

Methodology of Evaluation the Quality Level of Multimedia Content Based on the Emotional Perception of the Focus Group Respondent

Iryna Spivak¹, Svitlana Krepych¹, Oleksandr Fedorov¹ and Serhii Spivak²

¹ West Ukrainian National University, Lvivska str. 11, Ternopil, 46009, Ukraine

² Ternopil Ivan Puluuj National Technical University, Ruska str. 56, Ternopil, 46001, Ukraine

Abstract

The article deals with the problem of quality evaluation of multimedia content. The expediency of using to solve the specified problem taking into account the evaluations of respondents of focus groups is shown. Considering the inadequacy of the assessment, which was given by the respondents of the focus group only "verbally", it is impossible to unequivocally determine the level of quality of the presented multimedia content. Taking into account the indicated disadvantage, the task of developing a methodology for evaluation the quality of multimedia content based on the it's emotional perception by individual respondents of the focus group becomes actuality. On the basis of the proposed method of detection, fixation and classification of emotions, a software system for evaluating the quality of multimedia content has been developed.

Keywords

Multimedia content, evaluation of content quality, emotions, focus group, respondent

1. Introduction

Multimedia content, as a collection of interactive content data in video and photo format, in combination with audio, plays an important role in the everyday life of every person. Without thinking about it, every day we receive such information from various aspects of our lives in the form of advertisement, which we don't attach so much importance. Passing by billboards, listening the favorite radio station or watching TV, we get advertisement content. All these, in our opinion, insignificant factors leave an imprint in our memory. And at the next time when we will be in a store or looking for a necessary service, imprints of this media content will appear in our subconscious [1]. The main purpose of which is the formation a positive opinion about this product or company, and as a result – it is to encourage a potential buyer to purchase the product that was advertised.

2. Related works

Multimedia content, as known, is a combination of various forms of information presentation, such as text, graphics and sound [2-4]. In order to correctly assess a person's emotions while watching a video clip, we need high-quality multimedia content that has good recording quality, sound accompaniment, detailed information about the product and the absence of elements that would distract attention from the main thing. Such multimedia content is an important component for obtaining correct data when evaluating a given multimedia content by a formed focus group. The absence of one or more elements of high-quality content can lead to a misinterpretation of a person's emotions. The method of

COLINS-2023: 7th International Conference on Computational Linguistics and Intelligent Systems, April 20–21, 2023, Kharkiv, Ukraine
EMAIL: spivak.iruna@gmail.com (I. Spivak); msya220189@gmail.com (S. Krepych); fedorov.oleks@gmail.com (O. Fedorov); spivaksm@ukr.net (S. Spivak)

ORCID: 0000-0003-4831-0780 (I. Spivak); 0000-0001-7700-8367 (S. Krepych); 0000-0002-8080-9306 (O. Fedorov); 0000-0002-7160-2151 (S. Spivak)



© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

evaluating content by a focus group created for this purpose is widely used in marketing and sociological research. The purpose of the focus group is not to reach a group consensus, but to find out the directions of opinion of all participants of the process. Both positive and negative evaluation results are taken into account, as each person has their own individual vision for a given advertising product. The opinion of the focus group respondents is very important, as it reflects the percentage correlation of interviewed people with a positive result to those whose result is unsatisfactory for the customer. On the basis of these data, it is possible to make more complex calculations, from which it is possible to determine whether this product will be in demand among real buyers. The higher quality of the multimedia content, the better result of the focus groups, so these two factors are dependent on each other [5-8]. The disadvantage of such evaluation of multimedia content by respondents is the human factor, which is explained by giving a vague assessment of one's own needs. Therefore, the task of eliminating the factor of "vague", "verbal" and "overestimated" assessment by taking into account the respondents' emotional perception of the offered content becomes actuality.

3. Overview of the research

For the successful making of any marketing decision, first of all, work is carried out with the target audience interested in this content. The main goal of the audience is to receive quality content, which is reflected in the corresponding emotions of "satisfaction" from the content. The main goal of the customer is to make a profit from the implementation of its content, which directly depends on the satisfaction of the target audience with this content. So, in view of the above, the main lever of success is human emotions.

Let's mark E_m – the number of video fragments selected for analysis for a certain emotion; E_r – the number of video fragments proposed for analysis, rated by a specific respondent of the focus group with the satisfaction of the specified emotion.

Hence, the main goal function of the task of evaluating the quality of multimedia content based on the emotional perception of the respondent is the function:

$$E \xrightarrow{j} \max, \quad j = 1..N, \quad (1)$$

where j – one of videos fragment proposed for analysis and evaluation the respondent's emotional perception; N – total number of fragments.

The formula for calculating the value of the goal function is represented by the equation (2).

$$E = \frac{\sum_{j=1}^N E_s_j}{E_m} \times 100\% , \quad (2)$$

where $E_s_j = E_m_j \wedge E_r_j$.

4. Proposed approach

The process of evaluation of multimedia content represented in Figure 1.

All focus group respondents will be shown the same video content. At the first stage, those respondents whose interest is less than the established indicator (for example, 10-15% of attention) are separated out. To do this, first of all, we will convert the fixed duration of the video into a percentage. The next stage will be to establish the length of time during which the respondent looked away from the screen when this content was shown to him. If the respondent did not watch the content for more than 10-15% of the total duration of the video, indicates that the shown information is not interesting for his perception. Accordingly, the evaluation of his emotions will not give a positive result. Otherwise, when the system captures more than 85% of the respondent's attention capture, we will evaluate their emotions while viewing.

In the Figure 2 in red shows schematically the time when the respondent does not pay attention to the multimedia content shown to him. Accordingly, the time when the respondent follows the content is in black.

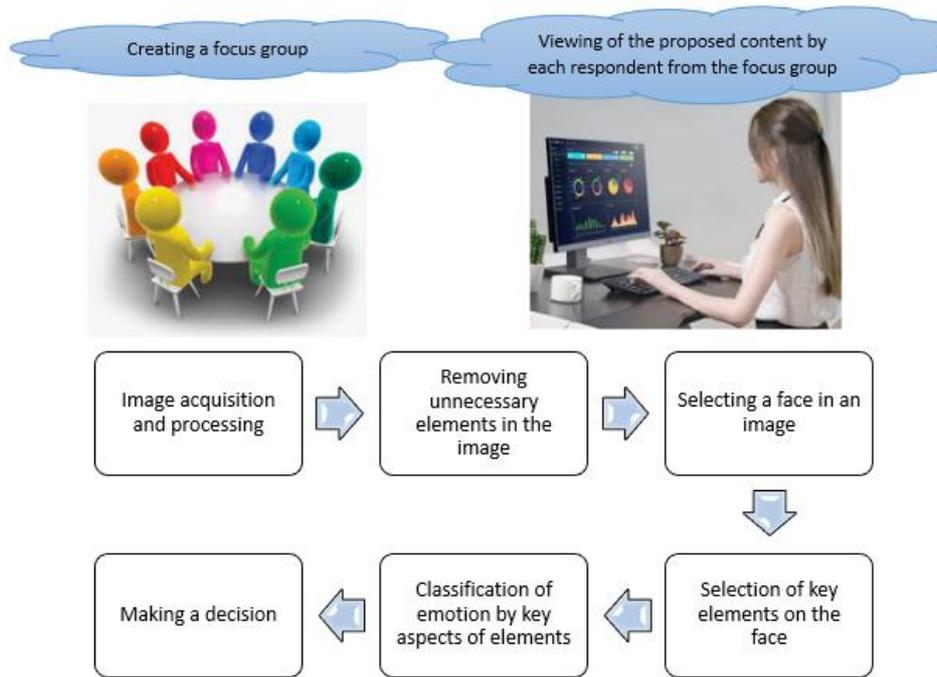


Figure 1: The basic algorithm of automatic recognition of emotions

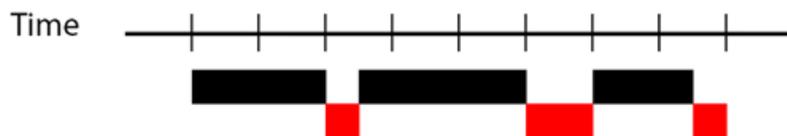


Figure 2: The time process of capturing the respondent's attention while viewing the content

To evaluation human emotions while viewing the content, we will capture the human face every 0.5-1seconds, using the functions of the CSS filter and the HSL (Hue Saturation Luminance) color model [9-12]. The duration of the video recording of a person from the focus group should correspond to the duration of the multimedia content.



Figure 3: Converting the image to grayscale

The next stage is the evaluation of the respondent's emotions, recorded at a certain moment, based on which the analysis of the obtained data will take place on seven key emotions [13-16].

By getting a grayscale image of the moment from the video, the next step is to convert it into a multi-level array, where each pixel will correspond to its value in the array, on the basis of which the evaluation of the shade of black by its illumination is carried out.

As we can see in Figure 5, graphically representation the face using black tint filtering, we were able to remove the unnecessary background, and also mark the main points on the face that interest us, namely the mouth, eyes, eyebrows. It is at these key points that a person's emotions are fixed.

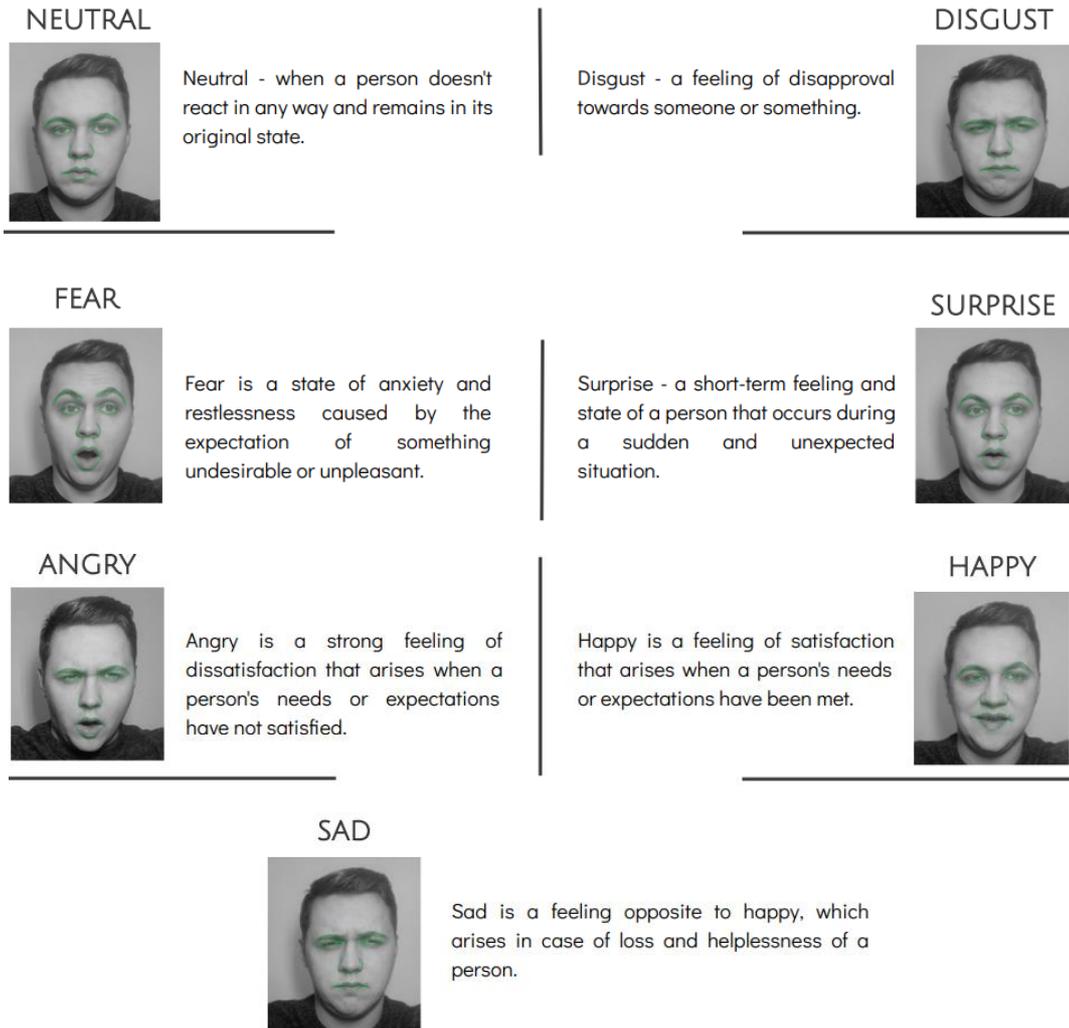


Figure 4: Detection of emotions on a human's face



Figure 5: Graphical presentation of a multi-level array using black tint filtering

The next stage is the recognition of a specific emotion. The development of emotion recognition in the vast majority of methods occurs in two steps. In the first step, functions are defined from fixed images, and in the second step, emotions are detected with the help of already developed classifiers. The most common are neural networks, local binary patterns (LBP) and Haar features. As a result, the system assigns the most appropriate emotion to the image. However, these approaches require a lot of time and resources for training, which limits their use in a large sample of input data. Therefore, for our software system, we use the method the nearest neighbor, which is a search algorithm for the closest neighboring node with a larger numerical value. The method starts with the central node (in our case, it is the smallest value), and then sequentially passes the neighboring nodes around it. This allows you

to find the nearest neighbor with the largest numerical value (darker shade of the pixel) by comparing the values with the previous ones.

The next step is to set a threshold for emotions. To do this, we need to analyze a person in his neutral state to determine the standard (sample image). The captured still image is compared with the standard (neutral emotion) and this method determines the change of the key elements of the face (Figure 6).



Figure 6: Sample image and captured still image

This figure shows the emotion of surprise (right) and neutral emotion (left). The method finds differences at defined key points, for example, more open eyes and mouth. Accordingly, the number of pixels from edge to edge will change, which will lead to a change in the state of a person from neutral to another. The following figure shows how the method sees the image.



Figure 7: The fragment of the table of numerical values for the recognition of the human's eye

Using the example of an open mouth, we will show how software security identifies what emotion is present on the face.

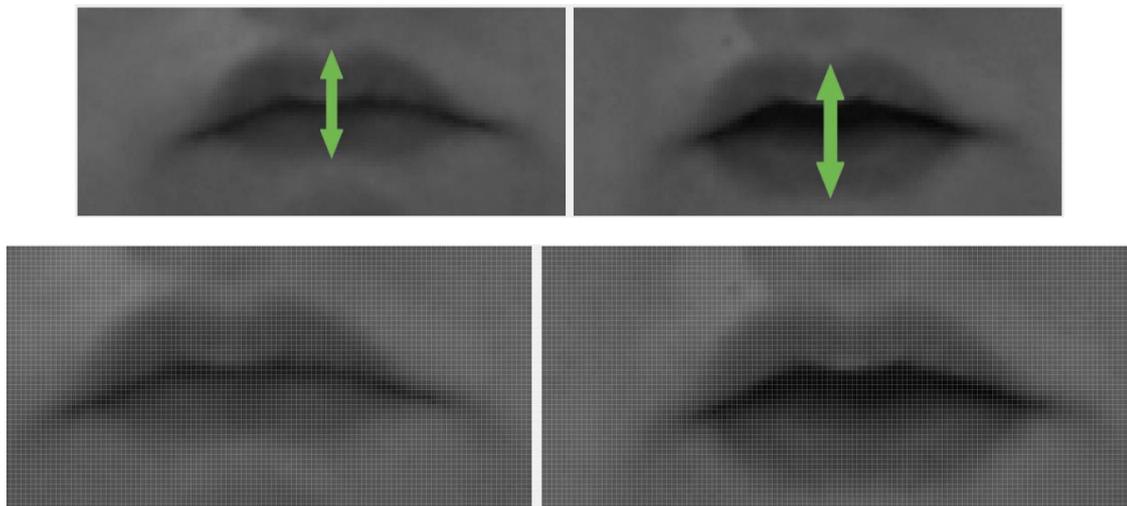


Figure 8: An example of an open mouth (pixel image) to identify an emotion

By comparing the still frame with the standard (left side) and the emotion that the person has at a certain moment, you can also see that the height of the vertical dark pixels increases from 127, which are present in a neutral state, to 157 during a certain emotion. An increase in the height of dark pixels around the mouth means the transition of a person's state from neutral to another. This increase refers not only to one emotion (it can also be fear). At the same time, the program rejects those options where the height does not increase (for example, sad).

For the final determination of which emotion is present on the face, the program determines changes in other key elements, such as the expansion or not of the eyes, a change in the position of the eyebrows [17-20].

Having completed the process of frame-by-frame recognition of emotions while viewing the offered multimedia content, the software system aggregates them to determine the percentage of representation of a certain emotion from the total set. Having such a percentage ratio, it is possible to assess the advantages of the offered content and, accordingly, they can be strengthened in the future. And also evaluate those moments of the content that did not interest or upset the respondents and, accordingly, make a decision on their elimination (Figure 9).

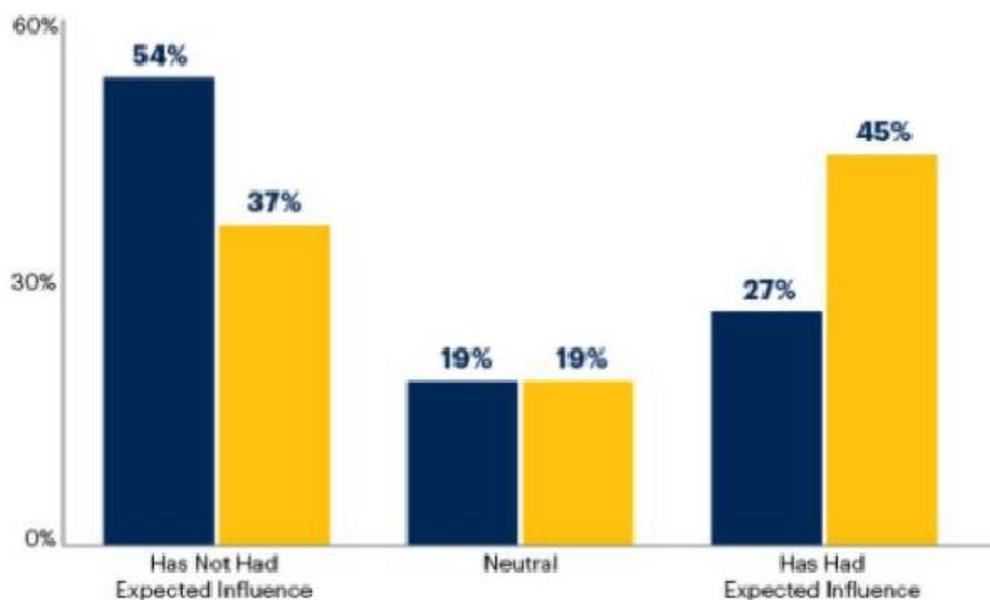


Figure 9: The influence of human emotions on marketing decision-making

For example, if we take a group of emotions such as happy or fear (blue columns in the diagram), we will see that they have a weak influence on the adoption of new or radical changes in proposed marketing solutions. At the same time, a group of emotions such as surprise or disgust have a significant impact on changing marketing decisions, as they immediately determine the strengths (surprise) and weaknesses (disgust) of the content.

5. Results and discussion

For our experiment, let's take a video of the presentation of a new mobile phone. When viewing this content, we will record a video of a person's face to capture emotional changes at a certain moment and equate them to the standard (neutral state). The duration of the video recording of a person from the focus group while viewing multimedia content in our case is 1 minute 32 seconds. However, the special interest of the customer lies in obtaining the desired reaction of the respondent to individual moments of the offered product, with the help of which his product will be a winner on the market. Therefore, special attention was paid to the following fragments of the video:

1. From 25 to 30 seconds – the expected reaction of "surprise";
2. From 51 to 70 seconds – the expected reaction of "sad";
3. From 73 to 78 seconds – the expected reaction of "surprise".



Figure 10: Fixation of viewing of multimedia content

Since the respondent knew that would be represented a phone to him, we do not see any emotions, except for a neutral one, in the first seconds of recording the viewing of multimedia content (Figure 10). At the moment when the video showed the back of the phone and how it lights up, the person's emotions changed (Figure 11).

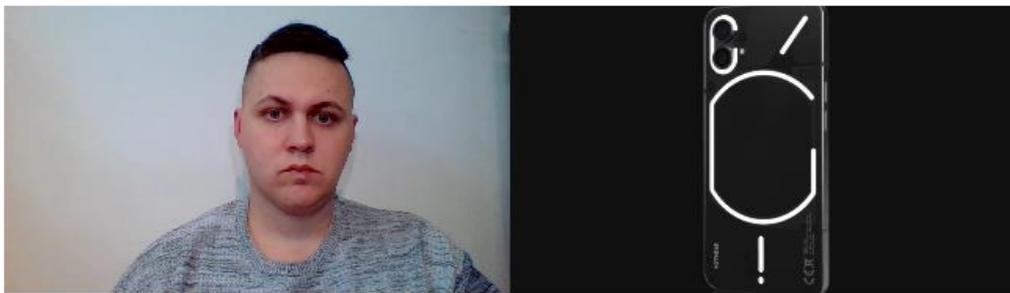


Figure 11: Fixation of the emotion of "surprise" while viewing multimedia content

In Figure 11, we can see that the person's eye circumferences have increased, and due to this, we can assume that the person has the emotion of "surprise".

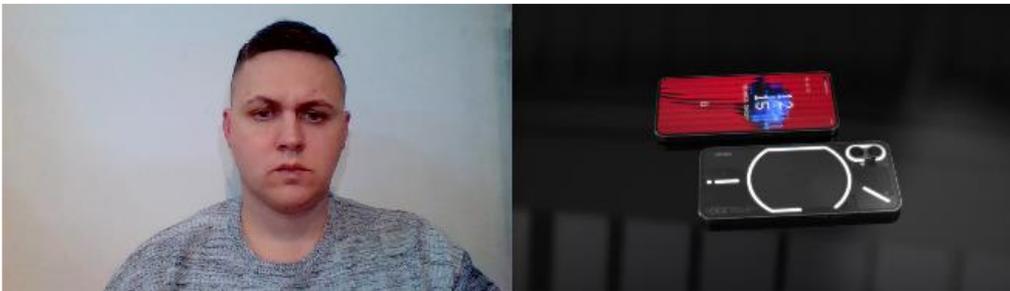


Figure 12: Fixation of the emotion of "sadness" while viewing multimedia content

From time to time viewing multimedia content, and receiving information from the viewer about the lack of radical changes in the design of a mobile device compared to other brands and models, a person reads the emotion of "sadness", when the eyebrows move down and to the center, appear vertical folds between the eyebrows.

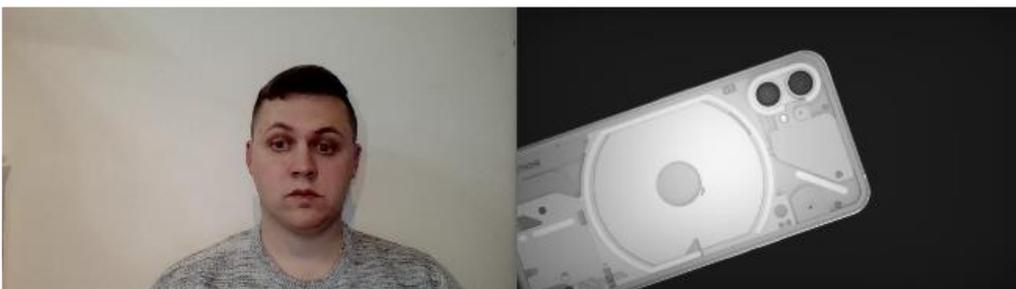


Figure 13: Repeating fixation of the emotion of "surprise" while viewing multimedia content

The emotion of "surprise" was very well read again, when the back side of the device was better shown on the multimedia content and it became clear to the person that the phone's case is transparent. Through it, we can look at all the internal details, wireless charging, soldering elements, and more. Here the person has a clearly open mouth, wide open eyes and raised eyebrows, which indicates only this emotion.

Comparing the duration of the video recording with the storyboard of the respondent's emotions, the following conclusions can be drawn that the main emotion while watching the video was "neutral". However, the most interesting thing is the assessment of the expected emotional perception of the proposed fragments of the video series.

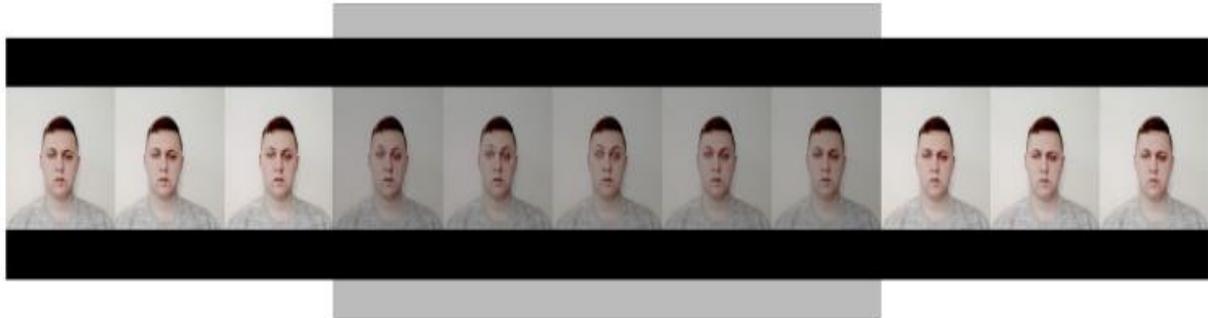


Figure 14: Frame-by-frame breakdown of fixed emotions

Figure 14 shows an example of a second-by-second photo of the respondent's reaction to the offered content from 23 to 33 seconds. The emotional series begins with a neutral reaction and changes to surprise at 26 seconds. This emotion lasts up to 30 seconds and then begins to change to neutral. That is, in a percentage ratio of 83%, the expected emotion was reproduced by the respondent.

The table below shows the results of the goal function calculation by formula 2 for this example.

Table 1

Tabular representation of the value of the goal function for certain video fragments

Interval, sec	j	Em_j	Er_j	Es_j	E
	25	1	0	0	
	26	1	1	1	
[25;30]	27	1	1	1	83,3%
	28	1	1	1	
	29	1	1	1	
	30	1	1	1	
	51	1	0	0	
[51;70]	...	1	0	0	40%
	70	1	0	0	
	73	1	0	0	
[73;78]	...	1	1	1	83,3%
	78	1	1	1	

According to the results of the experiment, the obtained data can be presented to the customer both in tabular (detailed) form and in graphic form, as shown in Figure 15.

On the graph presented, the customer can see a complete picture of the respondent's emotional perception of the entire video content. In the future, it is planned to strengthen this mechanism for evaluating respondents' emotions by introducing interval methods of data analysis [21-22], which will allow obtaining more accurate and adequate evaluation data that will take into account possible evaluation errors.

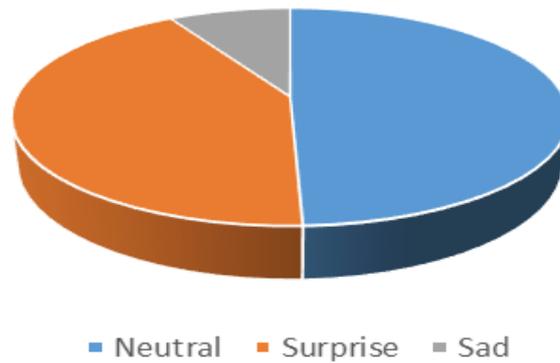


Figure 15: Demonstration of the results of evaluation of one respondent's emotions

6. Conclusions

The article considers the problem with improving the quality of multimedia content that could satisfy the expectations of both users and customers. It is noted that usually the quality of multimedia content evaluates by a focus group formed in a certain way, and this assessment is mainly “verbal” or “proposed”. Taken into account the fact that a person is usually guided by emotions, it is proposed to consider it to improve the process of assessing the content. Based on the algorithms for recognizing emotions, working with video content, etc., a method of evaluation the quality level of multimedia content based on the emotional perception of the focus group is proposed. According to the experiment based on the indicated method, it is obvious that the level of interest of proposed content is relatively lower than expected by customer. Of course, this is the result of only one respondent, and in the future, it is planned to compare the estimates obtained on the basis of the emotions of all respondents of the focus group to create a more adequate vision of “success” of the content. However, even by the example of one respondent, the client can form some proposals to improve the emotional component of the proposed content.

7. References

- [1] S. Porcu, S. Uhrig, J.-N. Voigt-Antons, S. Möller and L. Atzori, "Emotional Impact of Video Quality: Self-Assessment and Facial Expression Recognition," 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), 2019, pp. 1-6, doi: 10.1109/QoMEX.2019.8743186.
- [2] X. Wang, L. Cao, Y. Zhu, Y. Zhang, J. Jiang and S. Kwong, "Study of subjective and objective quality assessment for screen content images," 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 750-754, doi: 10.1109/ICIP.2017.8296381.
- [3] H. Yang, Y. Fang and W. Lin, "Perceptual quality assessment of screen content images", IEEE Transactions on Image Processing, vol. 21, no. 11, pp. 4408-4421, 2015.
- [4] K. Gu, S. Wang, H. Yang, W. Lin, G. Zhai, X. Yang, et al., "Saliency-guided quality assessment of screen content images", IEEE Transactions on Multimedia, vol. 18, no. 6, pp. 1098-1110, 2016.
- [5] I. Spivak, S. Krepych, V. Faifura, S. Spivak, Methods and tools of face recognition for the marketing decision making, in: Proceedings of IEEE International Scientific-Practical Conference: Problems of Infocommunications Science and Technology, PICS&T '19, Kyiv, Ukraine, 2019, pp. 212–216.
- [6] Mandal, P. C. (2021). Public policy issues and technoethics in marketing research in the digital age. *International Journal of Technoethics*, 12(1), 75-86. doi:10.4018/IJT.20210101.o7.
- [7] Maculotti, G., Ulrich, L., Olivetti, E. C., Genta, G., Marcolin, F., Vezzetti, E., & Galetto, M. (2022). A methodology for task-specific metrological characterization of low-cost 3D camera for face analysis. *Measurement: Journal of the International Measurement Confederation*.

- [8] M. Mozafari, R. Farahbakhsh and N. Crespi, "Content Similarity Analysis of Written Comments under Posts in Social Media," 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), 2019, pp. 158-165.
- [9] Y. Kuldeep, S. Joyeeta, Facial expression recognition using modified Viola-John's algorithm and KNN classifier, *Multimedia Tools and Applications* (2020). doi: 10.1007/s11042-019-08443-x.
- [10] Spivak, I., Krepych, S., Fedorov, O., Spivak, S. (2021). Approach to recognizing of visualized human emotions for marketing decision making systems. Paper presented at the CEUR Workshop Proceedings, 2870 1292-1301.
- [11] V. Chirra, S. Uyyala, V. Kolli, Facial Emotion Recognition Using NLPCA and SVM, *Traitement du Signal* (2019) 13-22. doi: 10.18280/ts.360102.
- [12] S. Adeshina, H. Ibrahim, S. Teoh, S. Hoo, Custom Face Classification Model for Classroom Using Haar-Like and LBP Features with Their Performance Comparison, *Electronics* 2021. doi: 10.3390/electronics10020102.
- [13] K. He, X. Zhang, Sh. Ren, J. Sun, Deep residual learning for image recognition in: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR '16*, Las Vegas, NV, USA, 2016, pp. 770–778.
- [14] V. Upadhyay and D. Kotak, "A Review on Different Facial Feature Extraction Methods for Face Emotions Recognition System", in *Fourth International Conference on Inventive Systems and Control (ICISC)*, 2020, pp.15-19.
- [15] Hasan, T. T., & Issa, A. H. (2020). Human facial aggressive detection system based on facial-width-to-height ratio. *IOP Conference Series: Materials Science and Engineering*, Volume 978, 3rd International Conference on Recent Innovations in Engineering (ICRIE 2020) 9-10 September 2020, Iraq.
- [16] Y. Tao, S. Huo and W. Zhou, "Research on Communication APP for Deaf and Mute People Based on Face Emotion Recognition Technology", in *2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, 2020, pp.547-552.
- [17] M. Shahabinejad, Y. Wang, Y. Yu, J. Tang and J. Li, "Toward Personalized Emotion Recognition: A Face Recognition Based Attention Method for Facial Emotion Recognition", in *16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, 2021.
- [18] G. G. Dordinejad and H. Çevikalp, "Face Frontalization for Image Set Based Face Recognition", in *30th Signal Processing and Communications Applications Conference (SIU)*, 2022.
- [19] J. Ueda and K. Okajima, "Face morphing using average face for subtle expression recognition", in *11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2019, pp.187-192.
- [20] G. Guo and N. Zhang, "What Is the Challenge for Deep Learning in Unconstrained Face Recognition?", in *13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp.436-442.
- [21] A. Dyvak, M. Dyvak, S. Krepych and I. Spivak "The method of providing of functional suitability of elements of the device of formation of signal in electrophysiological way of classification tissues surgical wound", in *Proceedings of XIIIth International Conference on Perspective Technologies and Methods in MEMS Design MEMSTECH'2017, – Lviv, 2017 – pp.183-186.*
- [22] I. Spivak, S. Krepych and R. Krepych, "Research of the agree of experts' evaluations in the estimation of software systems", in *CEUR Workshop Proceedings*, 2300, p. 203-206.