Emotion Elicitation with Stimuli Datasets in Automatic Affect Recognition Studies – Umbrella Review

Pawel Jemiolo, Dawid Storman, Barbara Giżycka, Antoni Ligęza

To cite this version:

This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Emotion Elicitation With Stimuli Datasets In Automatic Affect Recognition Studies – Umbrella Review

Paweł Jemioło[1], Dawid Storman[2,3], Barbara Giżycka[1], and Antoni Ligęza[1]

[1] AGH University of Science and Technology,
al. A. Mickiewicza 30, 30-059 Krakow, Poland
{pawljmlo,bgizycka,ligeza}@agh.edu.pl
[2] Jagiellonian University Medical College,
Department of Hygiene and Dietetics, Systematic Reviews Unit
[3] Department of Adult Psychiatry, University Hospital
dawid.storman@doctoral.uj.edu.pl

Abstract. Affect Recognition has become a relevant research field in Artificial Intelligence development. Nevertheless, its progress is impeded by poor methodological conduct in psychology, computer science, and, consequently, affective computing. We address this issue by providing a rigorous overview of Emotion Elicitation utilising stimuli datasets in Affect Recognition studies. We identified relevant trials by exploring five electronic databases and other sources. Eligible studies were those reviews identified through the title, abstract and full text, which aimed to include subjects who underwent Emotion Elicitation in laboratory conditions with passive stimuli presentation for Automatic Affect Recognition. Two independent reviewers were involved in each step in the process of identification of eligible studies. The discussion resolved any discrepancies. 16 of 1308 references met the inclusion criteria. The 16 papers reviewed 271 primary studies, in which 3515 participants were examined. We found out that datasets containing video, music, and pictures stimuli are most widely explored, while researchers should focus more on these incorporating audio excerpts. Five of the most frequently analysed emotions are: sadness, anger, happiness, fear and joyfulness. The Elicitation Effectiveness and techniques towards emotion assessment, are not reported by the review authors. We also provide conclusions about the lack of studies concerning Deep Learning methods. All of the included studies were of *Critically low* quality. Much of the critical information is missing in the reviewed papers, and therefore a comprehensive view on this research area is disturbingly hard to claim.

Keywords: Umbrella Review · Dataset · Stimuli · Emotion Elicitation · Automatic Affect Recognition · Affective Computing

1 Introduction

For several decades now, combined efforts in the field of Affective Computing and Artificial Intelligence have been proving to facilitate our day-to-day activi-

ties. Although Affective Computing aims to enhance our lives with emotionally intelligent technologies, many promising ideas and designs are still confined in research laboratories. Since its advent in the 1990s [51], Affective Computing has been promoting an approach that emphasises three areas of research in Human-Computer Interaction: Emotion Recognition, emotional data interpretation, and affective computer or system behaviour.

Affective data can come in different formats and from various levels, ranging from physiological to behavioural. Physiological data encompasses all of the body biological signals [23,68], while behavioural cues related to changes in one's emotional state comprise information conveyed by, e.g., facial expressions [2] and gestures [30]. To develop a system capable of handling this kind of information, more so a system equipped with Artificial Intelligence technology, it is fundamental to implement it with an emotion classification model [15]. Although such models can be trained with data harvested from people's everyday activities, such as cameras, GPS data, or the recordings of wearable technologies, in this paper, we focus on Emotion Elicitation performed in the laboratory setting studies because due to controlled conditions (such as limiting the influence of confounders, strict protocol), we obtain more accurate results. One of the essential steps in an experimental protocol for developing Emotion Recognition models is Emotion Elicitation, employing affect-inducing stimuli presentation. Type of the stimuli can affect not only participants emotions [14,75] but also Emotion Recognition efficiency [74]. The stimuli used in the experiments can either be prepared by the study author or be selected from one of the several available stimuli datasets.

The contemporary methodology of affective experiments is placed in a difficult situation. For at least a decade, the field of psychology has been suffering from a so-called replicability crisis [70], which refers to ongoing difficulties with the reproduction of scientific studies, rooted in poor methodological conduct and lack of robustness in scientific method descriptions in papers. Although the problem stems from social sciences and medicine, other areas – including Affective Computing – are severely affected by it as well. The persistent need for well-proven and meticulously verified solutions is reflected in the demand for Meta-Analyses and Systematic Reviews, which was our motivation for this work.

This paper aims at providing an Umbrella Review of the current state of Emotion Elicitation with stimuli datasets (comprised of at least two elements) by analysing recent reviews on the topic. We concentrate on Automatic Affect Recognition, meaning that the emotional responses are detected and recognised by an artificial system. We also analyse the methodological discrepancies between the reviews included in this Umbrella Review.

The rest of the paper is organised as follows. In Section 2, we provide the context for our work by a brief description of Emotion Recognition and Affect Elicitation, as well as a brief outline of several currently available emotional stimuli datasets. Section 3 specifies the characteristics of Systematic Reviews, emphasising the importance of using reviews to develop consistent research methodologies. In Section 4, we precisely outline the procedure of how our Umbrella Review was conducted. In Section 5, we put forward the results of our overview analysis, and

in Section 6, we discuss them. The paper ends with Section 7, where we provide general conclusions and suggestions for future works.

2   Affective Experiments Outline

After deciding on a theoretical approach, Emotion Recognition researchers may proceed to the experiment planning and design phase. Usually, the standard protocol involves a series of steps, including the experiment itself (stimuli presentation and experimental task), following data pre-processing, feature extraction, classification or regression, and finally validation (see Figure 1). This paper focuses on the first (excluding the design phase) step of the Emotion Recognition, meaning the stimuli set preparation and choice.
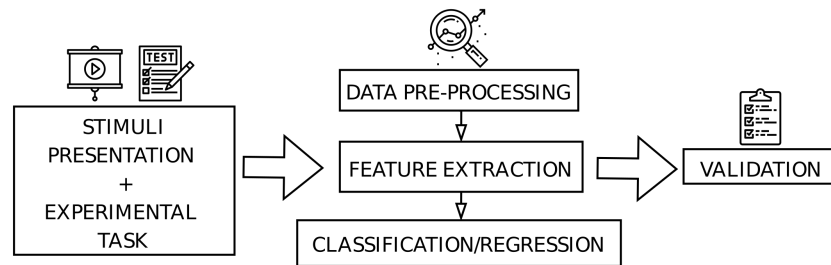

Fig. 1. An illustration of an experimental procedure for Emotion Recognition.

For reliable results of recognising emotions through automatic detection and identification, solid knowledge background in emotion theory is of great importance [67]. Unfortunately, for more than 150 years of scientific studies on this topic, researchers still cannot agree on one approach to the subject. However, in most modern experiments, it is possible to notice a distinctive contrast concerning the theoretical background.

2.1   Emotion Theories

Researchers may adopt one of the numerous theories of emotion for studies on emotion, developed throughout decades of psychological research. The vast range of available approaches includes, among others, a so-called *discreet* and *continuous* stance on emotions. In the discreet emotion theories, each of human emotions (usually including basic, i.e. happiness, sadness, anger, surprise, disgust, and fear) is a category of its own, having features and behaviours that are specific just for them, and which are triggered by a specific neural circuit in the brain [17, 47]. As for the continuous approach, emotional states are interpreted on a scale of level of arousal (from high to low intensity of emotion) and valence (from positive to negative) [54], sometimes with an additional dimension of dominance (from top to slight degree of feeling of being in control) [45].

2.2   Emotion Elicitation

To collect data on emotional states, a researcher needs proper tools for evoking affective responses in the study participant. Emotions can be induced in either an active or a passive manner [33]. The active approach engages the study participant to interact with the experimental setup (e.g., video games [49]) or a passive manner, where the stimuli are presented in a way that does not require any action from the participant. The passive materials include images [12,34,36,40,43], emotion-evoking videos [13,38,56], sounds [6], and music tracks [63].

Many static picture databases contain photographs of human facial expressions [1,16,37,42]. While selecting a proper dataset for an experimental procedure, one needs to focus on the stimuli quantity and length (in the case of video and audio databases). These features tend to vary among different databases. These characteristics need to be controlled to ensure that a specific emotion is effectively elicited in each subject, regardless of individual variations. In turn, an effective elicitation enables the generalization of results.

2.3   Emotional Datasets

The study authors are free to identify the stimuli that will best serve the aims of their experiments. Nonetheless, it is often convenient to use one of the publicly available affective stimuli datasets. The specific structure of each database may vary from one to another, but they comprise a collection of stimuli in general. Usually, those include pictures/photographs, audio tracks, or videos, which are adequately annotated or labelled concerning their content and specific emotion designed to induce. For example, the affective annotations are provided in terms of valence and arousal scores [34, 36] or basic emotion labels [1, 16, 37, 42] (or distributions of labels [50]), acquired through analysing the ratings performed by the subjects in initial studies during database development. Although the databases should be validated and standardised for research purposes to eliminate the possibility of incurring any uncontrolled variables, some do not meet this requirement. Several of the stimuli sets are listed and briefly described below.

- IAPS (International Affective Picture System) [36] is one of, if not *the*, most often used sets of standardised, emotionally evocative pictures. It includes 1192 images of various aspects of human experience, i.e., people, architecture, everyday objects, animals, landscapes, and unpleasant sightings such as mutilated bodies.
- IADS (International Affective Digitised Sounds) [8] is an IAPS counterpart for the auditory modality. Its latest version [7] consists of 167 sounds that one can encounter in their daily life, such as sounds of nature, people talking, laughing, etc. Each sound has a duration of precisely 6 seconds. Recently, Yang et al. [71] has proposed a revised and extended version of this set, called IADS-E.
- ANEW (Affective Norms for English Words) [6] is a collection of 1034 English words developed for emotion research. Other language adaptations of ANEW are also present, including, e.g., German [58] or Spanish [53].

- GAPED (Geneva Affective Picture Database) [12], with 730 pictures, was created to increase the availability of visual, emotional stimuli. The picture categories are strictly specified in terms of their negative (spiders, snakes, scenes of moral and legal norms violation), positive (human and animal babies, nature sceneries), and neutral (inanimate objects) content.
- NAPS (Nencki Affective Picture System) [43] attempts to provide the largest (to date) database of visual stimuli, with 1356 high-quality photographs of 5 categories: people, faces, animals, objects, and landscapes, annotated by mostly European population. Each picture physical properties are also provided with valence, arousal ratings, and approach-avoidance ratings (luminance, contrast, entropy).
- OASIS (Open Affective Standardised Image Set) [34] is an open-access online database of 900 colour images of various themes, described with valence and arousal normative ratings. The distinctive features of this collection are that it allows for free research use and has been assembled using online sources, unlike databases restricted by copyrights, such as IAPS images.
- CAPS (Chinese Affective Picture System) [40] is an answer to the need for a culturally accurate affective stimuli database, in this case, for the Chinese.
- POFA (Pictures of Facial Affect) [16] is (dating to 1993) presumably one of the oldest affective datasets. It consists of 110 black and white photographs of facial expressions of six basic emotions.
- LIRIS-ACCEDE (the Annotated Creative Commons Emotional DatabasE) [13] is a public database of 9800 video excerpts shared under Creative Commons licenses and annotated along affective dimensions of valence and arousal.

With the above examples being only a few from the variety of existing databases, researchers have a sound number of Emotion Elicitation tools from which they can choose. The question of which of these are used most often and how exactly they are applied to Emotion Recognition studies was one of our motivations for developing this review.

2.4   Emotion Elicitation Effectiveness

Another factor that motivated us to conduct this research was determining what emotions are exactly elicited in the studies and the effectiveness of different elicitation databases. One needs to remember that many features may affect the process of emotion induction in individuals. The personality [35] is one of them, and the mood [11] is the other. Due to this, the emotions that participants of the experiment experience could be different than assumed while designing the protocol (see the beginning of Section 2).

It is thus necessary to examine the effectiveness of the Emotion Elicitation process. It can be measured using Detection Theory metrics [21], e.g. accuracy.

Additionally, one must remember that emotional labels can be gathered using different methods. There is, for example, a standardised tool called SAM (Self-Assessment Manikin) [5]. Experimenters or other non-related people may also assure labels. What is more, one can also easily calculate the agreement between mentioned methods [11].

3   Reviews

Due to the rapidly increasing number of primary studies, the need to analyse and synthesise them grew [52].

Particularly in medicine, where clinical decisions should be based on credible data, this need to develop data identifying and synthesis methods in an unbiased, rigorous, and transparent approach is present. These are the basis for conducting Systematic Reviews, distinguishing them from traditional reviews.

The quality of Systematic Reviews depends mainly on the extent to which the methods are used to minimise the risk of error and bias while conducting a review [3]. The standards for producing high-quality Systematic Reviews are supposed to support best practices in publishing scientific studies. They indicate how to include appropriate sections and information in authors' manuscripts that can be understood by the readers or replicated by the researchers. Some organizations provide them, e.g., Cochrane Collaboration (health care), Campbell Collaboration (social, behavioural, and educational areas) or EQUATOR Network (Enhancing the Quality and Transparency of Health Research), which is an international collaboration that promotes quality and consistency in research publications [18]. These organizations publish their guidelines for conducting different types of reviews [24] and provide reporting tools [39, 46].

In Systematic Review, the fundamental research unit included in the synthesis is a primary study, e.g., cross-sectional, observational, experimental. Due to the growing popularity of secondary studies and an enormous number of published papers, it is impossible to follow their results. Due to this and the need to collect evidence in reduced timeframes, tertiary studies methodology (Umbrella Reviews, also known as Meta-review, Overview) was developed. The defining feature of Umbrella Review is considering a Systematic Review as the first and often the only study type for inclusion [52, 62]. Because the methodology of Systematic Reviews is not well-grounded in Artificial Intelligence, and there are only a few papers published as a Systematic Review, we decided to include all types of reviews in our article to collect as much interesting data as possible.

4   Materials and Methods

4.1   Protocol, Search Strategies, Inclusion Criteria

First, we conducted a pre-search before April 2020 to determine if any interesting papers could be eligible for our study. A protocol has been published [28] on the Open Science Framework (OSF) platform on 07 June 2020 before the data extraction stage. Electronic databases (MEDLINE Ovid, EMBASE, Web of Science, dblp, Cochrane Library), IEEE library, and preprint database arXiv were initially searched from inception to 09 May 2020 using predefined search strategies.

We included both free-text terms, such as *review, overview, dataset*, and MESH terms. We combined: 'dataset.ti, ab. or (video* or picture* or photo* or stimuli or audio or sound*).ti,ab' and 'exp emotions/ or (emotion*).ti,ab.' and

'exp Recognition, Psychology/ or (recogni* or classif* or regres* or clusteriz* or discriminat*).ti,ab.' and '(facial expression* or body movement* or gesture* or speech or behavio?r* or eye gaze or eye movement* or physiological or ECG or electrocardiograph* or EDA or electrodermal activity or GSR or galvanic skin response or EEG or electroencephalograph* or BVP or blood volume pressure* or HR or HRV or heart rate* or EMG or electromyograph* or temperature*).ti,ab.'

Search strategies are provided on the OSF [28] in Appendix 1. Language restrictions were imposed. We restricted to papers written in English and Polish.

Eligible studies were reviews including at least 50% healthy adult ($\geq$ 18 years old) subjects who underwent Emotion Elicitation in laboratory conditions with passive stimuli presentation for Automatic Affect Recognition. We did not exclude those reviews that omit the information about age, health and number of the subjects, and the setting. All the records containing information about post-conference books were excluded. However, the aforementioned does not apply to conference publications themselves.

4.2 Screening, Data Extraction, Quality Assessment

Identified references were checked for duplications using Endnote (Clarivate Analytics ®). We screened titles and abstracts using Rayyan software [48]. Then, full texts of included papers were assessed for meeting our eligibility criteria. Using a pre-specified extraction form, we extracted data from included articles. Following published protocol with research questions [28], we were particularly interested in the type of stimuli, their number and length, emotions – how they were elicitated and measured in primary studies and emotions. We also extracted information about the standardization, validation, and public availability of the datasets. We also focused on setting and Emotion Recognition procedures, e.g., pre-processing, features, algorithms, and validation process. Besides, we collected the necessary information about the population and bibliometric data. All the mentioned stages (identification, screening, eligibility assessment, and extraction) of conducting Systematic Reviews were done independently (PJ and BG).

Quality assessment was also done independently in pairs (partly by PJ and BG and partly by PJ and DS). QASR [27] tool was applied for quality assessment. It is based on a well-known method developed for healthcare, i.e., A Critical Appraisal Tool for Systematic Reviews (AMSTAR 2) [59]. This adapted instrument contains several generic domains about conducting Systematic Review, e.g., registration of protocol, comprehensive literature searching, reproducible and transparent assessment of included articles, or declaring funding and conflict of interest.

Pilot exercises preceded each phase (screening, full-text assessment, extraction, or quality assessment) of the presented Umbrella Review. These exercises were aimed at improving the common understanding of criteria. We worked separately on a material sample until achieving a 90% agreement. We resolved any conflicts through discussion and consensus.

4.3   Analysis

We analysed bibliographic and other most essential characteristics of included studies descriptively. We prepared a qualitative summary concerning the number of primary studies, population, outcomes (Emotion Elicitation in Automatic Affect Recognition), datasets, experimental procedures (Automatic Affect Recognition), and quality of reviews.

Additionally, we planned on conducting a quantitative analysis. However, due to insufficient reporting and poor quality, we could not conduct a planned statistical synthesis. The numbers and calculations are based on these reviews that report on specific factors for all primary studies, except interesting outcomes – data regarding them we extracted from primary studies.

5   Results

5.1   Study Selection

Primary electronic databases searches yielded a total of 869 references to screen after duplicates were removed. Of these, 54 full texts were obtained and screened. Altogether 16 studies [4,9,10,19,22,29,32,44,55,57,60,61,64–66,73] were included for qualitative synthesis.
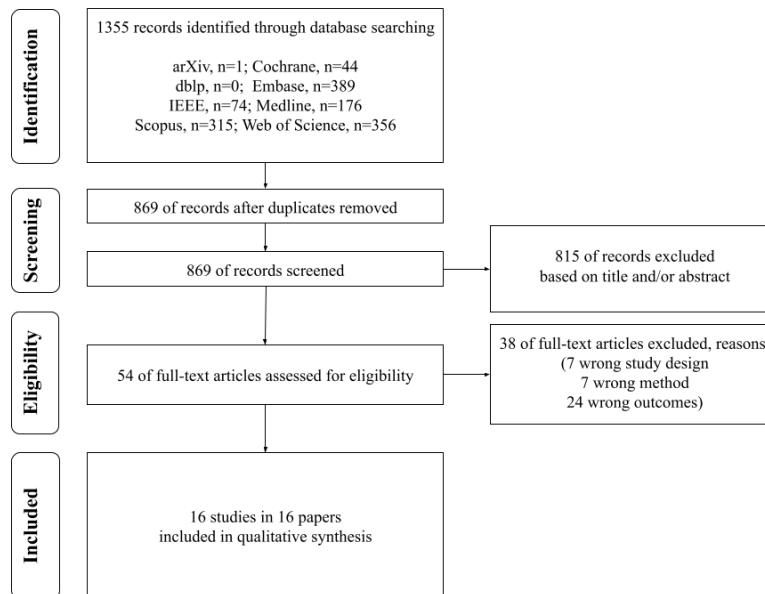

Fig. 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) study flow diagram.

The study flow is presented in Figure 2. Our reporting is consistent with PRISMA guidelines [39,46]. Lists of included and excluded studies (with reasons) and full details are available online on the OSF platform [28] in Appendices 2, 3, and 4, respectively. The spreadsheet forms are also available from authors on request.

5.2   Characteristics of Included Studies

Table 1 presents the essential characteristics of the included trials. 14 (87.5%) included studies [4, 9, 10, 19, 22, 32, 44, 55, 57, 60, 61, 64, 66, 73] were published after 2017. The included reviews were based on 530 primary studies and 271 interesting papers (mean: 83.33 and 16.94, respectively). However, 10 (62.5%) of considered reviews [4, 9, 10, 29, 44, 57, 60, 61, 65, 66] did not provide explicit information about the number of analysed papers.

Table 1. Most essential characteristics of the included studies.

Review ID	Included studies		Population			QASR [27][6]
	Total[1]	Analysed[2]	Total[3]	Age[4]	Gender[5]	
[22]	100	39	PI	PI	NR	5/3
[57]	U	6	PI	PI	NR	7/5
[61]	U	25	PI	NR	NR	7/5
[66]	U	22	526	NR	PI	6/4
[9]	U	4	112	NR	PI	8/5
[19]	62	19	702	NR	NR	7/4
[73]	40	2	47	NR	NR	8/5
[65]	U	8	245	NR	PI	8/5
[29]	U	10	PI	NR	NR	8/5
[10]	U	6	PI	NR	NR	8/5
[55]	137	45	1634	PI	PI	7/4
[64]	60	57	NR	NR	NR	5/3
[60]	U	2	NR	NR	NR	7/5
[44]	U	10	PI	NR	PI	7/4
[32]	131	4	92	PI	PI	7/5
[4]	U	12	157	NR	NR	8/5

PI – Partial Information (provided only for some primary studies);
U – Unclear; NR – Not Reported.
[1] Total number of studies included by a review.
[2] Total number of interesting studies in this review.
[3] Total number of the subjects.
[4] Age of the subjects.
[5] Gender of the subjects.
[6] QASR results: number of flaws/number of critical flaws.

There were at least 3515 participants altogether in included trials. Sample sizes ranged from 47 [73] to 1634 [55] subjects. 6 (37.5%) articles [10,22,29,44,57, 61] report on the number of participants only in part of the primary studies, and

2 (12.5%) publications [60,64] did not contain such information at all. Similarly, 12 (75%) included papers [4, 9, 10, 19, 29, 44, 60, 61, 64–66, 73] did not report on the age of subjects, and 11 (68.75%) trials [4, 10, 19, 22, 29, 57, 60, 61, 64, 73] on sex or gender. The rest provided only partial information.

Synthetic results regarding datasets are presented in Table 2. Studies utilising video (adverts, movies excerpts), music (different genres), and audio (sounds of nature, people talking, laughing) datasets were reported in 15 (93.5%) [4,9,10,19, 22, 29, 32, 44, 55, 57, 61, 64–66, 73], 12 (75%) [4, 10, 19, 22, 29, 32, 44, 55, 61, 64–66], and 7 (43.75%) [4, 19, 22, 32, 55, 61, 64] included reviews. Publications focused on papers using datasets containing only facial expressions were provided by 3 (18.75%) works [32, 55, 64], while 12 (75%) reviews [4, 9, 10, 19, 29, 32, 44, 55, 57, 60, 61, 64] attached information about datasets which comprise various pictures (architecture, landscapes, and unpleasant sightings such as mutilated bodies).

Table 2. Emotion Elicitation Datasets.

Type of Stimuli	Mean (Range)
No. of datasets used in primary studies	
Total	17.38 (2-65)
Video	6.13 (0-34)
Music	3.07 (0-21)
Audio[2]	1.07 (0-8)
Face	0.33 (0-4)
Picture	3.20 (0-18)
Other	0.31 (0-2)
Length[1] of stimuli used in primary studies [seconds]	
Total	129.03 (2-600)
Video	63.75 (15-393)
Music	157.08 (2-600)
Audio[2]	19.00 (6-30)

[1] Calculated only for primary studies with length provided.
[2] Sounds other than music.

Moreover, 4 (25%) studies [29, 44, 55, 65] analysed references incorporating other types of datasets, i.e., multimodal, 3D photorealistic models, and emotional narratives (text). The reviews did not explain what *multimodal* in this context means. 2 (12.5%) studies [22,64] employed own incoherent classification. To name it, authors did not differentiate photo and face or music, audio and video dataset. We, therefore, cannot assign their primary references accordingly and excluded them from the analysis.

It should be noted that the length of stimuli was based on only 2 (12.5%) reviews for video [9,57] and music [55,66] datasets (containing 5 and 25 primary studies respectively), and 1 (6.25%) in terms of audio [55] (4 primary articles). 8 (50%) interesting trials [19,22,29,32,44,60,61,64] did not contain information about the number of stimuli in datasets, while 7 (43.75%) [4,9,10,55,57,65,66] described it only for some of the included studies.

Finally, 1 (6.25%) work [73] analysed such data in full (2 primary studies). According to it, the number of elements in datasets varied from 15 to 40.
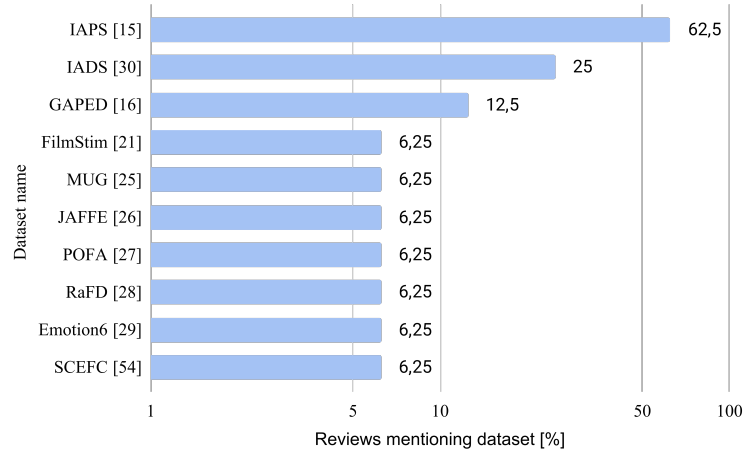

Fig. 3. Reviews with primary studies utilising publicly available datasets, log. scale.

Additionally, we checked which publicly available datasets were reported in the reviews (Figure 3). There were 10 datasets utilised in Emotion Elicitation for Automatic Affect Recognition. However, 6 of them were mentioned only by exactly 1 (6.25%) review [55]. Validation of datasets was partly checked by authors of 2 (12.5%) reviews [55,57], while only 1 (6.25%) study [4] partly focused on the availability of standardised instructions.

Regarding the outcomes, we focused on emotions elicited with stimuli datasets. There were 132 different states reported in the reviews. We found out that some of them overlap themselves, and that is why we merged them into the same categories, e.g. *relaxed*, *relax*, *relaxation* and *relaxing* fell into the same group. After this processing, we got 101 separate affective states.

The most common discrete emotions evoked in the studies included in 16 analysed reviews are presented in Figure 4. On the other hand, some researchers focused on dimensional emotions. 78.23% of included papers concerned discrete states, while only 9.59% was connected with dimensional ones; the rest did not report the nature of elicited emotions.

We also wanted to establish the most popular tools used for emotion assessment. However, the reviews provided information only about 4 out of 271 studies, two of them used SAM [5], and the other 2 utilised the questionnaire, but the name was not provided.

Only Sarma and Barma [55] provided information about the protocol of emotion elicitation with the indication of SAM assessment. On the other hand, they derived only 57.78% of included studies. All of them asked participants about emotions after each stimulus.
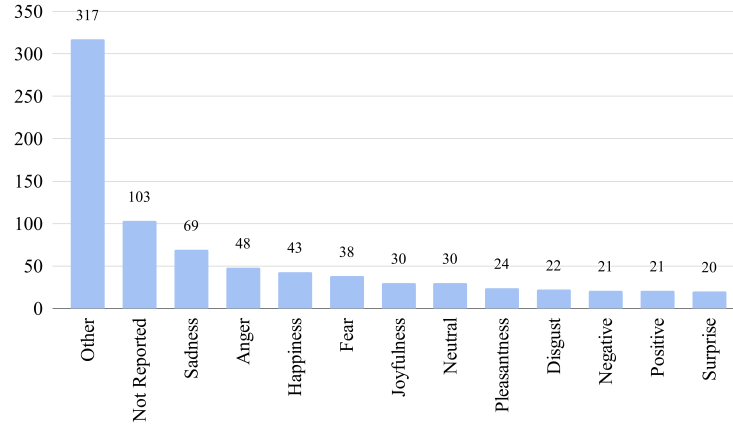
Fig. 4. Discrete emotional states evoked among included primary studies.

Table 3. Additional Information – Automatic Affect Recognition.

		N (%)			N (%)
Modalities			**Features**		
No.			Extraction		
	Single	8 (50)		Reported	10 (63.5)
	Multiple	8 (50)		PR[2]	2 (12.5)
Type				Not Reported	4 (25)
	Physiological	12 (75)			
	Both[1]	4 (25)			
AI methods			**Validation**		
Neural Networks			Methods		
	Reported	9 (56.25)		Reported	1 (6.25)
	Not Reported	4 (25)		PR[2]	3 (18.75)
	Not Applicable	3 (18.75)		Not Reported	12 (75)
Deep Learning			Metrics		
	Reported	4 (25)		Reported	6 (37.5)
	Not Reported	9 (56)		PR[2]	5 (31.25)
	Not Applicable	3 (19)		Not Reported	5 (31.25)

[1] Both – Physiological and Behavioral.

[2] PR – Partially Reported – provided only for some primary studies.

The remaining results concerning the Automatic Affect Recognition procedure are presented in Table 3. All the reviews focused only on physiological or behavioural modalities (or both). The setting was partly provided by 5 (31.25%) trails [32, 44, 55, 57, 61]. 13 (81.25%) of included studies [4, 9, 10, 19, 22, 29, 32, 44, 57, 60, 61, 65, 66] analysed multiple recognition methods.

5.3   Quality of Included Studies

Following the pre-specified method for overall quality assessment [27], we summed up the number of *Flaws* in crucial and non-crucial domains (Figure 5). According to this tool, the quality of all (100%) included papers was assessed as *Critically Low*. Regarding the accumulated number of critical flaws, 10 (62.5%) works [4, 9, 10, 29, 32, 57, 60, 61, 65, 73] contained 5, 4 (25%) articles [19, 44, 55, 66] – 4, and 2 (12.5%) reviews [22, 64] – 3.

The QASR items that most of the included references failed to meet were the following: 3 (independent screening), 4 (extraction process), 5 (justification for excluded studies), 6 (assessment of the Risk of Bias in individual studies), 7 (incorporating Risk of Bias when interpreting results).

The 8th criterion (financial issues) was the only one that was met by almost all reviews. All (100%) analysed studies reported on funding, and 15 (93.75%) [4, 9, 10, 19, 22, 29, 32, 44, 55, 60, 61, 64–66, 73] provided exact information about funding type.
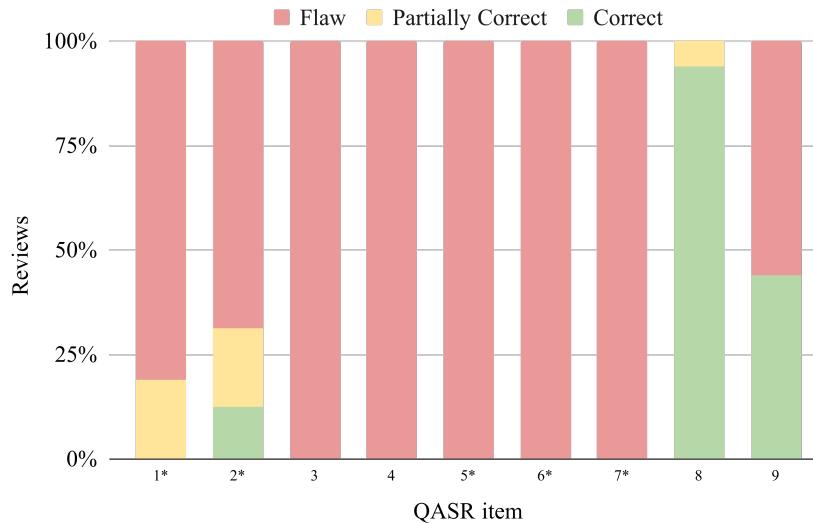

Fig. 5. Percentage of reviews with a given number of critical flaws; * means crucial items.

6   Discussion

6.1   Summary, Quality and Potential Biases in Results

Our overview aimed at analysing Emotion Elicitation with Stimuli Datasets in Automatic Affect Recognition. 16 reviews were included for a thorough qualitative analysis. Using VOSviewer [69] (tool enabling visualisation and frequency

assessment of words in examined records), we visualised the most common words among the early 822 (see Figure 6) and final 16 (see Figure 7) references. The use of VOSviewer confirmed that the included works belong to the area we were interested in this review.

We found that authors most willingly analyse video, picture, and music datasets (see Table 2). However, the last type was overestimated (modal: 1), as 1 included work [66] focused on musical stimulation of the subjects. Its authors described 21 sets, while others provided less than 7 in their overview. Described video and music datasets contained excerpts of various lengths. It did not apply to audio stimuli, which were shorter than 30 seconds. Included reviews mentioned 10 publicly available datasets (see Figure 3) utilised by primary studies. The number of elements within a particular set of stimuli varied from 15 to 40. It should be noted, though, that this range was based on one review [73]. Additionally, we reveal that only 4 (25%) reviews [10, 19, 57, 61] analyse Deep Learning methods, while 12 (75%) [4, 9, 22, 29, 32, 44, 55, 60, 64–66, 73] did not report on validation methods used in primary studies (see Table 3). Similarly, only 4 (25%) of the reviews [10, 44, 57, 73] included works focused on behaviour.

Regarding the outcomes, the authors focus on positive, negative and neutral emotions. In one-seventh of primary studies, the information about the types of evoked emotions was not provided. There were over 100 different names and variants of emotions. Two types of approaches to emotion elicitation (discrete and dimensional) were identified. Researchers more often choose the first approach. The four most explored discrete states correspond to Ekman's model [16], i.e. sadness, anger, happiness and fear. Disgust and surprise are also present but investigated slightly more seldom. Interestingly, researchers seem to distinguish more emotional granularity when it comes to positive states than Ekman, i.e. pleasantness, and joyfulness next to happiness.

Due to the authors skip information about elicitation effectiveness, assessment tools and protocol, and therefore unambiguous conclusions regarding them cannot be made.

Additionally, all of the interesting studies comprised a vast amount of other shortcomings. For example, 10 (62.5%) reviews [4, 9, 10, 29, 44, 57, 60, 61, 65, 66] did not provide the number of included primary studies (see Table 1). Similar applies to characteristics of the examined population, outcomes, and Automatic Affect Recognition process.

What is more, there were several methodological issues connected with the reviews. 4 out of 9 domains were violated by all (100%) studies. According to the QASR tool [27], the overall quality is *Critically Low* for all the reviews.

We, therefore, reckon that our findings should be interpreted with great care. The results say more about the issues concerning the current state of the art reviews than the relationships within Automatic Affect Recognition.

6.2   Suggestions

As for outcomes, we suggest that researchers should specify unequivocal emotion nomenclature. We encourage scientists to explore dimensional emotions often as
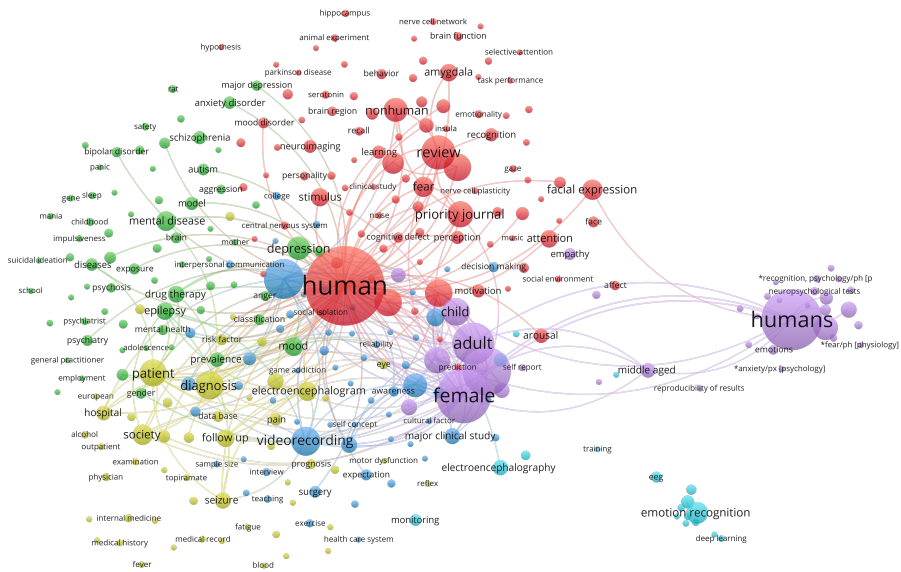
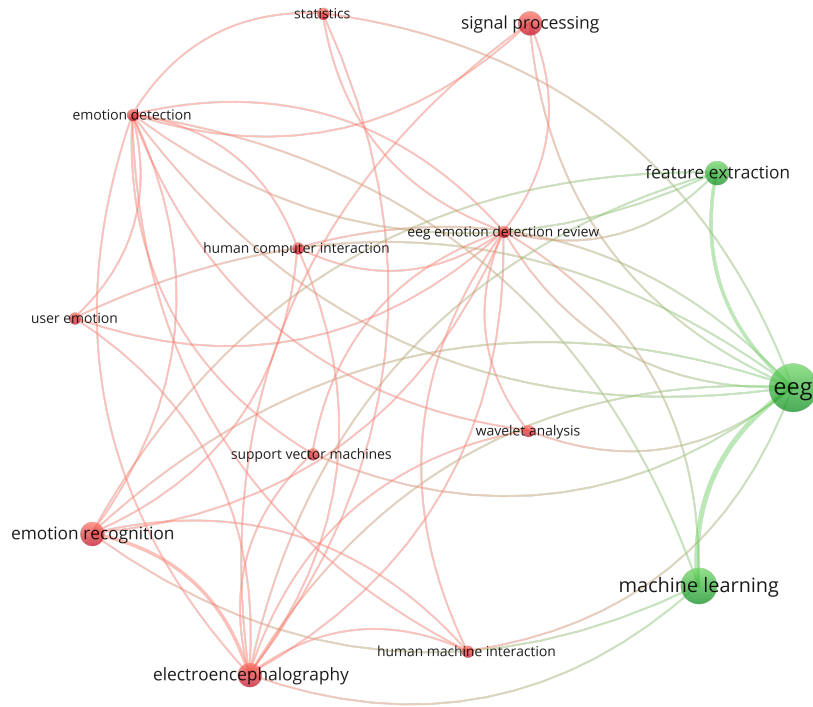**Fig. 6.** Most common words among early 882 studies.

Fig. 7. Most common words among final 16 studies.

they provide an alternative insight into the nature of affective states. What is essential, the authors of the reviews and primary studies ought to conduct and report on Emotion Elicitation effectiveness measurements. The same applies to assessment tools and protocol of the Emotion Elicitation experiments.

Regarding the datasets, we encourage authors to focus more on works utilising facial expressions and audio datasets. More effort should be put on other modalities, e.g., 3D models or emotional narratives. Authors ought to also explore the stimuli generated using Artificial Intelligence methods such as GANs (Generative Adversarial Networks [20]). They may be used when generating stimuli, e.g. pictures, videos, audio, and video games [26] or even signals, e.g. physiological if the data is missing [41]. Additionally, we also believe that researchers should pay attention to sound stimuli length as using more extended excerpts seems an unaddressed issue. We also encourage the reviewers and authors of primary studies to include works containing various publicly available datasets, e.g., [12, 13, 36, 63].

What is more, authors should not confuse stimuli datasets with databases containing physiological signals, e.g., in [55]. Of course, such sets are reused in recognition processes but serve as a benchmark, not a tool for inducing affect. It can be confusing and misleading for scientists who want to utilise specific stimuli in their Emotion Elicitation experiments. Additionally, authors of the reviews should more carefully report on the names of publicly available datasets. For example, *Film Clip* provided in [55] needs to be referred to as *Standard Chinese Emotional Film Clips Database (SCEFC)* [38].

We are aware that authors provide much information about particular sets within the background section, e.g., [55]. They list them but do not refer to the included studies. Since some of these datasets are not used in Automatic Affect Recognition but are incorporated in psychological experiments, it is uninformative from Affective Computing perspective and may lead to biases. As Affective Computing is embedded on the border of Psychology and Computer Science, authors should also provide information about the population [51], i.e., sex and age. Only then correct conclusions about the applicability of experiments in this area to everyday life could be drawn. Authors need to remember also that *Emotion Recognition* may refer to human ability. Thus, we sincerely encourage authors to use the name *Automatic Affect/Emotion Recognition* (at least in abstract and title) to eliminate unwanted misunderstandings.

Next, we believe that reviews should focus more on the current state of the art methods deriving from Deep Learning. This technique allows feature selection to be more flexible, less time-consuming and enhances the models performance than traditional models [31,72]. It is also highly significant to provide information about validation methods and compare the results accordingly [24]. The same applies to posthoc analyses.

Due to the low quality and completeness of the findings mentioned above, we suggest conducting new Systematic Reviews in the area of Emotion Elicitation (and Automatic Affect Recognition in general). The authors should thoroughly

report the utilised methodology with compliance to PRISMA guidelines [39, 46], publish the protocol, and provide synthetic results.

We encourage the authors to provide all the mentioned information as it helps determine whether the issue is addressed in primary studies or whether there is an unaddressed gap in current research. Such a methodological approach can lead to a reduction in resource and time-wasting [24]. Due to the growing number of works in Affective Computing, we reckon that upcoming reviews should focus on particular modalities, i.e., physiological and behavioural, or even better divided into individual signals, e.g., facial expressions, electroencephalography, cardiovascular. The last one is especially relevant due to the growing popularity of wearable devices able to measure such signals.

Following Cochrane guidelines for Systematic Reviews [24, 25], in this Section, we also wanted to address differences and similarities with Umbrella Reviews on the congruent subject. Using Endnote (Claritive Analytics ®), we have chosen 277 studies from 822 selected first. We obtained them by applying search in title and abstracts with keywords specific for Umbrella Reviews, e.g., *Overview, Meta-review*. However, none of the studies selected this way was connected with Affect Elicitation, Emotion Recognition, Affective Computing, Human-Computer Interaction, Artificial Intelligence, or Computer Science.

6.3 Limitations and Strengths

To the best of our knowledge, we are one of the first authors to provide Umbrella Review in the field of Emotion Elicitation (if not in the whole Affective Computing). This study strengths include the publication of protocol and incorporating solid methodological background for all we present here. It enabled us to synthesise the results in a comprehensive, unbiased, rigorous, and transparent way [52]. Despite the page limit, we share all the information gathered in the process of preparing this research via OSF [28], which is a platform for sharing research data with permanent DOI and funding secured for years from now. We also proposed a method for quality assessment (QASR [27]) – it can be incorporated as a checklist for other researchers or reviewers.

On the other hand, the proposed tool is not validated and standardised. Our assessment should thus be regarded as preliminary. However, the tool is an interpretation of AMSTAR 2 [59], which is respected and widely used among reviewers. We also did not search grey literature, and we omit the references to the whole post-conference books. Next, we imposed language limitations on the included publications. However, based on this criterion, we excluded only one paper. Finally, we did not check on the overlapping of interesting works. Nonetheless, following Cochrane guidelines [24], when one wants to *present and describe the current body of Systematic Review evidence on a topic, it may be appropriate to include the results of all relevant Systematic Reviews, regardless of topic overlap.*

7   Conclusions and Future Work

This overview aimed at gathering knowledge about the Emotion Elicitation with Stimuli Datasets in Automatic Affect Recognition experiments.

We found out that authors prefer discrete over-dimensional emotions. However, the nomenclature of affective states is still undeveloped. Review authors do not report on techniques that enable researchers assessment of emotional state. Its effectiveness is also neglected.

Datasets containing video, music, and pictures are most widely explored, while researchers should focus more on these incorporating audio excerpts. We also reveal that authors need to put more effort into analysing Deep Learning methods and incorporating more modalities, e.g. cardiovascular signals, into the process of classification of emotion.

However, considering the *Critically low* quality of studies included in the review, we believe that there is still a place for a comprehensive Systematic Review in discussed area. We suggest that authors follow the strict methodology and use a common language to avoid resource and time-wasting.

We are now working on providing a thorough analysis, including all these remarks. Next, we want to explore active Emotion Elicitation in Affective Computing and focus more on non-laboratory emotion examination. Additionally, we want to validate and standardise the QASR [27] tool utilised in this review.

Additional Information

Authors' Contributions

Paweł Jemioło (PJ) – all listed stages. Barbara Giżycka (BG) – conceptualization, investigation, validation, writing. Dawid Storman (DS) – conceptualization, formal analysis, investigation, methodology, supervision, writing. Antoni Ligęza (AL) – supervision.

Conflict of Interest

Authors declare that they have no conflict of interest.

Funding Sources

References

1. Aifanti, N., Papachristou, C., Delopoulos, A.: The mug facial expression database. In: 11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10. pp. 1–4. IEEE (2010)

2. Anderson, K., McOwan, P.W.: A real-time automated system for the recognition of human facial expressions. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **36**(1), 96–105 (2006)
3. Aromataris, E., Munn, Z.: Chapter 1: Jbi systematic reviews. Joanna Briggs Institute Reviewer's Manual. The Joanna Briggs Institute. (2017)
4. Baghdadi, A., Aribi, Y., Alimi, A.M.: A survey of methods and performances for eeg-based emotion recognition. In: International Conference on Hybrid Intelligent Systems. pp. 164–174. Springer (2016)
5. Bradley, M.M., Lang, P.J.: Measuring emotion: the self-assessment manikin and the semantic differential. Journal of behavior therapy and experimental psychiatry **25**(1), 49–59 (1994)
6. Bradley, M.M., Lang, P.J.: Affective norms for english words: Instruction manual and affective ratings. Tech. rep., The center for research in psychophysiology (1999)
7. Bradley, M.M., Lang, P.J.: The international affective digitized sounds (iads-2): Affective ratings of sounds and instruction manual. University of Florida, Gainesville, FL, Tech. Rep. B-3 (2007)
8. Bradley, M., Lang, P.: International affective digitized sounds: Stimuli, instruction manual and affective ratings. Center for Research in Psychophysiology (1999)
9. Chen, J., Mehmood, R.: A critical review on state-of-the-art eeg-based emotion datasets. In: Proceedings of the International Conference on Advanced Information Science and System. pp. 1–5 (2019)
10. Christensen, L.R., Abdullah, M.A.: Eeg emotion detection review. In: 2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). pp. 1–7. IEEE (2018)
11. Correa, J.A.M., Abadi, M.K., Sebe, N., Patras, I.: Amigos: A dataset for affect, personality and mood research on individuals and groups. IEEE Transactions on Affective Computing (2018)
12. Dan-Glauser, E.S., Scherer, K.R.: The geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance. Behavior research methods **43**(2), 468 (2011)
13. Dellandréa, E., Huigsloot, M., Chen, L., Baveye, Y., Xiao, Z., Sjöberg, M.: Predicting the emotional impact of movies. ACM SIGMM Records (2018)
14. Dhaka, S., Kashyap, N.: Explicit emotion regulation: Comparing emotion inducing stimuli. Psychological Thought **10**(2), 303–314 (2017)
15. D'Mello, S., Kappas, A., Gratch, J.: The affective computing approach to affect measurement. Emotion Review **10**(2), 174–183 (2018)
16. Ekman, P.: Pictures of facial affect. Consulting Psychologists Press (1976)
17. 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., et al.: Universals and cultural differences in the judgments of facial expressions of emotion. Journal of personality and social psychology **53**(4), 712 (1987)
18. EQUATOR: Enhancing the quality and transparency of health research (2014), https://equator-network.org
19. García-Martínez, B., Martinez-Rodrigo, A., Alcaraz, R., Fernández-Caballero, A.: A review on nonlinear methods using electroencephalographic recordings for emotion recognition. IEEE Transactions on Affective Computing (2019)
20. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. arXiv preprint arXiv:1406.2661 (2014)
21. Green, D.M., Swets, J.A., et al.: Signal detection theory and psychophysics, vol. 1. Wiley New York (1966)

22. Hamada, M., Zaidan, B., Zaidan, A.: A systematic review for human eeg brain signals based emotion classification, feature extraction, brain condition, group comparison. Journal of medical systems **42**(9), 162 (2018)
23. Hamdi, H., Richard, P., Suteau, A., Allain, P.: Emotion assessment for affective computing based on physiological responses. In: 2012 IEEE international conference on fuzzy systems. pp. 1–8. IEEE (2012)
24. Higgins, J., Lasserson, T., Chandler, J., Tovey, D., Thomas, J., Flemyng, E., Churchill, R.: Methodological expectations of cochrane intervention reviews. London: Cochrane **6** (2019)
25. Higgins, J.P., Thomas, J., Chandler, J., et al.: Cochrane handbook for systematic reviews of interventions. John Wiley & Sons (2019)
26. Jemioło, P., Giżycka, B., Nalepa, G.J.: Prototypes of arcade games enabling affective interaction. In: International Conference on Artificial Intelligence and Soft Computing. pp. 553–563. Springer (2019)
27. Jemioło, P., Storman, D.: Quality assessment of systematic reviews (qasr) (Jun 2020), https://osf.io/dhtw3/
28. Jemioło, P., Giżycka, B., Storman, D.: Datasets for affect elicitation in emotion recognition (2020), https://osf.io/vdbqg/
29. Jerritta, S., Murugappan, M., Nagarajan, R., Wan, K.: Physiological signals based human emotion recognition: a review. In: 2011 IEEE 7th International Colloquium on Signal Processing and its Applications. pp. 410–415. IEEE (2011)
30. Kapur, A., Kapur, A., Virji-Babul, N., Tzanetakis, G., Driessen, P.F.: Gesture-based affective computing on motion capture data. In: International conference on affective computing and intelligent interaction. pp. 1–7. Springer (2005)
31. Khalil, R.A., Jones, E., Babar, M.I., Jan, T., Zafar, M.H., Alhussain, T.: Speech emotion recognition using deep learning techniques: A review. IEEE Access **7**, 117327–117345 (2019)
32. Khosla, A., Khandnor, P., Chand, T.: A comparative analysis of signal processing and classification methods for different applications based on eeg signals. Biocybernetics and Biomedical Engineering (2020)
33. Kory, J.M., D'Mello, S.K.: Affect elicitation for affective computing. The Oxford handbook of affective computing p. 371 (2015)
34. Kurdi, B., Lozano, S., Banaji, M.R.: Introducing the open affective standardized image set (oasis). Behavior research methods **49**(2), 457–470 (2017)
35. Kutt, K., Drazyk, D., Jemioło, P., Bobek, S., Giżycka, B., Rodriguez-Fernandez, V., Nalepa, G.: Biraffe: Bio-reactions and faces for emotion-based personalization. CEUR Workshop Proceedings (2019)
36. Lang, P.J., Bradley, M.M., Cuthbert, B.N., et al.: International affective picture system (iaps): Technical manual and affective ratings. NIMH Center for the Study of Emotion and Attention **1**, 39–58 (1997)
37. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H., Hawk, S.T., Van Knippenberg, A.: Presentation and validation of the radboud faces database. Cognition and emotion **24**(8), 1377–1388 (2010)
38. Liang, Y., Hsieh, S., Weng, C., Sun, C.: Taiwan corpora of chinese emotions and relevant psychophysiological data – standard chinese emotional film clips database. Chinese Journal of Psychology **55**(4), 597–617 (2013)
39. Liberati, A., Altman, D., Tetzlaff, J., Mulrow, C., Gøtzsche, P., Ioannidis, J., et al.: The prisma statement for reporting systematic and meta-analyses of studies that evaluate interventions. PLoS Medicine **6**(7), 1–28 (2009)
40. Lu, B., Hui, M., Yu-Xia, H.: The development of native chinese affective picture system – a pretest in 46 college students. Chinese Mental Health Journal (2005)

41. Luo, Y., Cai, X., Zhang, Y., Xu, J., Yuan, X.: Multivariate time series imputation with generative adversarial networks. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. pp. 1603–1614 (2018)
42. Lyons, M.J., Akamatsu, S., Kamachi, M., Gyoba, J., Budynek, J.: The japanese female facial expression (jaffe) database. In: Proceedings of third international conference on automatic face and gesture recognition. pp. 14–16 (1998)
43. Marchewka, A., Żurawski, Ł., Jednoróg, K., Grabowska, A.: The nencki affective picture system: Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. Behavior research methods **46**(2), 596–610 (2014)
44. Maria, E., Matthias, L., Sten, H.: Emotion recognition from physiological signal analysis: a review. Notes in Theoretical Computer Science **343**, 35–55 (2019)
45. Mehrabian, A.: Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Current Psychology **14**(4), 261–292 (1996)
46. Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G.: Prisma 2009 flow diagram. The PRISMA statement **6**, 97 (2009)
47. Moors, A.: Theories of emotion causation: A review. Cognition and emotion **23**(4), 625–662 (2009)
48. Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A.: Rayyan—a web and mobile app for systematic reviews. Systematic reviews **5**(1), 210 (2016)
49. Pallavicini, F., Ferrari, A., Pepe, A., Garcea, G., Zanacchi, A., Mantovani, F.: Effectiveness of virtual reality survival horror games for the emotional elicitation: Preliminary insights using resident evil 7: Biohazard. In: International Conference on Universal Access in Human-Computer Interaction. pp. 87–101. Springer (2018)
50. Peng, K.C., Chen, T., Sadovnik, A., Gallagher, A.C.: A mixed bag of emotions: Model, predict, and transfer emotion distributions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 860–868 (2015)
51. Picard, R.W.: Affective Computing. MIT Press (1997)
52. Pollock, M., Fernandes, R.M., Becker, L.A., Pieper, D., Hartling, L.: Chapter v: overviews of reviews. Cochrane Handbook for Systematic Reviews of Interventions version **6** (2018)
53. Redondo, J., Fraga, I., Padrón, I., Comesaña, M.: The spanish adaptation of anew. Behavior research methods **39**(3), 600–605 (2007)
54. Russell, J.A., Barrett, L.F.: Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. Journal of personality and social psychology **76**(5), 805 (1999)
55. Sarma, P., Barma, S.: Review on stimuli presentation for affect analysis based on eeg. IEEE Access **8**, 51991–52009 (2020)
56. Schaefer, A., Nils, F., Sanchez, X., Philippot, P.: Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. Cognition and Emotion **24**(7), 1153–1172 (2010)
57. Schmidt, P., Reiss, A., Dürichen, R., Laerhoven, K.V.: Wearable-based affect recognition – a review. Sensors **19**, 4079 (2019)
58. Schmidtke, D.S., Schröder, T., Jacobs, A.M., Conrad, M.: Angst: Affective norms for german sentiment terms, derived from the affective norms for english words. Behavior research methods **46**(4), 1108–1118 (2014)
59. Shea, B.J., Reeves, B.C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., et al.: Amstar 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. bmj **358**, j4008 (2017)

60. Shoumy, N.J., Ang, L.M., Seng, K.P., et al.: Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals. Journal of Network and Computer Applications **149**, 102447 (2020)
61. Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X., Yang, X.: A review of emotion recognition using physiological signals. Sensors **18**(7), 2074 (2018)
62. Smith, V., Devane, D., Begley, C.M., Clarke, M.: Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC medical research methodology **11**(1), 15 (2011)
63. Soleymani, M., Aljanaki, A., Yang, Y.: Deam: Mediaeval database for emotional analysis in music (2016)
64. Spezialetti, M., Cinque, L., Tavares, J.M.R., Placidi, G.: Towards eeg-based bci driven by emotions for addressing bci-illiteracy: a meta-analytic review. Behaviour & Information Technology **37**(8), 855–871 (2018)
65. Szwoch, W.: Using physiological signals for emotion recognition. In: International Conference on Human System Interactions (HSI). pp. 556–561. IEEE (2013)
66. Tandle, A.L., Joshi, M.S., Dharmadhikari, A.S., Jaiswal, S.V.: Mental state and emotion detection from musically stimulated eeg. Brain informatics **5**(2), 14 (2018)
67. Thanapattheerakul, T., Mao, K., Amoranto, J., Chan, J.H.: Emotion in a century: A review of emotion recognition. In: Proceedings of the 10th International Conference on Advances in Information Technology. pp. 1–8 (2018)
68. Valenza, G., Citi, L., Lanata, A., Scilingo, E.P., Barbieri, R.: A nonlinear heartbeat dynamics model approach for personalized emotion recognition. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 2579–2582. IEEE (2013)
69. Van Eck, N., Waltman, L.: Software survey: Vosviewer, a computer program for bibliometric mapping. scientometrics **84**(2), 523–538 (2010)
70. Witkowski, T.: Is the glass half empty or half full? latest results in the replication crisis in psychology. Skeptical Inquirer **43**(2), 5–6 (2019)
71. Yang, W., Makita, K., Nakao, T., Kanayama, N., Machizawa, M.G., Sasaoka, T., Sugata, A., Kobayashi, R., Hiramoto, R., Yamawaki, S., et al.: Affective auditory stimulus database: An expanded version of the international affective digitized sounds (iads-e). Behavior research methods **50**(4), 1415–1429 (2018)
72. Zhang, Q., Chen, X., Zhan, Q., Yang, T., Xia, S.: Respiration-based emotion recognition with deep learning. Computers in Industry **92**, 84–90 (2017)
73. Zhao, Y., Zhao, W., Jin, C., Chen, Z.: A review on eeg based emotion classification. In: 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). vol. 1, pp. 1959–1963. IEEE (2019)
74. Zhou, F., Qu, X., Jiao, J., Helander, M.G.: Emotion prediction from physiological signals: A comparison study between visual and auditory elicitors. Interacting with computers **26**(3), 285–302 (2014)
75. Zupan, B., Babbage, D.R.: Film clips and narrative text as subjective emotion elicitation techniques. The Journal of social psychology **157**(2), 194–210 (2017)