

Role of Contextual Knowledge in Paralinguistic Affect Recognition by Humans

Rôle de la connaissance contextuelle dans la reconnaissance des effets paralinguistiques

Andreas H. Marpaung¹, Avelino J. Gonzalez²

¹ University of Central Florida, Florida, USA, amarpaung@knights.ucf.edu

² University of Central Florida, Florida, USA, avelino.gonzalez@knights.ucf.edu

ABSTRACT. It has long been thought by cognitive science researchers that context plays a large role in human cognition – from memory recall to linguistics and problem solving. However, the role of context in identifying the emotional state of a human interlocutor, while thought to be important, has never been evaluated experimentally. This paper describes a study involving human test subjects that sought to gauge how well they could identify the emotion being expressed by a speaker using only paralinguistic signals (i.e., speech without understanding the spoken words), with and without knowing the speaker's context. The first part of the study entailed asking the test subjects to identify the emotion expressed by a speaker who utters unintelligible sounds in a context-free basis. The test subjects were provided with knowledge about the context in which the speaker(s) uttered the same sounds heard before on a context-free basis. This allowed for the measurement of the impact of knowing the context upon their ability to correctly identify the expressed emotion. The results of our study indicate that knowing the context under which the speakers uttered the expressions indeed improved the ability of the test subjects to infer the correct emotion being expressed by the speakers. This paper describes the study and its results.

RÉSUMÉ. Les chercheurs en sciences cognitives ont longtemps pensé que le contexte joue un rôle important dans la cognition humaine - du rappel de la mémoire à la linguistique et à la résolution de problèmes. Cependant, le rôle du contexte dans l'identification de l'état émotionnel d'un interlocuteur humain, bien que considéré comme important, n'a jamais été évalué expérimentalement. Cet article décrit une étude impliquant des sujets de test humains qui cherchaient à évaluer dans quelle mesure ils pouvaient identifier l'émotion exprimée par un locuteur en utilisant uniquement des signaux paralinguistiques (c'est-à-dire, la parole sans comprendre les mots prononcés), avec et sans connaître le contexte du locuteur. La première partie de l'étude a consisté à demander aux sujets du test d'identifier l'émotion exprimée par un locuteur qui émet des sons inintelligibles sans contexte. Les sujets de l'étude ont reçu des connaissances sur le contexte dans lequel le (s) locuteur (s) ont émis les mêmes sons entendus auparavant sur une base sans contexte. Cela a permis de mesurer l'impact de la connaissance du contexte sur leur capacité à identifier correctement l'émotion exprimée. Les résultats de notre étude indiquent que la connaissance du contexte dans lequel les locuteurs ont prononcé les expressions a en effet amélioré la capacité des sujets de test à déduire l'émotion correcte exprimée par les locuteurs. Cet article décrit l'étude et ses résultats.

KEYWORDS. affect recognition, paralinguistic, acoustic, affective computing, contextual modelling, contextual knowledge.

MOTS-CLÉS. reconnaissance d'affect, paralinguistique, acoustique, calcul affectif, modélisation contextuelle, connaissance contextuelle.

1. Introduction

Can a system that automatically recognizes a speaker's emotions be context-independent? In 1995 Picard answered this question unequivocally when she said that to be context-independent or not is not a questionable statement in the Affective Computing domain [1]. Many researchers have agreed with her on that point, stressing that we need to integrate contextual information into our traditional affect recognition approach [2][3]. In this paper we describe a study that supports Picard's statement.

Over the past three decades, research in automated affect recognition has shifted from a single modality to multimodal approaches [4][5][6]. As deep learning algorithms gain more acceptance

and wider use, researchers have employed them for recognizing facial expression [7][8] and physiological signals [9]. Recent publications about multimodal research with deep learning include [10][11][12][13]. Nevertheless, because the expression of emotion is not produced in a vacuum and has some causal chains, we hypothesize that contextual information affects a human's expression of emotion. To our knowledge, Kosti et al. [14] have been the only researchers who have integrated contextual information in their affect recognition algorithm from facial expressions.

Speech has become an essential means to share opinions, ideas, observations, and feelings in human-to-human interaction, and conveys emotional content in two ways: (1) the linguistic aspect of speech (what words are said), (2) and the paralinguistic aspect of speech (how are the words acoustically expressed). This article describes our human study to understand the effects of contextual knowledge on a human's ability to recognize the emotions expressed by a speaker when only considering paralinguistic speech signals. We sought not only to understand but also to determine which type of context could better help a person infer the emotional state of a conversant when paralinguistic information is the only other signal available to her/him. We focus on paralinguistic signals because automatically inferring a speaker's emotions has proven more difficult with paralinguistic than with linguistic signals. Furthermore, there are several potential applications, such as when the other person is too far away to understand the words, the conversant does not speak our language, or the conversant speaks our language very poorly, and with a thick accent.

The structure of this paper is as follows: Section 2 focuses on the importance of contextual knowledge from as the standpoint of human psychology, linguistics, memory recall, problem solving and other functions that humans perform well. Section 3 describes our human study and the data used in this study. Section 4 describes the study results, and finally, section 5 concludes the work and suggests future research.

2. Why Context?

Over the past 50 years, research has shown that “reasoning with and about context is an essential aspect of human cognition, permeating language, memory, and reasoning capabilities.” [15]. In human cognition, the *reasoning context* (derived from the environment) and the *perceived context* (recognized in the environment), give humans the ability to think both logically and critically when problem solving, as well as to comprehend different components of a problem to derive its overall solution [16]. In language, contextual processing is utilized in linguistic comprehension, both in written and spoken language as indicated by the surrounding words and the general topic of the conversation [17][18][19]. In the human memory system, context helps someone remember information and recognize what situations caused its retrieval (e.g., “This reminds me of ...”). Contextual processing facilitates the encoding, storage, and retrieval of some cognitive frameworks known as schemas [20][21]. Therefore, it is no surprise that it is considered easier for humans to interpret emotions expressed by other people in a non face-to-face interaction when the context of conversation is known [2]. In summary, the knowledge of context (the environment, observed objects, and/or the current task) is critical in social interaction in order to structure activities, to navigate the world around us, to organize information, to adapt to conditions, to convey ideas to one another and to react appropriately to situations.

3. Approach

The research described in this paper sought to answer two questions of interest. First, can adding contextual knowledge improve a human's affect recognition ability when only the paralinguistic aspect of speech is otherwise available? Secondly, if indeed contextual knowledge helps, what kind of contextual knowledge can more effectively improve a human's affect perception capability?

Inspired by other researchers on the benefit of contextual knowledge in disambiguating different emotions [22-25], we conducted a study to explore whether and how humans make use of context to infer the emotional state of a conversant when only paralinguistic information is available. Our hypothesis is that with the additional contextual cues, the accuracy of the test subjects in discerning the correct emotions expressed by the speaker/actors would improve progressively as the cues were separately and sequentially provided. The results of this study generally support our hypothesis and serve as evidence of the benefit of contextual cues in affect recognition.

The study involved providing audio recordings made by professional actors who uttered meaningless imaginary words (in no known language) in a way that its paralinguistic qualities expressed a pre-determined emotion. The test subjects were first asked to indicate the emotion that they thought was being conveyed by the utterances in the audio files strictly from the paralinguistic qualities on a context-free basis. Then, two contextual cues about the situation that provoked the expressions uttered were given separately and sequentially to the test subjects, and they were asked to once again determine the emotions that they thought were being communicated in the same recordings. One of the two contextual cues involved describing the actions being undertaken by the speaker/actor at the time of his/her recorded utterance, while the other involved knowing the relationship of the speaker/actor with whomever he/she was speaking, and/or a description of the environment. One of the cues was provided after the context-free query, and the second cue was provided after the query related to the first contextual cue. Therefore, after receiving the second contextual cue, the test subjects now knew about both contextual cues. The order of the provision of the two different cues was reversed for the next recording to that the value of each type of cue could be isolated. Our expectation was that the accuracy of the test subjects in selecting the correct emotion would progressively improve after each successive receipt of contextual information.

3.1. *Speech Material*

Audio files from the Geneva Multimodal Emotional Portrayal (GEMEP) database [26] were used for this study. These audio files include 10 (five male and five female) professional French-speaking actors who had previously worked in theatrical productions (age range: 25-57 years old; mean age of 37.1). These actors portrayed several affective states by saying one of these two pseudo speech sentences: “nekal ibam soud molen!” (equivalent to a declarative statement or an exclamation) or “koun se mina lod belam” (equivalent to a question) for each affective state. The use of pseudo-speech in this corpus eliminated any potential effects attributable to linguistic content.

In this study, we focused on analyzing Ekman’s six universal emotions: anger, disgust, fear, joy, sadness, and surprise [27]. With the exception of disgust and surprise, each emotion had ten audio files, spoken by all (five) male and (five) female actors (10 individual recordings in total for each emotion); disgust was expressed by only two males and two females (total of 4) while surprise was expressed by three males and two female actors (total of 5). Readers interested in the emotion elicitation technique used by the actors, the recording procedures and setups used, and the statistical analysis for the verification of the perceptual accuracy for the believability, authenticity, and plausibility can refer to [26] for details.

3.2. *Contextual Knowledge*

To induce and elicit the assigned emotions, the actors used the *felt enacting technique*. This technique is typically used in theatrical settings where the actors recall or have mental imagery of relevant personal experiences similar to those found on the written scripts given to them prior to the recordings in order to enact the desired emotion. The scripts describe the setting and other aspects of the environment in which the actors would utter the pseudo-speech “words”. In effect, these scripts described the context experienced by the actors, although in an imaginary way only. For our study, we translated these original scripts from French to English using Google Translate [28].

From these translated scripts, we manually extracted two pieces of contextual knowledge: (a) the *action context*, defined as any information about what happens to a person or a group of people involved in the conversation, and (b) the *relationship/environment (R/E) context*, defined as any information that relates an object to a person or relates a person to another person (or a group of people).

In our study, test subjects were first asked to listen to the audio recording and infer the emotion expressed by the speaker, without the benefit of any contextual information (i.e., context-free). Then, they were given one piece of contextual knowledge (action or R/E) and allowed to listen to the same recording again, after which they could reconsider and amend their prior choices if they so desired. Lastly, they were provided with the second piece of contextual knowledge (R/E or action) and were asked to again reconsider their choice (after once again listening to the same recording). These choices were then compared to the ground truth as provided by the GEMEP database creators [26]. The order in which the contextual cues were presented (i.e., action-then-R/E or R/E-then-action) was determined by a random function. Therefore, all emotions had two sequences each: (1) no-context→action→R/E, and (2) no-context→R/E→action. The complete list of the contextual knowledge that we had can be found in [29].

3.3. Data Collection

Fig. 1 shows our user interface, written in Microsoft Visual C++, used to collect the data from our human participants. The test setup was installed in only one laptop computer, so all test subjects had to use that same laptop. For audio purposes, a universal 3.5 mm earbud headset was connected to the laptop’s audio jack. Sixty volunteers were recruited for our study – 22 males 18 to 57 years old (average age of 26.4) and 38 females 17 to 54 years old (average age of 23.9). There was no time limit placed on the test subjects for completion of any step or of the entire process. All test subjects undertook their assignments in a quiet and comfortable office environment. A short training session on how to operate the GUI was given prior to their participation.

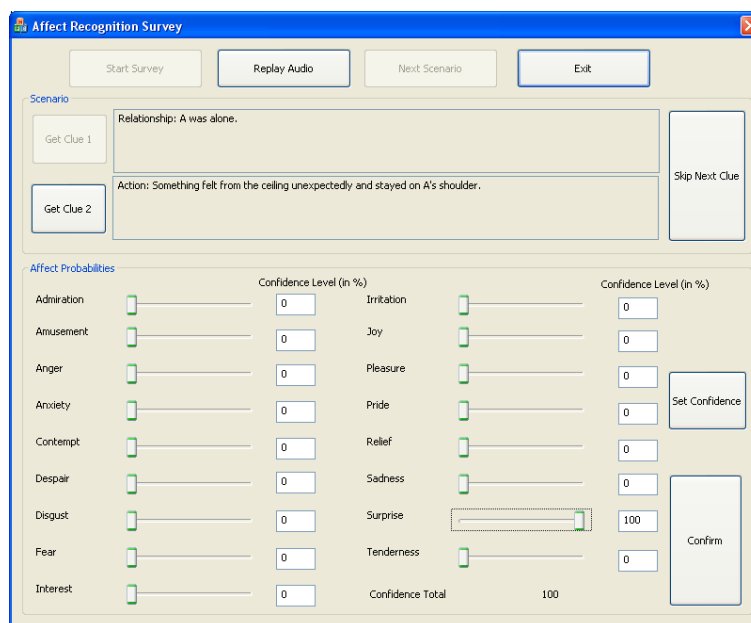


Fig. 1. Our Graphical User Interface (GUI)

The seventeen emotions listed in Fig. 1 were included in the GEMEP corpus and were presented to the test subjects as possible emotions expressed by the actors. However, the audio recordings actually presented to the test subjects (as well as the associated contextual cues) only involved Ekman’s six universal emotions (i.e., joy, sadness, surprise, disgust, fear and anger).

Each participant was subjected to twelve different randomly generated scenarios/recordings. Two recordings per emotion were presented to each participant. The two recordings were randomly selected from the pool of recordings in GEMEP for each emotion. Two different combinations of context presentation sequences were presented to each participant for each emotion: (1) no-context→action→ R/E and (2) no-context→R/E→action. For the first combination, after hearing the audio file and without any further information, the participants selected the emotion (or the emotions) that they thought the speaker was trying to convey, and set their confidence levels associated with each choice. This allowed the test subjects to indicate their level of certainty about each of the 17 possible emotions (they were not informed that only six emotions would be conveyed in the recordings). Once this was set, the participants moved on to the first contextual cue stage. After hearing the same audio file, the action context was revealed. With this newly-revealed contextual knowledge, the participants made any changes (if they so desired) to both emotion(s) and confidence level(s). Once the second stage was completed, the participants moved on to the third stage, where they heard the same audio file and were presented with the second contextual cue - the R/E context in this case. Using the same method, changes could be made to both emotion(s) and confidence level(s) as the participants felt was warranted. For each scenario/recording, the “winning” emotion was the one with the highest confidence at the end of the third stage. Once the third stage was finished for this recording/scenario, the next of the 12 recordings/scenarios was presented in the same three-stage manner except the order of presentation of the contextual cues was reversed (R/E followed by action). Upon completion of the 12 recordings/scenarios procedures, the data collection process for a test subject was completed.

4. Results

For our study, the confidence level of each emotion for each scenario/recording was logged; a total of 612 data points were recorded for each participant - 12 scenarios/recordings (two for each of the six emotions presented) x 17 emotions x 3 contextual situations (after no-context, after contextual cue #1 and after contextual cue #2).

Table 1 shows the average confidence levels when the contexts are presented in the contextual cues in the two separate sequences.

Emotion Expressed in Recording	Avg. Confidence: NC→Action→R/E			Avg. Confidence: NC→ R/E→Action		
	After No Context	After Action Context	After Action+R/E Contexts	After No Context	After R/E Context	After R/E+ Action Contexts
Anger	21.7%	53.0%	49.0%	38.6%	30.9%	54.6%
Disgust	9.0%	39.5%	42.7%	12.0%	21.2%	49.5%
Fear	28.1%	33.6%	45.8%	32.2%	34.3%	46.3%
Happiness	9.2%	41.1%	46.1%	4.2%	9.8%	49.0%
Sadness	3.3%	33.3%	27.5%	49.1%	44.6%	56.7%
Surprise	1.8%	16.5%	20.3%	1.4%	1.2%	26.9%

Table 1. Average Confidence Levels for the Tests Performed

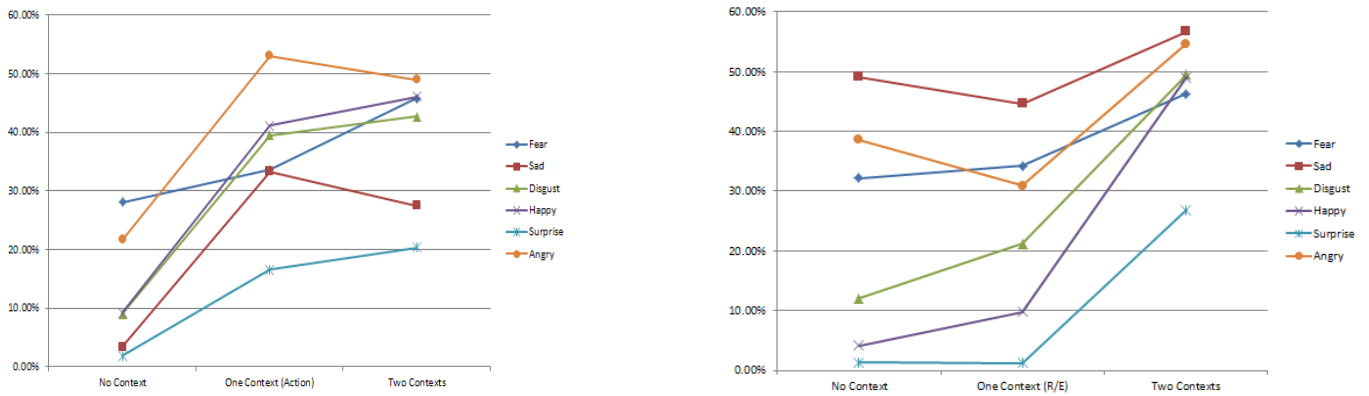


Fig. 2. Graphical Representation of Table 1

Left: Average Confidence: NC → Action → R/E;

Right: Average Confidence: NC → R/E → Action

For anger, the R/E context scripts were: (1) “had a neighbor I didn’t like”; (2) “had a friendship relation”; and (3) “had a co-worker relationship”. The action context scripts were: (1) “vandalized the neighbor’s car”; (2) “abandoned the apartment and owed money”; and (3) “the co-worker seduced his wife”. The test subjects were initially able to rely on the paralinguistic features alone to identify anger reasonably well without contextual knowledge. This may be explained by the fact that anger is often easy for humans to spot. Nevertheless, the confidence levels for anger improved from 21.7% to 53.0% and from 30.9% to 54.6% when the action context was provided.

For disgust, the R/E context scripts were: (1) “had a decayed dead fish”; (2) “had a dead cockroach laid out in the hallway”; and (3) “had a summer job as a janitor”. The action context scripts were: (1) “remembered the previous bad experience”; (2) “stepped on the dead cockroach and the sticky substance stuck on one’s foot”; and (3) “remembered the job where one had to clean many dirty and smelly toilets”. The initial confidences were low after the no-context stage. However, they increased robustly after the contextual cues were presented, as can be seen in Table 1. These numbers reflect increases of 30.5% and 28.3% attributed to the action context.

In the case of fear, the R/E context scripts were: (1) “stood by a stranger”; (2) “owned a car”; and (3) “was alone”. The action context scripts were: (1) “a stranger pointed a knife”; (2) “could not stop the car due to brake failure”; and (3) “something unexpectedly felt on one’s shoulder”. Similarly to anger, humans were able to rely reasonably well with only the paralinguistic features. Nevertheless, the confidence levels for fear improved after provision of contextual information (see Table 1).

For happiness (joy), the R/E context scripts were: (1) “had a lottery ticket”; (2) “had many friends”; and (3) “had a long-distance friend”. The action context scripts were: (1) “won the lottery”; (2) “received a beautiful gift”; and (3) “spent time together”. The confidence levels for happiness improved from 9.2% to 41.1% and from 41.1% to 46.1% for the no-context→action→R/E case and from 4.2% to 9.8% and from 9.8% to 49.0% for the no-context→R/E→action case. For this emotion, the data suggests that the action context had a stronger effect on the human subject’s affect perception when the action context was presented after the R/E context, as the confidence level proceeded to increase significantly, from 5.6% to 39.2%.

For sadness, the R/E context scripts were: (1) “had final exams to pursue the dream career”; (2) “was in a marriage relationship”; and (3) “had a dog”. The action context scripts were: (1) “failed test and resigned from a dream job”; (2) “received divorce paperwork”; and (3) “had to euthanize the dog”. The confidence levels for sadness changed positively from 3.3% to 33.33% but then

negatively from 33.3% to 27.50% for the no-context→action→R/E case. Similarly, provision of the R/E context first had a negative effect in the no-context→R/E→action case, decreasing from 49.1% to 44.6%. However, the confidence recovered after the action context, going from 44.6% to 56.7%. It is clear that for the sadness emotion, the R/E context only served to confuse the test subjects.

Finally, the R/E context scripts for surprise were: (1) “was in an office”; (2) “was at the waterfall park”; and (3) “was in the dining room”. The action context scripts were: (1) “saw a pigeon fly into the office”; (2) “the waterfall stopped flowing”; and (3) “heard loud noise from next door”. The confidence levels for surprise increased from 1.8% to 16.5% and from 16.5% to 20.3% for the no-context→action→R/E case, and decreased slightly from 1.4% to 1.2% before significantly increasing from 1.2% to 26.9% for the no-context→R/E→action case. Once again, the action context showed to be a stronger positive influence than the R/E context.

The data and discussion above clearly suggest that in general, the action context had significantly more positive influence than did the R/E context. The latter even had occasional negative effects. Now, let’s look at the percent of test subject evaluations that showed the correct emotion as the one having the highest confidence as assigned by the test subjects. For this analysis, we combine all assessments for all emotions, making a data set of 720 data points – 60 test subjects, each evaluating 12 recordings with and without contextual information. Furthermore, we compare how many test subjects selected the correct emotion on a context-free basis vs. how many did so after knowing both contextual cues (regardless of the order in which these were provided). See Table 2 below.

Two things stand out about the results shown on Table 2. The first is how difficult it is for humans to discern the emotion expressed by someone through paralinguistic information only, with or without context. From this data, only fewer than 23% of the test subjects were able to correctly identify the emotion being conveyed by a speaker when given only the paralinguistic information (a recording). While that number more than doubled to nearly 53% when the test subjects were provided with contextual cues, the latter is still a remarkably low success rate. This validates the current trend in the research community to use multiple modes of signals to identify a person’s emotional state as he/she utters unintelligible sounds (e.g., facial expressions, body language, hand/arm gestures).

	Correct Selections Made (highest confidence assigned)	Percentage of Correct Selections Made (out of 720)
Without Context Information	161	22.4%
With Full Context Known	377	52.4%

Table 2. *Percent of the 720 evaluations in which the test subject assigned the highest confidence to the correct emotion, with and without contextual information.*

The results of Tables 1 and 2 serve to validate our hypothesis that humans’ affect recognition ability was improved when the context under which the speakers uttered the expressions was known.

An interesting question would be how the test subjects would perform if the paralinguistic signals and the contextual information were contradictory. For example, if the R/E context was a child’s birthday party and the action context was blowing out the candles on the cake as everyone applauded, yet, the utterances emitted by the actor were of fear or anger. We intuitively believe that this would result in a significantly lower rate of correct emotion selection with full context known (bottom right cell in Table 2) by the test subjects. However, it would be interesting to confirm this hypothesis experimentally, and if confirmed, how much exactly would be the decrease. Even more interesting will be to determine if contradictory/misleading contextual information lowers the

correct selection rate to a level worse than the context-free rate (right-most cell on the middle line in Table 2).

5. Conclusion and Future Research

Our study has shown the generally positive impact of incremental contextual knowledge being presented to the test subjects in their task to infer the emotion expressed by the actors through paralinguistic features of speech. Between the two contextual knowledge cues identified, we found that the action context had a more significant as well as consistent positive impact on the confidence levels expressed by the test subjects than did the R/E context. Our results also showed that in general, humans had limited ability to recognize emotion based solely on paralinguistic speech with no contextual knowledge, and to some degree, even with contextual knowledge.

Our next objective will be to integrate context modeling into the traditional automatic affect recognition with paralinguistic speech processes currently available from the published literature. Supported by our findings in this study, we believe that context-based affect recognition from paralinguistic signals will be more robust and have improved recognition ability.

An objective of our future research will be to investigate why the action context was consistently more effective than the R/E context in influencing the test subjects towards identification of the correct emotion expressed by the actor.

Some of the remaining questions that are left for future research include why the R/E context decreased the confidence in some cases while it contributed positively to the identification of some other emotions. Because the expression of an emotion is not produced in a vacuum but has some causal chains, we also want to investigate other contextual knowledge information that can give positive impact in the affect recognition algorithm and can be gathered and analyzed automatically by an automated system.

Bibliography

- [1] PICARD R. *Affective Computing*. The MIT Press (1997).
- [2] CALVO R., D'MELLO S. *Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications*, IEEE Transactions on Affective Computing. 1(1) pp. 18-37. (2010).
- [3] SCHULLER B., BATLINER A., BERGLER C., POKORNY F., KRAJEWSKI J., CYCHOSZ M., VOLLMANN R., ROELEN S., SCHNIEDER S., and BERGELSON E., *The INTERSPEECH 2019 computational paralinguistics challenge: Styrian dialects, continuous sleepiness, baby sounds & orca activity*. Proceedings of the INTERSPEECH 2019 conference. (2019).
- [4] D'MELLO S., BOSCH N., and CHEN H., *Multimodal-multi sensor affect detection*. The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition - Volume 2, Association for Computing Machinery and Morgan & Claypool, New York, NY. (2018).
- [5] SADOUGHI N., BUSSO C., *Speech-driven animation with meaningful behaviors*. Speech Communication. 110, pp. 90–100. (2019).
- [6] REN M., NIE W., LIU A., and SU Y., *Multi-modal correlated network for emotion recognition in speech*. Visual Information, 3(3), pp. 150–155, (Sep. 2019).
- [7] ZHANG S., ZHANG S., HUANG T., GAO W., and TIAN Q., *Learning Affective Features With a Hybrid Deep Model for Audio–Visual Emotion Recognition*. IEEE Transactions on Circuits and Systems for Video Technology 28(10): 3030-3043. (Oct. 2018).
- [8] LI S., DENG W., *Deep facial expression recognition: A survey*, arXiv preprint arXiv:1804.08348, (2018).
- [9] RIM B., SUNG N., MIN S., HONG M., *Deep Learning in Physiological Signal Data: A Survey*. Sensors, 20(4), 969, (2020).

- [10] SARIYANIDI E., GUNES H., and CAVALLARO A., *Automatic analysis of facial affect: A survey of registration, representation and recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) conference. (2015).
- [11] SAMIRA P., SAAD S., YILIN Y., HAIMAN T., YUDONG T., MARIA R., MEI-LING S., SHU-CHING C., and IYENGAR S., *A Survey on Deep Learning: Algorithms, Techniques, and Applications*. ACM Computing Surveys (CSUR), 51(5):92:1–92:36, (2018).
- [12] TZIRAKIS P., CHEN J., ZAFEIRIOU S., SCHULLER B., *End-to-end multimodal affect recognition in real-world environments*. Information Fusion, 68:46–53, (2021).
- [13] ROUAST P., ADAM M., *Learning deep representations for video-based intake gesture detection*. IEEE J. Biomedical Health Information, 24(6): 1727–1737, (Jun. 2020)
- [14] KOSTI R., ALVAREZ J., RECASENS A., and LAPEDRIZA A., *Context based emotion recognition using EMOTIC dataset*, IEEE Transaction on Pattern Analytic Machine Intelligence, early access, (May 2019).
- [15] HOLLISTER D., GONZALEZ A., and HOLLISTER, J., *Contextual Reasoning in Human Cognition and the Implications for Artificial Intelligence Systems*, International and Interdisciplinary Conference on Modeling and Using Context, pp. 599-608. (2017).
- [16] KOKINOV B., *A dynamic approach to context modeling*. Proceedings of the IJCAI-95 workshop on modeling context in knowledge representation and reasoning. LAFORIA, 11, pp. 199-209, (1995).
- [17] GLUCKSBERG S., KREUZ R., and RHO S., *Context can constrain lexical access: Implications for models of language comprehension*. Journal of Experimental Psychology: Learning, Memory, and Cognition. 12(3). (1986).
- [18] KIDD E., LUM J., *Sex differences in past tense over-regularization*. Developmental Science. 11(6): 882- 889. (2008).
- [19] RICE M., TOMBLIN J., HOFFMAN L., RICHMAN W., and MARQUIS J., *Grammatical tense deficits in children with SLI and nonspecific language impairment: Relationships with nonverbal IQ over time*. Journal of Speech, Language, and Hearing Research 47(4): 816-834. (2004).
- [20] PIAGET J., *The Development of Thought: Equilibration of Cognitive Structures*. New York: Viking. (1976).
- [21] TREISMANN A., *Contextual cues in contextual listening*. Quarterly Journal of Experimental Psychology, 12(4). (1960).
- [22] RUSSELL J., *Core Affect and the Psychological Construction of Emotion*. Psychological Review 110: 145-172. (2003).
- [23] PANKSEPP J., *Affective Neuroscience: The Foundations of Human and Animal Emotion*. Oxford Univ. Press. 1998.
- [24] KELTNER D. and HAIDT J., *Social Functions of Emotions*. Emotions: Current Issues and Future Directions, T.J. Mayne and G.A. Bonanno, eds. Guilford Press. pp. 192-213. (2001).
- [25] BACHOROWSKI J., OWREN M., *Vocal Expression of Emotion—Acoustic Properties of Speech Are Associated with Emotional Intensity and Context*. Psychological Science. 6: 219-224. (1995).
- [26] BANZIGER T., MORTILLARO M. and SCHERER K., *Introducing the Geneva Multimodal Expression corpus for experimental research on emotion perception*. Emotion. Advance online publication. (2011).
- [27] EKMAN P., *An argument for basic emotions*. Cognition & Emotion. 6:169–200. (1992).
- [28] Google, Inc. <https://www.google.com/> (2021).
- [29] MARPAUNG A., *Context-Centric Affect Recognition From Paralinguistics Features of Speech*, Ph.D. Dissertation. University of Central Florida. (2019).