A Survey on Usage of Multimedia Databases for Emotion Elicitation: A Quantitative Report on How Content Diversity Can Improve Performance

M. Horvat* and P. Jerčić**

* University of Zagreb, Faculty of Electrical Engineering and Computing, Department of Applied Computing, Zagreb, Croatia

** Graz University of Technology, Institute of Interactive Systems and Data Science, Graz, Austria

marko.horvat3@fer.hr, pjercic@tugraz.at

***Abstract* – Affective picture databases provide a standardized set of images to elicit controlled and consistent emotional responses in research participants. They are a valuable tool for studying various emotion-related phenomena across several research domains. These domains include emotion perception, emotion regulation, and the neural basis of emotion. However, affective picture databases have diverse schemas, structures, and content, making them difficult to use. Searching and retrieving optimal pictures relevant to affective stimulation may be challenging and time-consuming. In this context, we surveyed domain experts about their practices and experiences working with affective multimedia databases such as IAPS, NAPS, OASIS, GAPED, and others. The survey identified a need for novel data observatory software. This finding motivates the authors' intention to develop and validate such software platform that relies on AI. Such a platform would describe better, retrieve, and integrate various semi-structured affective multimedia datasets. The results prominently indicate the overwhelming dissatisfaction regarding stimuli content diversity and cultural bias, specifically regarding emotional and semantic context. The main driver of satisfaction from users of existing automated retrieval software is the quality of semantic descriptors available. This points to the direction AI should take in novel data observatory software. This survey follows up on a similar survey conducted ten years ago and explores the differences in researchers' opinions and experiences during that time. The complete aggregated results are publicly available at https://github.com/mhorvat/stimdbsurvey.**

Keywords - affective computing, emotion, databases, emotion stimulation, document retrieval

I. Introduction

Digital multimedia, apart from having a format, semantic content, and context, also can provoke emotions in observers. When people engage with various forms of digital media, such as movies, music, books, or virtual reality, they are inherently emotionally stimulated. The process of eliciting emotional responses through digital multimedia is complex and multifaceted, with underlying neuroanatomical mechanisms [1][2]. The type and intensity of the elicited emotional states can be modeled probabilistically for a given stimulus [3]. Affective multimedia databases, also known as multimedia stimuli databases, are a particular type of collection containing processed multimedia documents specifically designed to elicit emotions in a controlled laboratory environment [4]. Documents are often referred to as stimuli, depending on their purpose, while images and videos are typically referred to as visual stimuli [4][5]. These databases store the semantics, context, and affect of the stimuli and categorize them according to common emotion models [4][5][6]. Although traditionally considered non-interchangeable, recent research has challenged this notion [7][8]. Affective multimedia databases have practical applications not only in the study of human emotion mechanisms but also in perception, memory, attention, and reasoning [9].

Affective multimedia databases, for example, International Affective Picture System (IAPS) [10], Nencki Affective Picture System (NAPS) [11], Open Affective Standardized Image Set (OASIS) [12], The Geneva Affective PicturE Database (GAPED) [13], NimStim Face Stimulus Set [14], are designed to be used in research of emotion processing, attention, stress resilience, and mental health, yet a lot can be done to improve their models, facilitate their usage and expand prevalence in the field.

Our previous work showed that multimedia documents are sparsely annotated, making semantic retrieval difficult and resulting in low recall and precision [15]. Furthermore, high-level semantic content descriptors are informal and ambiguous due to the insufficient annotation methods which rely on unrestricted keywords. This limited meta-retrieval framework has no unified stimuli content description dictionary, underlying knowledge base, concept taxonomy, or terminology. These problems contribute to a prolonged and work-intensive construction of stimuli sequences, difficulties in finding optimal elicitation, and emphasize the need to thoroughly structure affective stimuli databases and develop new tools for searching and retrieving emotion stimuli [7][15].

Currently, stimuli are often manually extracted from multimedia databases through a time-consuming visual inspection of each stimulus and its associated database manuals. Complicating matters further, multimedia stimuli databases are structured differently, may describe emotional and semantic data differently, and may contain different media formats. In addition, there is no agreement in the research community on an optimal structure or standard implementation of a multimedia stimulus database. For these reasons, a typical user in any database must be skilled in both emotion capture and technology-related tasks such as stimulus selection and extraction. Unfortunately, it is difficult and time-consuming to master such a fusion of different skills. Furthermore, because databases differ in structure, these skills must be learned separately for each database.

These difficulties suggest that a novel AI-based computer system is required to assist experts in finding the most appropriate stimulus and completing it quickly. A system of this type must be as universal as possible regarding database models and media formats and efficient and user-friendly. Furthermore, to aid the expert, the system should ideally include empirically derived rules for decision support and the automatic generation of stimuli sequences. Again, such a stimuli generator should produce personalized stimuli tailored to specific emotional and semantic parameters [7][15].

To determine the need for such a software tool in the professional community, we conducted an online survey to examine how researchers in psychology, neurology, cognitive science, and related fields use multimedia stimuli databases to elicit emotional responses. In addition, the survey identified current problems in the use of affective multimedia databases and areas where further improvement is needed. The study follows up on a similar survey on the usage patterns of stimuli databases undertaken ten years ago [16].

The remainder of this paper is organized as follows: Section 2 lists all the survey questions and systematically describes the methodology and the mixed-effects model used to explain the research questions. The aggregated survey results are presented in Section 3. Section 4 presents the statistical analysis of the survey results and shows which semantic descriptors (i.e., tags) influence users' satisfaction with affective multimedia databases. In addition, in Section 4, the results of this survey are compared to the previous survey's results, and the key differences and similarities between the two studies are highlighted. Finally, Section 5 summarizes the paper and suggests future directions for developing affective multimedia databases.

II. SURVEY METHOD

Between 1 December 2022, and 15 January 2023, an invitation to participate in an online survey was sent to 336 e-mail addresses of authors of published papers using at least one emotionally annotated database. Invitations were sent to each author twice within 2-4 weeks. The authors were demographically diverse from different higher education and research institutions. IAPS, NAPS, OASIS, NimStim, and GAPED were the databases addressed. In addition, some invitees used Karolinska Directed Emotional Faces (KDEF) [17], International Affective Digital Sounds (IADS) [18], and Pictures of Facial Affect (POFA) [19]. Relevant publications and author contact information were manually searched via Web of Knowledge, PubMed, IEEE Xplore, and Google Scholar by tracking citations in the corresponding databases. Of 336 invitations, 87 were bounced because of invalid emails, of which 54 were resent. In total, N=37 invited individuals completed the survey (12.13%). Individual responses were automatically recorded and aggregated using the Google Forms service for creating online forms and surveys. The anonymity of the participants was guaranteed by the service provider, as stated in the e-mail invitation. Participation in the survey was completely anonymous. Individuals participating in the survey were not identifiable, and no personal information was stored. Multiple responses were not permitted.

The survey consisted of 19 questions: 5 questions (1, 13, 15, 18, 19) were free text, and 14 others with predefined answers, including the Likert scale with a range of 1–5. Questions 15, 16, and 19 were optional. Likert scale questions also had the "Not Applicable/Prefer Not to Answer" option as one of the possible answers.

The survey questions were:

Q1. What best describes your research topic?

Q2. Please rate the image retrieval process from multimedia stimuli databases such as IAPS, NAPS, GAPED, OASIS, KDEF, POFA, JAFFE, and IADS?

Q3. How satisfied are you with the above-mentioned level of difficulty in the image retrieval process from the database?

Q4. How much time was necessary to effectively search the database and construct one picture sequence that was used in your research?

Q5. On a scale of 1 to 5, with one being very inadequate and 5 being very adequate, please rate the quality of the documentation (guide instructions and support)?

Q6. Have you at any time felt that the picture set you were using is missing images with a particular emotion that would be useful for your stimuli sequence?

Q7. Have you at any time felt that the picture set you were using is missing images with semantic content that would be useful for your stimuli sequence?

Q8. Have you felt that the picture set you were using addressed the cultural values in your target group?

Q9. On a scale of 1 to 5, with one being very insufficient and 5 being very sufficient, please rate the diversity of semantic and emotional content in the picture sets you used.

Q10. On a scale of 1 to 5, with one being very inadequate and 5 being very adequate, please rate how inadequate (insufficient) did you find the predefined semantic descriptors (e.g., keywords or tags) of the images you used?

Q11. On a scale 1 to 5, with one being extremely useless and 5 being extremely useful, please rate how helpful would a user-friendly software tool for intelligent retrieval of emotionally annotated images be to your research?

Q12. Have you at any time during your research wanted to find the most appropriate emotionally-annotated images faster and more efficiently?

Q13. Can you give a rough indication of the duration in minutes of one of your test sessions (the period in which one participant is tested continuously)?

Q14. Did you construct the sequence manually or with a help of a software tool?

Q15. Which software tool did you use (if any)?

Q16. Did your group actually develop the tool used in your experiment or acquired it elsewhere?

Q17. How useful or useless a stimuli database with realistic and immersive Virtual Reality (VR) images instead of just still ones, would be to your work?

Q18. Please list the names of multimedia stimuli databases you have used in your research (e.g., IAPS, NAPS, GAPED, OASIS, KDEF, POFA, JAFFE, IADS).

Q19. Do you wish anything to add that was not covered in the survey?

Addressing the ambiguity on what influences user ratings concerning affective multimedia databases, we wanted to investigate the impact of the following elements mentioned in the survey: 1) the quality of documentation (manuals, online help), 2) the diversity of available stimuli, and 3) the expressiveness of the content descriptions (keywords). Furthermore, considering the difference between users in terms of previous use of automatic software to create stimuli sequences and building it manually, we proposed the following research question:

Which factors impact user satisfaction with affective multimedia databases – the quality of documentation (manuals, online help), diversity of available stimuli, or expressiveness of the content descriptions (keywords)?

We have designed a linear mixed-effect model to investigate the drivers (i.e., factors) for user ratings of affective multimedia databases. More specifically, we considered the quality of the documentation (manuals, online help), diversity of available stimuli, and quality of semantic descriptors as independent variables, affecting satisfaction ratings with affective multimedia databases as dependent variables. The analysis aims to estimate the rating scores of the users of affective multimedia databases based on the survey data we have collected. First, we grouped the users based on the previous usage of automatic retrieval software and those retrieving it manually. Following this, we focused on the variables that significantly affect the satisfaction ratings while still considering the variability from other factors. We used a mixed-effect model because the rating scores from the same user groups regarding previous experience in using automated retrieval software were correlated. Moreover, it makes sense to structure the data given our interest in automated retrieval software.

Given the small number of users involved in this high-dimensional data, it was considered correct to use the restricted maximum likelihood as an estimation method since the bias to this method might over or underestimate the true variance.

III. SURVEY RESULTS

Keywords best describing participants' research topics (Q1) are visualized as a tag cloud in Figure 1. Stop words were removed, and keywords were lemmatized.

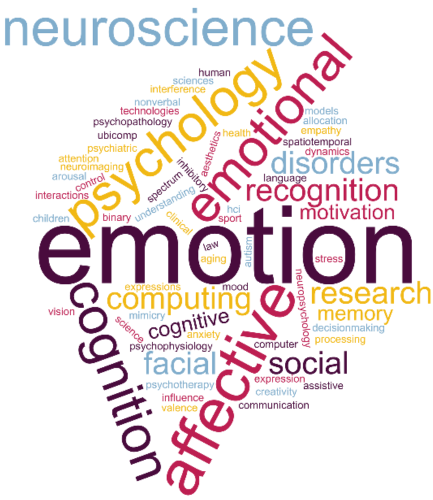

Figure 1. Keywords that best describe survey participants' research topics. Normalized answers to Q1.

The names of multimedia stimuli databases the participants used in their research (Q18) are shown in Figure 2. The most frequent are IAPS (27 answers), NAPS (15), KDEF (11), GAPED (8), OASIS (8), IADS (5), NimStim (3), JACFEE (2), and MSFDE (2). Further, 32 different databases were mentioned one time each.

Figure 2. Keywords that best describe survey participants' research topics. Answers to Q18.

The aggregated survey results to Likert scale questions are displayed as individual charts in Figure 3 and Table 1 for better visibility. Because of the restrictions of the survey tool, the standard answer "Not Applicable/Prefer Not to Answer" is shown in a different color between the charts in Figure 3.

All survey results are freely available to the academic community and for research purposes on the GitHub online repository (https://github.com/mhorvat/stimdbsurvey) and can be downloaded in CSV, Excel, and other standard formats.

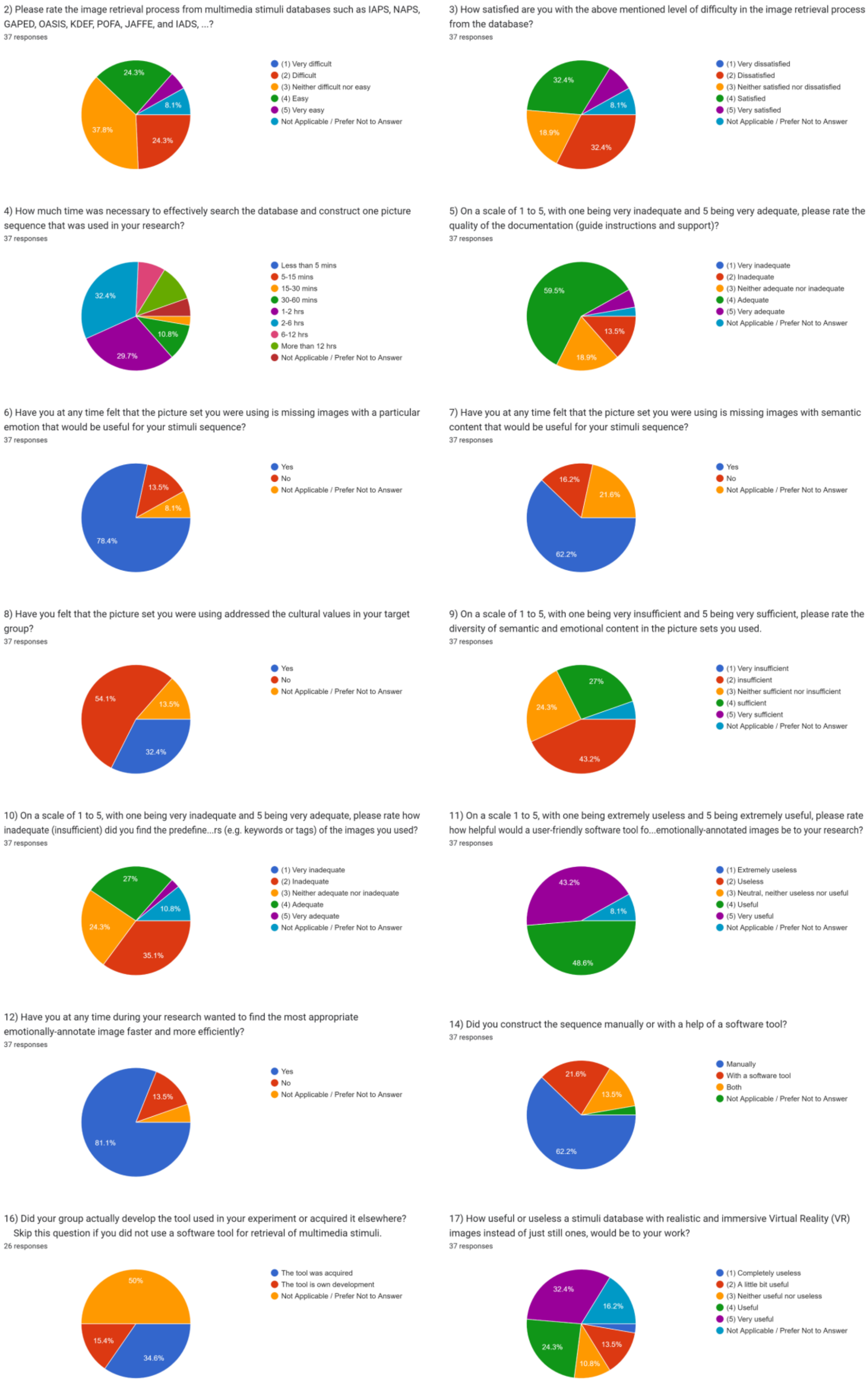

Figure 3. Overview of the aggregated survey results for Likert-scale questions.

TABLE I. TABLE OVERVIEW OF THE AGGREGATED SURVEY RESULTS FOR LIKERT-SCALE QUESTIONS.

Question	Answers
Q2	Very difficult = 0%, Difficult = 24.3%, Neither difficult nor easy = 37.8%, Easy =24.3%, Very easy = 5.4%, NA = 8.1%
Q3	Very dissatisfied = 8.1%, Dissatisfied = 32.4%, Neither satisfied nor dissatisfied = 18.9%, Satisfied = 32.4%, Very satisfied = 8.1%, NA = 8.1%
Q4	Less than 5 mins = 0%, 5-15 mins = 0%, 15-30 mins = 2.7%, 30-60 mins = 10.8%, 1-2 hrs = 29.7%, 2-6 hrs = 32.4%, 6-12 hrs = 8.1%, More than 12 hrs = 10.8%, NA = 5.4%
Q5	Very inadequate = 0%, Inadequate = 13.5%, Neither adequate nor inadequate = 18.9%, Adequate = 59.5%, Very adequate = 5.4%, NA = 2.7%
Q6	Yes = 78.4%, No = 13.5%, NA = 8.1%
Q7	Yes = 62.2%, No = 16.2%, NA = 21.6%
Q8	Yes = 32.4%, No = 54.1%, NA = 13.5%
Q9	Very insufficient= 0%, Insufficient= 43.2%, Neither sufficient nor insufficient = 24.3%, Sufficient = 27%, Very sufficient = 0%, NA = 5.4%
Q10	Very inadequate = 0%, Inadequate = 35.1%, Neither adequate nor inadequate = 24.3%, Adequate = 27%, Very adequate = 2.7%, NA = 10.8%
Q11	Extremely useless = 0%, Useless = 0%, Neutral, neither useless nor useful = 0%, Useful = 48.6%, Very useful = 43.2%, NA = 8.1%
Q12	Yes = 81.1%, No = 13.5%, NA = 5.4%
Q14	Manually = 62.2%, With a software tool = 21.6%, Both = 13.5%, NA = 2.7%
Q16	The tool was acquired = 34.6%, The tool is own development = 15.4%, NA = 50%
Q17	Completely useless = 2.7%, A little bit useful = 13.5%, Neither useful nor useless = 10.8%, Useful = 24.3%, Very useful = 32.4%, NA = 16.2%

Regarding the main drivers for user satisfaction with using these databases, semantic descriptors reached a statistical significance coefficient, while diversity and documentation did not (95% confidence interval). The results are presented in Figure 4 below, and detailed results are in Table 2 (Appendix).

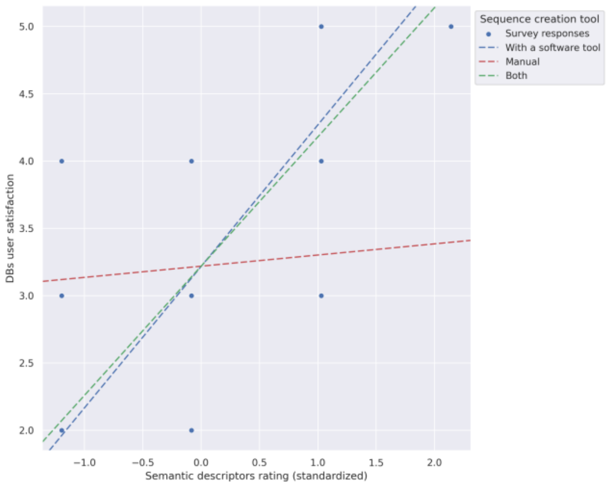

Figure 4. Keyword semantic descriptors were influencing user satisfaction with affective multimedia databases. Data is grouped based on experience with sequence creation tools (automatic retrieval software, manual creation, or both).

IV. DISCUSSION AND RESULTS ANALYSIS

While researchers are divided on the difficulty in retrieving images from stimuli datasets (Q2 and Q3), the survey clearly showed that sequences are still predominantly constructed manually, with experts visually searching multimedia repositories for prolonged periods (Q4). The process of stimuli construction takes 2-6 hours (32.4%), 1-2 hours (29.7%), more than 12 hours (10.8%), 30-60 minutes (10.8%), and 6-12 hours (8.1%). Therefore, 51.3% need 2 hours or more to construct an emotion elicitation sequence. Dataset documentation (Q5), when available, was primarily judged to be adequate (59.5%) and very adequate (5.4%), while a minority (13.5%) believed that the documentation was inadequate or neutral (18.9%).

A large majority of participants (78.4%) think that images with a specific emotion are missing from the dataset they use (Q6), and 62.2% believe that images with semantic content that could be useful for their work are missing (Q7). In addition, 54.1% feel that the imagery they use does not address the cultural values of the target group, and only 32.4% feel that it does (Q8). Finally, the diversity of semantic and emotional content was rated as insufficient by 43.2% of the participants (Q9).

Semantic information, keywords, or tags describing the stimuli were considered inadequate by 35.1% and neither adequate nor inadequate by another 24.3% of the researchers (Q10). Although no one (0%) thought that the existing stimuli labels were very inadequate; on the other hand, only 2.7% felt that they were very adequate. Clearly, the description of stimuli information should be improved.

An overwhelming percentage of participants (total of 91.8%) declared that a user-friendly software tool for intelligent retrieval of affective images would be very useful (43.2%) or useful (48.6%) for their research (Q11). Very importantly, none (0%) indicated that an intelligent software tool for stimuli retrieval would either be highly useless or had a neutral opinion, believing it would be neither useless nor useful. This (Q11) indicates that the professional community would welcome such a tool.

In addition, 81.1% of examinees stated that they would like to find the most appropriate images more quickly and efficiently. However, only 13.5% reported this was not the case and were satisfied with the current stimuli search speed (Q12).

Responses to Q13 (Figure 5) show that one test session lasts 30 minutes for 29.7%, 60 minutes for 18.9%, and 20 minutes for 16.2% of most researchers. Outliers are 1 (min) and 120 minutes (max).

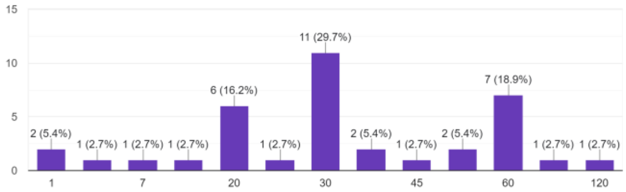

Figure 5. Duration of one test session in minutes. Answers to Q13.

As many as 62.2% of respondents constructed stimuli sequences manually, and only 21.6% with the aid of a software tool, while another 13.5% used both methods (Q14). The tools used for constructing stimuli sequences are (Q15): Psychopy, E-Prime, Psychtoolbox in Matlab, PicRotator (v1.0 and v2.0), Qualtrics, Testable, Presentation, and an online random sequence generator. In answering Q15, some survey respondents indicated tools intended only for presenting stimuli. As a result, 15.4% of researchers (N=26) developed the tool, and 34.6% acquired it elsewhere (Q16).

Only 2.7% of respondents consider immersive Virtual Reality (VR) completely useless (Q17). On the other hand, a large majority (70.2%) support introducing new computer-based methods in creating and delivering emotion-provoking content at any level, while 13.5% are neutral (Q17). Together with the previous answers, this once again underscores the need to continue producing fresh multimedia content and creating new databases with rich and personalized stimuli.

In comments on various topics (Q19), participants called for more consistency in documentation. They also pointed out that more racial diversity of images and more emotion types would be helpful (e.g., bored, disgusted). Some asked for neutral stimuli that are as semantically related as the emotional stimuli and positive arousing stimuli and emotional stimuli that do not occur in distant, exotic locations. The number of available images, as is the ability to assign images based on various features other than valence, is essential. It was also pointed out that there have yet to study children. Some were concerned about the quality of the ground truth data and how the dataset authors collected it. Some researchers found it easy to construct sequences based on published normative ratings but found it challenging to work with data dimensions beyond those covered by the normative ratings (i.e., beyond valence, arousal, and basic emotions).

Notably, regarding user satisfaction with these databases, we found that semantic descriptors have the most significant influence on predicting user satisfaction and significantly affect rating scores. Furthermore, the effect was strong for users with previous experience with automated retrieval software, while it was weaker for those who retrieved manually. This effect was significant even after controlling for the variance from other confounding factors in the survey.

A. *Comparison to the previous survey results*

In the earlier similar survey reported in 2013 [16], most respondents felt that the descriptions of the stimuli were inadequate and that it generally takes 1-2 hours or more than 24 hours to complete a single stimulation sequence. In the earlier survey, a large majority (84%) of participants indicated that real-life videos would be useful in their research. The experts recognized the need for an intelligent stimuli retrieval application to assist them in their experiments. Almost all experts then agreed that such software tools could be helpful in their work.

In light of the ten years between the two surveys, there were some shifts in the researchers' opinions. The image retrieval process has become more difficult in recent years. While at the same time, satisfaction with this difficulty increased. With more invested effort, possibly higher gains can be obtained using multimedia databases in recent years. This fact is crucial if researchers invest more time effectively searching the database and constructing one picture sequence. More specifically, there has been a 3-fold increase in spending 2 - 6 hours on this task. By segregating semantic or emotional content questions in the new survey, we observed that both are a problem, with emotional content being a more significant issue than the semantic one. At the same time, the adequacy of the descriptions of the images remained stable. Half the researchers responded that diversity or underrepresented cultural values might be responsible for this dissatisfaction.

On the other hand, researchers have positively consolidated their opinions on a user-friendly software tool for the intelligent retrieval of emotionally-annotated images. Increased difficulty in the image retrieval process might shift some respondents towards expressing a need for such a tool. However, the need to find the most appropriate emotional annotated image faster and more efficiently remained stable. Furthermore, researchers who used software tools in the past to accompany their manual construction of the sequence have shifted towards fully manual construction. This fact might indicate a need for a more intelligent software tool, as research slightly sifted towards acquiring software rather than developing their own.

V. Conclusion

The survey provided valuable insights into the opinion, motivation, and experience of experts who use affective multimedia databases to elicit emotional responses from subjects. Although the survey was not large in terms of the questions' complexity or the number of participants, it provided statistically significant results. Although email invitations certainly limited the number of responses received, it was the only method available to obtain the valuable opinions of a wide range of professionals.

Users creating the stimuli sequence manually had no significant effect of the investigated variables on their rating scores compared to the users of automatic software for stimuli sequence creation. Furthermore, their user ratings were remarkably stable over the investigation space, generally fixed on the middle rating.

The results indicate that automatic retrieval software for stimuli sequence creation has a clear advantage over manual creation. This is because users may give higher score ratings for the availability of good semantic descriptors. However, on the other hand, poor semantic descriptors in such automatic software drive their ratings very low, lower even than manual creation. Potentially, the non-availability of good semantic descriptors renders such automatic software unusable to create a satisfying sequence, which users with manual pipelines would still be able to operate in such adverse affective multimedia databases.

Considering these findings, an intelligent software platform for automatic sequence generation from affective multimedia databases should focus on providing quality semantic descriptors. This idea feeds well into our initial goal of developing more effective methods and procedures for generating semantic descriptors based on knowledge graphs to better describe, retrieve, and integrate different semi-structured stimuli datasets. We hope that such a tool will help all researchers in the field of affective computing, psychology, neurology, and cognitive science to find the most appropriate multimedia stimuli with minimal effort and in the shortest possible time. This should improve the quality of personalized emotion elicitation and reduce the experts' workload, allowing them to focus more on the subject and stimulation procedures.

The collected results indicate the need to develop new stimuli databases with fresh multimedia content. Generative AI graphics and VR delivery are good research directions, as they offer many advantages over acted or real visualizations. However, creating more databases further highlights the need for intelligent software to help manage multifaceted affective multimedia content.

The survey's authors would like to cordially thank all experts who have taken part in the survey.

References

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

[2] M. Lewis, J. M. Haviland-Jones, L. F. Barrett (Eds.), "Handbook of emotions," Guilford Press, 2010.

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

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

[5] Y. Wang, W. Song, W. Tao, A. Liotta, D. Yang, X. Li, G. Shuyong *et al.*, "A systematic review on affective computing: Emotion models, databases, and recent advances," Information Fusion, 2022.

[6] M. Horvat, A. Stojanović, Ž. Kovačević, "An overview of common emotion models in computer systems," In 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), pp. 1008–1013, 2022.

[7] M. Horvat, A. Jović, K. Burnik, "Investigation of Relationships between Discrete and Dimensional Emotion Models in Affective Picture Databases Using Unsupervised Machine Learning," Applied Sciences, vol. 12, no. 15, p. 7864, 2022.

[8] S. Zhao, G. Ding, Y. Gao, X. Zhao, Y. Tang, J. Han, ... Q. Huang, "Discrete probability distribution prediction of image emotions with shared sparse learning," IEEE Transactions on Affective Computing, vol. 11, no. 4, pp. 574-587, 2018.

[9] T. Brosch, K. Scherer, D. Grandjean, D. Sander, "The impact of emotion on perception, attention, memory, and decision-making," Swiss medical weekly, vol. 143, no. 1920, w13786−w13786, 2013.

[10] P. J. Lang, M. M. Bradley, B. N. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," Technical Report A-8, University of Florida, Gainesville, FL, 2008.

[11] A. Marchewka, Ł. Żurawski, K. Jednoróg, A. Grabowska, "The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database," Behavior research methods, vol. 46, pp. 596−610, 2014.

[12] B. Kurdi, S. Lozano, M. R. Banaji, "Introducing the open affective standardized image set (OASIS)," Behavior research methods, vol. 49, pp. 457−470, 2017.

[13] E. S. Dan-Glauser, K. R. Scherer, "The Geneva Affective PicturE Database (GAPED): A new 730 picture database focusing on valence and normative significance," Behavior Research Methods, vol. 43(2), pp. 468–77, 2011.

[14] N. Tottenham, J. W. Tanaka, A. C. Leon, T, McCarry, M. Nurse, T. A. Hare, D. J. Marcus, A. Westerlund, B. J. Casey, C. Nelson, "The NimStim set of facial expressions: Judgments from untrained research participants," Psychiatry Research, vol. 168:3, pp. 242–249, 2009.

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

[16] M. Horvat, S. Popović, K. Ćosić, "Multimedia stimuli databases usage patterns: a survey report," in Proceedings of the 36nd International ICT Convention MIPRO, pp. 993–997, 2013.

[17] D. Lundqvist, A. Flykt, A. Öhman, "The Karolinska Directed Emotional Faces – KDEF, " CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 1998.

[18] P. J. Lang, M. M. Bradley, B. N. Culthbert, "International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings (Tech. Rep. No. B-2)," The Center for Research in Psychophysiology, University of Florida, U.S.A., 1999.

[19] The Paul Ekman Group, LLC, retrieved from: https://face.paulekman.com/face/productdetail.aspx?pid=1

VI. Appendix

TABLE II. Parameter of the linear mixed-effects model and the formula: "c2 ~ q10 + 1", groups= "q14", re_formula="~q10 + q5 + q9 + 1". The questions used in the model are listed in Section 2.

Parameter name	Beta	Std. Err.	Lower-95	Upper-95	Random effect (SD)
Intercept	3.221	0.177	2.873	3.568	
Q10	0.741	0.312	0.130	1.352	
Q5					0.191
Q9					0.021