprise	Conversation, Interaction, Emotion-specific activities	Spontaneous/ Natural	Multiple (unspecified)	Adulthood		Emotion judgments by trained raters, Interrater agreement on emotion categories	Continuous intensity, valence, and authenticity ratings	No	Comprised of 8 data subsets, incl. BNED, Videos extracted from television or recorded outdoors/in studio, Audio-visual recordings, Multimodal (speech, language, gestures, faces)
MMI ¹⁸	AU sequences (79), incl. AUs for anger, bored, disgust, fear, happy, sad, sleepy, surprise (set I-III) Disgust, happiness, surprise (set IV+V)	Instruction to perform expression and AUs, Emotion-specific videos/sounds (set IV+V)	Posed + Spontaneous (set IV+V)	Asian, Caucasian, South American	19-62	78		EMFACS coding by trained raters	Yes	Frontal and side viewpoints, Frame-level FACS coding (gradation), 5 datasets, Audio-visual recordings (set V)
MPI ¹⁹	Emotional and conversational expressions (55), incl. anger, contempt, disgust, embarrassment, fear, happiness, pain, sadness	Scenarios enactment	Posed	Caucasian	20-30	19	Emotion judgments by observers	Emotion-specific scenarios validation by observers, Naturalness ratings	No	2 intensities, 3 viewpoints, 3D facial scans, Audio-visual recordings
MPI Bio ²⁰	AU sequences (46), incl. AUs for 8 expressions	Instruction to perform AUs	Posed	Caucasian	Adulthood	1	Emotion judgments by observers		No	6 viewpoints, Unilateral expressions
STOIC ²¹	Anger, disgust, fear, happiness, neutral, pain, sadness, surprise (8)	Instruction to perform expression	Posed	Caucasian	20-45	10	Emotion judgments by observers		No	3 intensities, Professional actors
UT Dallas ²²	Anger, boredom, disbelief, disgust, fear, happiness, laughter, neutral, puzzlement, sadness, surprise (11)	Emotion-specific videos	Spontaneous	African-American, Asian, White, Hispanic, other	18-25	284	None reported		No	Conversational and compound expressions, 9 viewpoints

Table 2. Practical Features of 22 Dynamic Facial Expression Datasets

Database	Stimuli		Elements [#]		Resolution [#]			Format		Access Info	
	<i>N</i> (videos / sequences)	Duration	Visible	Controlled	Nominal (w x h)	Face-box (<i>SD</i>)	% (Area)	Modality	Color / Gray	Contact Email (Web Address)	Payment
ADFES	648	5.6-6.5s	HD, NE, SH	BG°, LI°, CL°	720 x 576	358 ² (20)	31	V(.mpg), S(.jpg)	Color	a.h.fischer@uva.nl (http://psyres.uva.nl/research/content/programme-group-social-psychology/adfes-stimulus-set/stimulusset.html)	No
BINED	1400	5s, 30s, 60s, 3 min	HD, NE, SH+UT, AM	BG°, LI	720 x 576; 384 x 288 (Peru)	173 ² (45); 123 ² (Peru; 28)	7;14	V(.mp4), A	Color	g.mckeown@qub.ac.uk (www.psych.qub.ac.uk/BINED/)	No
BNED	298	10-60s	HD, NE, SH+UT	BG°	352 x 288	124 ² (39)*	15	V(.mpg), A(.wav)	Color	g.mckeown@qub.ac.uk (belfast-naturalistic-db.sspnet.eu)	No
BP-4D Spontaneous	328	20s	HD, NE, SH	BG, LI, HR, CL, AS°	1040 x 1392 (V); 37K verts. (3D)	722 ² (41)*	36	V(uncompr.), 3D	Color	lijun@cs.binghamton.edu (http://www.cs.binghamton.edu)	\$200
BU-4DFE	606	ca.4s	HD, NE	BG, LI, HR, CL, AS°	1040 x 1329 (V); 35K verts. (3D)	766 ² (33)	41	V(.jpg), 3D(.wrl)	Color	lijun@cs.binghamton.edu (http://www.cs.binghamton.edu)	\$250
CAM Face-Voice Battery	2472	3-5s	HD, NE, SH	BG, LI°, CL°, AS	320 x 240	118 ² (3) [†]	18	V(.mov), A	Color	golanofy@gmail.com (http://www.jkp.com/mindreading)	£75
CK	486	9-60 frames	HD, NE, SH	BG	640 x 490	288 ² (25)	26	S(.png)	Gray	ner3@pitt.edu (http://www.pitt.edu/~emotion/ck-spread.htm)	No
CK+	593	9-60 frames	HD, NE, SH	BG°	640 x 490	290 ² (26)	27	S(.png)	Gray, Color	ner3@pitt.edu (http://www.pitt.edu/~emotion/ck-spread.htm)	No
D3D-FACS	519	5-10s	HD, NE (processed); HD, NE, SH (raw)	BG, LI, HR	1280 x 1024; 30K verts. (3D)	664 ² (44)	34	V(D3D-FACS browser), S(.bmp), 3D(.obj)	Color	dpc@cs.bath.ac.uk (https://vision.cs.bath.ac.uk/~hg299/d3dfacs_hg/)	No
DaFEX	1008	4-27s	HD, NE, SH	BG, LI, CL, AS	360 x 288	140 ² (12)	19	V(.avi/.mpg), A	Color	mana@fbk.eu (http://i3.fbk.eu/resources/)	No
DISFA	54	4 min	HD, NE, SH	BG°	1024 x 768	346 ² (24)	15	V(.avi)	Color	mmahoor@du.edu (enr.du.edu/mmahoor/DISFA.htm)	No
DynEmo	358	6s-4.7 min	HD, NE, SH	BG°, LI	768 x 576	308 ² (6)	21	V(.mpg), A	Color	anna.tcherkassof@upmf-grenoble.fr (https://dynamo.upmf-grenoble.fr/)	No
EU-Emotion Stimulus Set	418	2-52s	HD, NE, SH	BG, LI, CL°, AS	640 x 360	154 ² (12)*	10	V(.mov, .mpg), A	Color	heo24@medschl.cam.ac.uk (http://www.autismresearchcentre.com/projects/Emoticons.aspx)	No

Table 2. (continued)

Database	Stimuli		Elements [#]		Resolution [#]			Format		Access Info	
	<i>N</i> (videos / sequences)	Duration	Visible	Controlled	Nominal (w x h)	Face-box (<i>SD</i>) [§]	% (Area)	Modality	Color / Gray	Contact Email (Web Address)	Payment
FG-NET FEEDtum	399	30s-5 min	HD, NE, SH	BG [°] , LI	640 x 480	248 ² (14)	20	V(.avi)	Color	fgnet@mmk.ei.tum.de (http://cotesys.mmk.e-technik.tu-muenchen.de)	No
GEMEP Core Set	3780	ca.1-5s	HD, NE, SH	BG, LI	720 x 576	205 ² (22)	10	V(.avi), A	Color	GEMEP@unige.ch (www.affective-sciences.org/gemep)	No
Hi4D-ADSIP 3-D	3360; 80 3D models	3-10s	HD, NE, SH (2D); FA (3D)	BG, LI	2352 x 1728 (V); ca. 20K vertices (3D)	1269 ² (39) [‡]	34	V(HD), 3D(.obj), A	Color, Gray	bmatuszewski1@uclan.ac.uk (n/a)	No
HUMAINE	63	4s-30 min	HD, NE, SH, UT, AM	BG [°]	384 x 288	82 ² (19) [*]	6	V(.avi), A	Color	g.mckeown@qub.ac.uk (http://humaine-db.sspnet.eu/)	No
MMI	848 V, 740 SQ	0.5-80s	HD, NE, SH	BG, LI [°]	720 x 576; 576 x 720; 640 x 480; 1200 x 1600 (S)	182 ² (32); 379 ² (26); 236 ² (22); 983 ² (S; 184)	8;35; 18;50	V(.avi), S, A	Color	mmi_face_db@mahnob-db.eu (http://mmifacedb.eu/)	No
MPI	439 (validated); 19152 (total)	1-10s	HD, NE	BG, LI, HR, CL	384 x 288 (V); ca. 133 ² (10) 75K vertices (3D)		16	V(.avi, .mpg), A	Color	kathrin.kaulard@tuebingen.mpg.de (ftp://aedb:UurXMs3WtOkh1F@ftp.tuebingen.mpg.de/)	No
MPI Bio	324	0.68-7.76s	HD, NE	BG, LI, CL, HR, AS [°]	384 x 288 (V); ca. 174 ² (9) 75K vertices (3D)		27	V(.avi, .mpg), 3D	Color	vdb@tuebingen.mpg.de (http://vdb.kyb.tuebingen.mpg.de/)	No
STOIC	80	0.5s	FA	BG, LI, HR, CL, AS	256 x 256	193 ² (16)	57	V(.mov)	Gray	frederic.gosselin@unmontreal.ca (mapageweb.umontreal.ca/gosselif/STOIC.rar)	No
UT Dallas	2556	5-10s	HD, NE, SH	BG, LI [°] , CL, AS [°]	720 x 480	259 ² (23)	19	V(.dv), S(.tif)	Color	mgh100020@utdallas.edu (bbs.utdallas.edu/facelab/database/)	£100

Note. # For databases containing multiple subsets, analyses pertain only to face-focused subsets (excluding social or postural subsets). [°] This aspect was controlled only to a limited degree. [§] Face-box was measured by means of a custom software OpenCV (Bradski, 2000) and a Haar classifier (Viola & Jones, 2001) which locate the human face in the image and return the width and height of the face bounding box. It can be described as either the absolute number of pixels² or as the relative proportion of the visible facial area in comparison to the absolute image size (% Area). * Box estimate based on a subset of (available/suitable) videos. † Box estimate based on one sample video only. ‡ Box estimate based on image samples extracted from the article itself (pdf).

Key descriptions: Stimuli: V= Video; SQ= Sequences; Format: V= Video; A= Audio; S= Still images; 3D= 3-dimensional object files; Visible Elements: FA= Face; HD= Head, NE= Neck, SH= Shoulders; UT= Upper Torso; AM= Arms; Controlled Elements: BG= Background; LI= Lighting; HR= Hair; CL= Clothing; AS= Accessories.

Table 3. Citation and Usage Frequencies of Dynamic Facial Expression Datasets

Dataset	Citing Reference				Keyword or Full title				Dataset Usage		
	Google Scholar	Web of Science	Scopus	PsycArticle, PsycInfo, MedLine, PubMed	Google Scholar	Web of Science	Scopus	PsycArticle, PsycInfo, MedLine, PubMed	Social Sciences	Computer Sciences	Total
ADFES	40	18	21	5	137	3	3	5	13	4	17
BINED	24	9	14	0	28	138	101	0	1	10	11
BNED	148	52	0	0	76	0	0	0	1	18	19
BP-4D Spontaneous	21	0	0	0	9	1	1	0	1	77	78
BU-4DFE	176	38	3	0	209	15	27	0	1	5	6
CAM Face-Voice Battery	159	72	82	39	182	4	3	39	16	6	22
CK*	1642	676	862	0	2280	206	334	0	22	773	795
CK+*	413	139	167	0				0	0	187	187
D3D-FACS	23	5	7	0	15	1	1	0	0	2	2
DaFEx	29	4	5	0	378	13	22	0	2	20	22
DISFA	29	4	17	0	22	2	8	0	0	19	19
DynEmo	6	1	0	0	8	17	0	0	1	0	1
FG-NET FEEDtum	98	0	0	0	687	61	123	0	6	121	127
GEMEP	488	159	101	16	480	15	25	16	28	66	94
Hi4D-ADSIP 3-D	24	11	13	0	22	3	3	0	0	7	7
HUMAINE	162	53	33	0	187	3	1	0	0	32	32
MMI	431	165	92	0	371	16	29	0	5	172	177
MPI	44	20	8	4	233	9	5	4	3	2	5
MPI Bio	3	0	0	0	561	1	2	0	4	3	7
STOIC	13	15	0	0	15	1	0	0	10	4	14
UT Dallas	110	42	57	5	182	2	3	6	10	15	25
Total	4083	1483	1482	69	6082	511	691	70	124	1543	1667

Note. Citing Reference: Papers which referenced the dataset. Keyword or Full title: Papers in which the abstract, full title of each dataset, or any known acronyms (if applicable) were mentioned in conjunction with combinations of the keywords 'face', 'facial', 'expression' and 'database' using the Boolean 'AND' and 'OR' operators as appropriate. Dataset Usage: Papers in which a dataset was used as stimulus material in research with human participants (Social Sciences) or for the training and testing of machine learning algorithms (Computer Sciences).