An overview of common emotion models in computer systems

M. Horvat*, A. Stojanović** and Ž. Kovačević**
* University of Zagreb, Faculty of Electrical Engineering and Computing, Department of Applied Computing, Zagreb, Croatia
** Zagreb University of Applied Sciences, Department of Computer Science and Information Technology, Zagreb, Croatia
marko.horvat3@fer.hr, aleksandar.stojanovic@tvz.hr, zeljko.kovacevic@tvz.hr

Abstract **– Emotions are an omnipresent and important factor in the interaction and communication between people. Since emotions are an indispensable part of human life, it would accelerate the progress of artificial intelligence and other fields of science that require data about emotions if they could be adequately described by computer systems. Today there are many different theories of affect, but few of them are used in affective computing. Other areas of computing also benefit from structured and expressive data models of the affective domain, such as human-computer interaction and brain-computer interfaces. Typical tasks include automated recognition and analysis of emotional states, mental fatigue, individual motivation, vigilance and stress resilience. In this paper four often used models of emotion and cognitive behavior are listed and their properties explained: discrete, dimensional, appraisal and action tendency models. For each model, algorithms are provided for similarity measures that can be used to determine the relatedness between different stimulation and estimation artefacts in their respective emotion spaces. The goal of this article is to help professionals find the optimal emotion model for their research and quickly become familiar with data modelling of affective states.**

Keywords - affective computing, emotion, data model, emotion stimulation, emotion estimation

I. Introduction

Emotions are ubiquitous and indispensable for communication, including the expression of personal opinions, moods, or reactions in social networks, e-mails, video conferencing, instant messaging, voice communication, etc. But in practice, the computer models used to capture this complex wealth of data are simplistic. For example, attitudes toward digital content are typically expressed with upvoting and downvoting (e.g., "like" and "dislike") or with a small, constant set of specific emoticons. Although these methods are inadequate and cannot convey all the important affective data. Fortunately, there are semantically well-defined and expressive emotion models. Although deeply rooted in complex psychophysiological theories, they can be easily used to annotate media and other digital content thanks to already developed and standardized metaformats.

This paper provides a compact yet comprehensive overview of common contemporary emotion models and annotation metaformats in computer systems. The paper intends to serve researchers as a concise roadmap in the complex problem of tagging digital content with affective information.

Affectively annotated multimedia purposely developed for regulated elicitation of emotional states represents a special type of digital content. Because of the purpose of their usage, these multimedia files are often referred to as stimuli and they are stored in affective multimedia databases [1]. In addition to the study of human emotion mechanisms, generation and appraisal of emotions, such databases have many other practical applications in the study of perception, memory, attention, and reasoning [1].

The remainder of this paper is organized as follows; Section 2 systematically describes emotion models that may be employed in computer systems. The section is divided into separate subsections explaining in detail discrete, dimensional, and cognitive models with examples for illustration. Sections 3 provides information on existing computer metaformats for conveying affective information. Section 4 provides information about the W3C Emotion Markup Language as the most comprehensive of these metaformats. Finally, Section 5 concludes the paper and proposes guidelines for future development of emotion computer models.

II. Emotion models

To capture the complete meaning of media in computer systems in addition to labelling the semantic content, it is also necessary to describe media affective meaning. Therefore, it is essential to use emotion models to convey the affective meaning in a way that is readable by humans and computers.

It is very important to note that knowledge about emotions is inherently imperfect [2] [3]. The facts about the affective meaning of media are uncertain because of incomplete knowledge of the real world. This ignorance arises from the nature of the phenomenon of emotion itself, but also from the inability to accurately measure the emotion experienced. Therefore, any statement about the affective content of multimedia is subject to probabilistic judgement [2] [3].

Identify applicable sponsor/s here. If no sponsors, delete this text box.

Emotion models can be divided into three large groups:

1.Theories of discrete emotions

2.Dimensional theories of emotions

3.Cognitive theories of emotions

Most multimedia stimuli databases use dimensional models, then discrete models, while cognitive models are exclusively represented in research that indirectly uses data from multimedia stimuli databases – the documents in the databases are not labeled with cognitive models, but these models are used to reason about the content of the stimuli, that is, for reasoning about the connection between cognitive, behavioral, and emotional components. The temporal aspect is also important since a good computational model of emotion must explain both the rapid dynamics of some emotional reactions as well as the slower responses that follow cognitive decision-making processes.

Following the recommendations of the W3C consortium for ease of understanding [4], the terms "emotion" and "affect" are often used interchangeably as synonyms. However, it should be noted that in the professional literature these terms do not have the same meaning and affect includes phenomena such as moods, interpersonal stances, preferences and attitudes, affect dispositions, emergent emotions and others [5] [6].

A. Discrete models

Discrete or categorical theories of emotion claim that dimensional emotion models do not accurately reflect the neural systems underlying emotional responses. Instead, proponents of these theories propose that there are many emotions that are universal across cultures and have an evolutionary and biological basis [7]. Which discrete emotions, or emotion norms as they are also called, are included in these theories is controversial. Most supporters of discrete emotion theories agree that six primary emotions exist: happiness, sadness, surprise, anger, fear, and disgust.

A clear example of how a discrete model is used in computer systems is the NimStim set of facial expressions [8] as illustrated in Fig. 1.

Pictures in the NimStim set are designated with named identifications which contain human model ID, sex, discrete emotion and level of emotion expression. The level of emotional expression is represented as ordinal value. Pictures are tagged with one of the following emotional facial expressions: neutral, angry, disgust, surprise, sad, calm, happy, and afraid. Neutral and calm expressions are included in the set as comparison conditions for usage in different studies, particularly in neuroimaging. Calm expression is perceptually similar to neutral but is perceived as having a less negative polarity.

Another well-known everyday example of the discrete model are emoticons or emojis on social networks such as the Facebook. They are standardized and easy to use on the web as plug-ins. The one-dimensional discrete model with likes (“thumb up”) and dislikes (“thumb down”), available on Twitter and Reddit, may also be used in emoticon analysis as a measure of sentiment.

Figure 1. Examples of 6 discrete emotions represented in the NimStim picture database [8]. From top to bottom, and left to right: happiness, sadness, surprise, anger, fear, disgust.

Discrete models OCC categories [9], FSRE categories [10] and Frijda’s categories [11] are currently used only in psychological research of emotion and not for content tagging or estimation of emotion. However, specific categories from different models may be used as named entities to expand or modify the basic Big six model vocabulary for digital content tagging [12].

Because of the described properties of the discrete affect space cosine distance may be used to measure similarity between two discrete emotions a^{dis} and b^{dis} where each discrete emotion contains *n* components:

$$sim(a^{dis}, b^{dis}) = \cos(\Theta) = \frac{a^{dis} \cdot b^{dis}}{\|a^{dis}\| \|b^{dis}\|} = \frac{\sum_{i=1}^{n} a_i^{dis} b_i^{dis}}{\sqrt{\sum_{i=1}^{n} a_i^{dis}} \sqrt{\sum_{i=1}^{n} b_i^{dis}}} \quad (1)$$

1) Big six model

The so called “Big six” set of emotion norms (happiness or joy, sadness, surprise, anger, fear, disgust) is the most often used discrete model of emotion [12]. It is commonly employed in affective multimedia databases, emotion estimation software and in general for tagging any digital content with affective information. It was postulated by Ekman in his seminal paper [12].

2) OCC categories

Ortony, Clore and Collins as a part of their more encompassing cognitive appraisal model [9] have also proposed a set of 22 different emotion categories. In alphabetical order these are: admiration, anger, disappointment, distress, fear, fears-confirmed, gloating, gratification, gratitude, happy-for, hate, hope, joy, love, pity, pride, relief. remorse, reproach, resentment, satisfaction, and shame [9].

3) FSRE categories

Similarly to OCC categories, Fontaine, Scherer, Roesch and Ellsworth have suggested a wider reaching model encompassing different dimensionalities of emotion space [10]. In regard to discrete models, they have proposed a model with 24 categories. These are, alphabetically: anger, anxiety, being hurt, compassion, contempt, contentment, despair, disappointment, disgust, fear, guilt, happiness, hate, interest, irritation, jealousy, joy, love, pleasure, pride, sadness, shame, stress, and surprise [10].

4) Frijda's categories

In his action tendencies theory (see subsection II.C.2), Frijda has identified a set of 12 emotion categories which are related to action tendencies: anger, arrogance, desire, disgust, enjoyment, fear, humility, indifference, interest, resignation, shock, and surprise [11].

5) WordNet-Affect

The WordNet-Affect is a so-called affective extension of WordNet and includes a subset of synsets suitable to represent affective concepts correlated with affective words [13]. It was developed with the aim of studying the relation between natural language and affective information, and representation of such information as hierarchically organized and functionally related named concepts [13]. An excerpt from the WordNet-Affect synset hierarchy is shown in Fig. 2. In the WordNet synset is a set of synonymous words (i.e. synonyms) that express the same semantic concept [14]. The WordNet-Affect provides additional hierarchy of "affective labels" (a-labels), independent from the semantic domain hierarchy, with which the synsets representing affective concepts may be annotated. The a-labels are: positive, negative, ambiguous, and neutral. For example, synsets such as joy#1 and enthusiasm#1 are marked with the first a-label, anger#1 and sadness#1 with the second, surprise#1 because it's valence depends on semantic context with the third, and synsets referring to mental states that are generally considered affective but are not characterized by valence are marked as neutral [15].

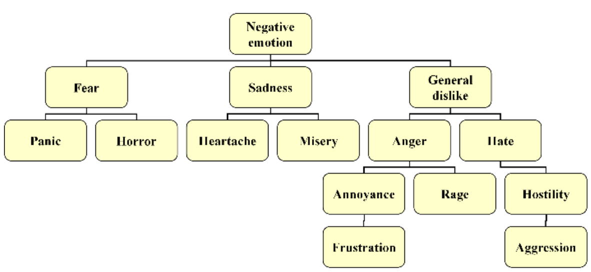

Figure 2. An excerpt from the WordNet-Affect semantic network hierarchy.

It has already been shown that WordNet may be successfully used for high-level representation of picture semantics and retrieval of emotionally-annotated pictures [16] [17]. With the WordNet-Affect it could be possible to categorically describe overall affective content of specific pictures or, for that matter, other multimedia formats.

B. Dimensional models

The dimensional model is simple yet efficient. In the literature it is also called the circumplex model of emotion, the PAD (Pleasure-Arousal-Dominance) model, or the Russell model of emotion [18] [19]. The dimensional model is the most often employed model of emotion for annotation of multimedia stimuli [1]. The underlying dimensional theory of emotion proposes that affective meaning can be well characterized by a small number of dimensions. Dimensions are chosen on their ability to statistically characterize subjective emotional ratings with the least number of dimensions possible [20]. In creating the model Russell estimated approximate central coordinates of specific discrete emotions in the dimensional model's space [18]. He hypothesized that these locations are not fixed but rather change during a person's lifetime, and also differ from one person to another, or between homogenous groups of persons based on their character traits.

The dimensional model is built around three emotion dimensions that are mutually orthogonal: valence (Val), arousal (Ar), and dominance (Dom). Positivity and negativity of a stimulus are specified by valence, while arousal describes the intensity or energy level, and dominance represents the controlling and dominant nature of the emotion. In practice dominance is frequently omitted from description of emotion space because it was shown to be the least informative measure of the elicited affect [19]. All three emotion dimensions are described with continuous variables normalized in interval [1, 9]: $val \in [1,9] \in Val$, $ar \in [1,9] \in Ar$, $dom \in [1,9] \in Dom$. In some instances emotion values are scaled and represented in percentages, or authors use a smaller Likert-scale. In any case, in the dimensional emotion model, a single emotionally-annotated picture can be projected onto the two-dimensional emotion space $\Omega = Val \times Ar$, as exemplified in Fig. 3, with each data point representing one picture from the Open Affective Standardized Image Set (OASIS) database containing 900 picture stimuli [21].

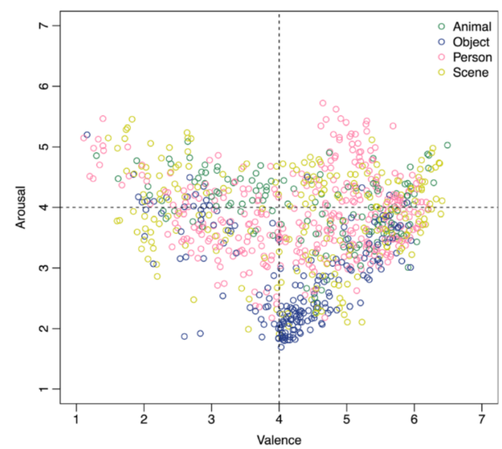

Figure 3. Picture stimuli from the OASIS dataset according to the dimensional emotion model (measured on a 1–7 Likert scale) with valence on the x-axis and arousal on the y-axis. The colors denote different image categories. Adapted from [21].

Euclidian distance metrics is commonly used to measure similarity between two emotions in the dimensional space. Therefore, similarity between two dimensional emotions a^{dim} and b^{dim} with each emotion vector containing components of valence, arousal and dominance is:

$$\begin{gathered} sim(a^{dim}, b^{dim}) = 1 - dist(a^{dim}, b^{dim}) \\ dist(a^{dim}, b^{dim}) = \\ = \sqrt{\left(a_{val}^{dim} - b_{val}^{dim}\right)^2 + (a_{ar}^{dim} - b_{ar}^{dim})^2 + \left(a_{dom}^{dim} - b_{dom}^{dim}\right)^2} \end{gathered} \quad (2)$$

C. Cognitive models

Cognitive models are fundamentally different from discrete and dimensional models of emotion. All cognitive models share the common goal of defining a general computational model of the mechanisms underlying human emotions [22]. In the context of computer systems they may be used to aid in the development of human-like autonomous agents that must mediate the interaction between emotion and cognition as core aspects of human behavior [23]. Examples of applications include a range of specific behaviors that must be modeled by virtual humans, such as facial expressions, dialog delivery, planning, reacting, and social understanding. Also, to model strategic decision- making, action selection, facial animation, and social intelligence. Cognitive models can be generally divided in two groups: cognitive appraisal and action tendency models.

1) Appraisal models

A central claim of any appraisal theory is that emotions are both elicited and distinguished on the basis of a person's subjective evaluation of the personal significance of a situation, object, or event [24]. Additionally, authors have suggested that the nature of an emotional response is best predicted on the basis of the person's subjective evaluation of a previous personally significant event [24]. Moreover, in various emotion stimulation experiments the process of appraisal and reappraisal have been shown to provide an explanation for occurrence of stress related mental disorders.

It is important to note that contemporary appraisal theories define emotions as processes rather than states. This is reflected in the fact that the term emotion is often used as a synonym for an emotional episode [25]. Appraisal theories are also referred to as component theories because they view an emotional episode as a change in a set of organic subsystems or components. The subsystems are viewed in an abstract manner but can be mapped to different regions of the central nervous system. The components include an appraisal component with evaluations of the environment and person-environment interaction, a motivational component with action tendencies or other forms of readiness to act, a somatic component with peripheral physiological responses, a motor component with expressive and instrumental behavior, and an emotional component with subjective experiences or feelings. The emotion process is continuous and recursive. Changes in one component affect other components as well. For example, changes in appraisal can lead to changes in physiological and behavioral responses. These, in turn, may lead directly or indirectly (via a change in the stimulus situation) to changes in appraisal. As a result, multiple emotional episodes may occur in parallel.

The key differences between appraisal theories and other theories of emotion include: *i*) the definition of appraisal, both in terms of content and nature of the process, *ii*) the role of appraisal in emotions and predictions about the relationship between changes in appraisal and changes in other components, and *iii*) predictions about individual, cultural, and developmental differences [27]. A schema of an appraisal model from Scherer connecting emotion stimulus, emotion response, cognitive appraisal and action tendencies with primary and secondary appraisal mechanisms is provided in Fig. 4.

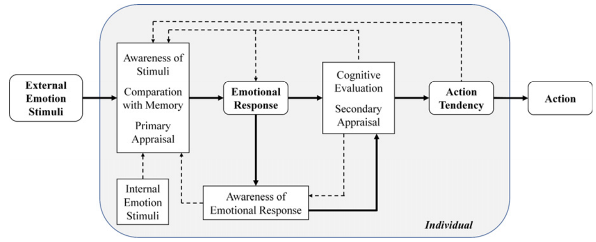

Figure 4. Model of individual's recursive effects between emotion stimulus, emotion response, cognitive appraisal and action tendencies adapted from [26].

Common appraisal models, which are supported in the EmotionML schema, are: model of emotions cognitive structure, also called OCC model [9], emotion component process model [24], and domain-independent framework for modeling emotion also known as EMA appraisals [23].

2) Action tendency models

The group of action tendency models is represented by only one vocabulary system from Frijda [11]. By definition, action tendency is an urge to carry out certain behaviors that are linked to a specific emotion [28]. For example, the action tendency of fear or worry involves an urge to flee, and that of anger or hostility incorporates an urge to attack or fight. The link between emotion and urge to perform action has been apprehended since the antiquity [29]. However, the current main point of contention among emotion theorists is whether physical activities that elicit explicit emotional behavior can be characterized according to one of two main perspectives, namely, the discrete action program perspective and the motivational basis perspective. The first states that there are discrete action or affect programs with specific profiles of the autonomic nervous system (ANS) that are explored and stereotyped by neuroscientists. The second perspective assumes that physical activity is expressed in relatively nonspecific ANS states whose influence on behavior is structurally and functionally based on and constrained by motivational and environmental contextual factors. Both perspectives argue that the action tendency of an emotional response should be considered as its essential defining feature [29]. In the context of emotion models in computer systems, Frijda [30] suggests that differences in autonomic activity may represent contingent patterns of action readiness. In this view, many different behaviors may manifest following a particular emotion-relevant event. For example, a threat stimulus associated with a state of anxiety may trigger fight or flight tendencies. Moreover, a given behavior may be triggered by different emotional events, although its manifestation may alternately be due to activation of the appetitive or defensive systems. In [30] a hierarchy of general appetitive and defensive behaviors was identified that provide elaborated states of action tendency. The action tendency terms defined in the EmotionML are ready to be used for annotation of media and other content in digital systems [4]. The terms with their related discrete emotional states in parenthesis are: approach (desire), avoidance (fear), being-with (enjoyment), attending (interest), rejecting (disgust), nonattending (indifference), agonistic (anger), interrupting (shock and surprise), dominating (arrogance), and submitting (humility and resignation).

III. Metaformats for Affective Annotations

Currently several standardized meta formats for tagging emotions exist. These are: Emotion Markup Language (EmotionML) [4], Synchronized Multimedia Integration (SMIL) [31], Speech Synthesis Markup Language (SSML) [32], Extensible MultiModal Annotation Markup Language (EMMA) [33], Emotion Annotation and Representation Language (EARL) [34], and Virtual Human Markup Language (VHML) [35].

All emotion description metaformats are stored in formatted text files that are used to store emotion data in other documents. The tagged media content can be in any digital format. None of the metaformats are based on higher abstraction schemes such as formal logic.

The description of emotions in standardized metaformats usually includes a set of data, such as the beginning, end, intensity, and category of the emotion, or the dimensions of the emotion, such as the level of comfort, arousal, and dominance. It is also possible to specify different affective terms in the multimedia data, such as appraisal and action tendency. In addition to describing emotions, metaformats can also define data about the associated digital document, such as its URI, multimedia format, time of creation, recent changes, information about the owner, and more.

Metaformats for describing emotions are the simplest and most widely used formalisms for defining emotional content in computer systems. They are readable by humans and computers alike, have well-defined semantics, and allow information to be exchanged between different systems. However, none of the metaformats have complex syntax and do not inherently support reasoning and derivation of new knowledge from existing knowledge. To enable these services, an intelligent system must be constructed that uses one of these formalisms for data exchange and storage. Of the metaformats mentioned above, EmotionML, described in the next section, is the most comprehensive.

IV. EmotionML

The W3C Emotion Markup Language (EmotionML) is currently the most expressive and complex declarative computer language for describing emotions in digital content and multimedia [4]. It has been developed as a plug-in language for use in three distinct application domains: 1) manual data annotation, 2) automated emotion estimation, and 3) management of computer systems with emotions. EmotionML provides users with the ability to select and combine different annotation types of various aspects of emotional behavior such as discrete emotions, emotion dimensions, appraisals, and action tendencies. EmotionML allows users to select the dictionary of emotions to be used in a particular application or to expand the corpus with a new custom dictionary.

EmotionML is a markup language based on XML/XML Schema notation and does not have constructs of imperative or procedural computer languages. It can be easily interpreted by a computer, but since it is not based on logic reasoning about emotions using EmotionML is not directly supported. This requires, for example, the transformation of records into more complex ontologically supported models of knowledge representation [3].

EmotionML distinguishes four types of applications or relationships of emotions (so called *emotion references*) within a specific the application domain [4]:

- expression (*expressedBy*): perceptual behavior or expression (e.g., physiology, facial expressions, etc.) that expresses emotional states.
- experience (*experiencedBy*): subject who "has", i.e. feels, emotion.
- cause (*triggeredBy*): an event that caused an emotion that caused an emotional reaction.
- goal (*targetedAt*): an object to which an emotional reaction or action is directed. In EmotionML it is allowed to use URIs or references to parts of video and audio files to reference annotated objects.

An example how affective data in pictures can be annotated with the EmotionML markup is shown in Fig. 5. Here the top-left picture from Fig. 1. (ID in the NimStim 01F_HA_C.jpg) is tagged with the Big Six vocabulary.

```
<emotionml xmlns="http://www.w3.org/2009/10/emotionml"
           xmlns:meta="http://www.example.com/metadata"
           category-set=
           "http://www.w3.org/TR/emotion-voc/xml#big6">
    <info>
        <meta:media-type>image</meta:media-type>
        <meta:media-id>01F_HA_C.jpg</meta:media-name>
        <meta:media-set>NimStim set of facial expressions
        </meta:media-set>
        <meta:doc>Example adapted from (Tottenham et al. 2009)
        https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3474329/pdf
        /nihms408635.pdf</meta:doc>
    </info>
    <emotion>
        <category name="happiness" value="0.978723404"/>
    </emotion>
</emotionml>
```

Figure 5. Discrete emotion data of the top-left NimStim picture 01F_HA_C.jpg in Fig. 1 annotated with EmotionML syntax.

Further information with examples of use is provided in the W3C standards definition [36].

V. Conclusion and Future Work

The ability to efficiently describe, store, and use reliable information about emotions is important in computer systems. Some of the applications that can benefit most from such information are multimedia repositories, text sentiment analysis tools, search engines, social networks, emotion recognition software, etc. The importance of the ability of computer systems to successfully manage affective data will certainly increase in the future.

However, much work remains to be done. The most important future work is to enable the functional integration of various emotion formats and the transformation of information between different emotion models. In addition, new software tools need to be developed for automated annotation of content and retrieval of information using these models. It is also vital to increase the semantic expressiveness of metaformats and to introduce either formal or probabilistic reasoning. For the introduction of formal knowledge representation methods, it is necessary to use computer ontologies and develop a general ontology of emotions and many different domain ontologies for specific applications [37] [38]. To support probabilistic inference, it is necessary to integrate differently structured affective datasets and increase their information quality.

The described area of emotion phenomena research is large and complex. However, emotion theories when combined with digital metaformats can be successfully used for design of computer systems. This paper provides interested professionals with valuable and structured content, and carefully selected references, as a steppingstone for further exploration of this important and multifaceted field of study.

REFERENCES

[1] M. Horvat, "A Brief Overview of Affective Multimedia Databases," In Central European Conference on Information and Intelligent Systems (CECIIS 2017), pp. 3−11, 2017.

[2] R. W. Picard, "Affective computing: challenges", International Journal of Human-Computer Studies, vol. 59, no. 1, pp. 55−64, 2003.

[3] M. Horvat, Generation of multimedia stimuli based on ontological, affective and semantic annotation (Doctoral thesis), University of Zagreb, 2013.

[4] F. Burkhardt, C. Pelachaud, B. W. Schuller, and E. Zovato, "EmotionML," In Multimodal interaction with W3C standards, Springer, Cham, pp. 65−80, 2017.

[5] K. R. Scherer, "What are emotions? And how can they be measured?", Social science information, vol. 44, no. 4, pp. 695–729, 2005.

[6] P. Wilhelm and D. Schoebi, "Assessing mood in daily life: Structural validity, sensitivity to change, and reliability of a short-scale to measure three basic dimensions of mood", *European Journal of Psychological Assessment*, vol. 23, no. 4, pp. 258–267, 2007.

[7] P. Ekman, "Are there basic emotions?", *Psychological Review*, vol 99, pp. 550–553, 1992.

[8] N. Tottenham, J. W. Tanaka, A. C. Leon, T. McCarry, M. Nurse, T. A. Hare, D. J. Marcus, A. Westerlund, B. J. Casey, and C. Nelson, "The NimStim set of facial expressions: judgments from untrained research participants," *Psychiatry research*, vol. 168, no. 3, pp. 242–249, 2009.

[9] A. Ortony, G. L. Clore, and A. Collins, "The cognitive structure of emotions," Cambridge university press, 1990.

[10] J. R. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth, "The World of Emotions Is Not Two-Dimensional," *Psychological Science*, vol. 18, no. 12, pp. 1050–1057, 2007.

[11] N. H. Frijda, "The Emotions," Cambridge, UK: Cambridge University Press, 1986.

[12] P. Ekman, "Facial expressions of emotion: New findings, new questions," 1992.

[13] C. Strapparava and A. Valitutti, "Wordnet affect: an affective extension of wordnet," In *Lrec*, vol. 4, no. 1083–1086, p. 40, 2004.

[14] G. A. Miller, "WordNet: a lexical database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[15] C. Strapparava, A. Valitutti, and O. Stock, "The affective weight of lexicon," In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, 1–83, 2006.

[16] M. Horvat, A. Grbin, and G. Gledec, "Labeling and retrieval of emotionally-annotated images using WordNet," *International Journal of Knowledge-based and Intelligent Engineering Systems*, vol. 17, no. 2, pp. 157–166, 2013.

[17] M. Horvat, D. Duvnjak, and D. Jug, "GWAT: The Geneva Affective Picture Database WordNet Annotation Tool," arXiv preprint arXiv:1505.07395, 2015.

[18] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and psychopathology*, vol. 17, no. 3, p. 715, 2005.

[19] I. Bakker, T. van der Voordt, P. Vink, and J. de Boon, "Pleasure, arousal, dominance: Mehrabian and Russell revisited," *Curr. Psychol*, vol. 33, pp. 405–421, 2014.

[20] M. M. Bradley and P. J. Lang, "Measuring emotion: The Self-Assessment Manikin and the semantic differential," *Journal of Behavior Therapy & Experimental Psychiatry*, vol. 25, pp. 49–59, 1994.

[21] B. Kurdi, S. Lozano, and M. R. Banaji, "Introducing the open affective standardized image set (OASIS)," *Behavior research methods*, vol. 49, no. 2, pp. 457–470, 2017.

[22] J. Gratch and S. Marsella, "A domain-independent framework for modeling emotion," *Cognitive Systems Research*, vol. 5, no. 4, pp. 269–306, 2004.

[23] S. C. Marsella and J. Gratch, "EMA: A process model of appraisal dynamics," *Cognitive Systems Research*, vol. 10, no. 1, pp. 70–90, 2009.

[24] K. R. Scherer, "Appraisal theory," 1999.

[25] K. R. Scherer, "Emotions are emergent processes: they require a dynamic computational architecture," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3459–3474, 2009.

[26] H. C. Traue, et al. "A framework for emotions and dispositions in man-companion interaction," *Converbal Synchrony in Human-Machine Interaction*, pp. 99–140, 2013.

[27] A. Moors, P. C. Ellsworth, K. R. Scherer, and N. H. Frijda, "Appraisal theories of emotion: State of the art and future development," *Emotion Review*, vol. 5, no. 2, pp. 119–124, 2013.

[28] APA Dictionary of Psychology, https://dictionary.apa.org/action-tendency, last accessed Feb 2022.

[29] R. Lowe, "The feeling of action tendencies: on the emotional regulation of goal-directed behavior," *Frontiers in psychology*, vol. 2, p. 346, 2011.

[30] N. H. Frijda, "Impulsive action and motivation," *Biological psychology*, vol. 84, no. 3, pp. 570–579, 2010.

[31] D. C. Bulterman, "SMIL: Synchronized multimedia integration language," In *MediaSync*, Springer, Cham, pp. 359–385, 2018.

[32] P. Taylor and A. Isard, "SSML: A speech synthesis markup language," *Speech communication*, vol. 21, no. 1–2, pp. 123–133, 1997.

[33] M. Johnston, "Extensible multimodal annotation for intelligent interactive systems," In *Multimodal Interaction with W3C Standards*, Springer, Cham, pp. 37–64, 2017.

[34] M. Schröder, L. Devillers, K. Karpouzis, J-C. Martin, C. Pelachaud, C. Peter, H. Pirker, B. Schuller, J. Tao, and I. Wilson, "What should a generic emotion markup language be able to represent?", *Affective Computing and Intelligent Interaction, Lecture Notes in Computer Science,* vol. 4738, pp. 440–451, 2007.

[35] A. Marriott, "VHML–Virtual Human Markup Language," In *Talking Head Technology Workshop*, *at OzCHI Conference*, pp. 252–264, 2001.

[36] W3C Recommendation Emotion Markup Language (EmotionML) 1.0, https://www.w3.org/TR/emotionml/, last accessed Feb 2022.

[37] M. Horvat, N. Bogunović, and K. Ćosić, "STIMONT: a core ontology for multimedia stimuli description," *Multimedia tools and applications*, vol. 73, no. 3, pp. 1103−1127, 2014.

[38] M. Horvat, "StimSeqOnt: An ontology for formal description of multimedia stimuli sequences," In *MIPRO 2020 Proceedings of the 43rd International Convention, IEEE*, pp. 1378−1383, May 2020.