How to evaluate emotional experiences in television drama series: Improving viewer evaluations by psychophysiological measurements and self-reports

Lene Heiselberg
Danish Broadcasting Corporation
Emil Holms kanal 20
0999 Copenhagen C, Denmark
lehe@dr.dk

Thomas Bjørner
Aalborg University
A.C. Meyersvænge 15,
2450 Copenhagen SV, Denmark
tbj@dr.dk

ABSTRACT

This paper describes experiences with new methodological approaches utilized in early viewer evaluations of a public service television drama series. We would like to challenge self-reports as the only source of data in viewer evaluations of television drama series. Based on two viewer evaluation studies of drama series, we outline the results and pros and cons of electroencephalography (EEG, EMOTIV Epoc, 14 channels) and skin conductance (SC, varioLAB-mini) measurements tested in combination with in-depth interviews and written self-reports. The main contribution within this study is how a one-hour television drama can be measured in an applied research setting. We suggest a research design consisting of skin conductance measurements combined with written self-reports as the preferred method for viewer evaluations of television drama series.

CCS CONCEPTS

• **General and References** → **Cross computing tools and techniques**; empirical studies; measurement, evaluation, validation • **Hardware** → Methodologies for EDA

KEYWORDS

Emotions, evaluation methods, EEG, skin conductance, mixed methods, self-reports, television

ECCE'18, September 5–7, 2018, Utrecht, Netherlands

ACM ISBN 978-1-4503-6449-2/18/09…$15.00
https://doi.org/10.1145/3232078.3232093

1 INTRODUCTION

Drama series from the public service institution the Danish Broadcasting Corporation (DR) include *Borgen* (2010-2013), *The Killing* (2007-2012), *The Legacy* (2014-2017), *The Bridge* (2011-present), and *Rides Upon the Storm* (2017-present). The success of DR's drama series has already been well-documented [1]. Little attention, however, has been given to DRs early viewer-evaluation process, which is part of the creative development process before a series airs. For many years, DR Audience Research used focus groups for viewer evaluations of new drama series, but when analyzing statements from the focus group respondents it became clear that they were generally too rational (discussion of the plot, the acting, mistakes in scenes) and rarely included respondent's emotional experiences. Hence, the research question within this paper is; how can early viewer evaluations of emotional experiences in television drama series be improved?

Emotions are often the most important focus in viewer evaluations of drama series, as it can be crucial for maintaining viewer interest. However, with self-reports as the only source of data, researchers get a limited picture of the emotional experience the television drama series provides. The problem is that emotional experiences and self-reports are, in many ways, a mismatch. There are three main reasons for this: First, it can be extremely difficult getting respondents to register, evaluate, or simply talk about their emotional experiences, as they are not always readily accessible from their consciousness [2-3]. Second, the understandings of viewers' emotional minds are most often based on verbal stimuli as the gateway to emotions or verbal reports to measure emotions [3]. Third, respondents usually provide self-reports after their exposure to stimuli, and the self-report thus represent a summary of the whole experience, not the here-and-now emotional experience [4].

2 PREVIOUS RESEARCH

Psychophysiological methods have been used for various research topics related to media effects, especially watching violent television [5-6] and responses to television commercials and brand memory (including loyalty and identification of target groups) [7-10]. There is an emerging area of viewer evaluation research using a mixed methods approach, combining self-reports and neuro-scientific methods [8]. However, there are only a few examples of research consisting of a psychophysiological approach in combination with self-reports in which a full-length television drama series is the stimulus [9]. Besides, research in methodological reflections within early stage viewer evaluations of television drama series is limited [3, 11].

3 STUDY DESIGN

This study is based on two different viewer evaluations. Viewer evaluation 1 used EEG measurement and in-depth interviews. Viewer evaluation 2 used SC measurements and written self-reports. We initiated with the psychophysiological measurements followed by self-reports, giving equal weight to both methods. The main argument for this order was to use the EEG or SC data to qualify and structure the subsequent self-reports about preselected scenes from the drama series, followed by final analysis, interpretation, and reporting. This means that the interviewer could follow up on scenes of the drama series that activated the respondents emotionally and those that did not, based on the EEG or SC data. An advantage was also to have the psychophysiological measures and interviews in separated phases.

The recruitment procedure was the same for both viewer evaluations. Respondents were recruited by quota sampling through DR Audience Research's own Internet panel consisting of 10,000 self-recruited viewers. The criteria for selection were age, geography, and genre preferences. This means that we wanted both males and females in different age categories (18–35, 36–60, and 61+), all of whom should have positive genre preferences for television drama series. To minimize geographic spread, participants were recruited from the Copenhagen area. Viewer evaluation 1 had 28 respondents (16 females, 12 men. Age: Mean = 48.2). Viewer evaluation 2 had 46 respondents (23 females, 23 men. Age: Mean = 44.0). Informed consent were obtained, and we followed strict ethical guidelines. We also made sure to recruit respondents without a pacemaker.

4 VIEWER EVALUATION 1: EEG AND INTERVIEWS

4.1 Stimulus

The stimulus was the first episode of the television drama series The Legacy (DR1, 2014-2017) in full-length (one-hour). The Legacy is a modern family portrait and takes place at a farm estate that is the residence of the internationally renowned artist Veronika.

4.2 Procedure

Viewer evaluation 1 consisted of an EEG measurement (EMOTIV Epoc, 14 channels, sampled to 128 Hz, following the 10-20 standard for electrode placement system), followed by narrative-inspired in-depth individual interviews. The EEG measurement provided a measure of arousal—the intensity of the emotional experience [4, 8]. For the EEG measurements, each electrode was disinfected with salt water before being attached to the scalp. Respondents were seated in a comfortable chair that was 60 cm from the 24 inches computer screen. A calm and undisturbed environment was created for the screening. Data corrections (noise filtering) were made with independent component analysis (ICA), and sample frequencies by fast Fourier transform (FFT), and MATLAB for the data processing.

In the interviews, the respondents rescreened seven scenes, which were selected based on the aggregated data from the EEG measurement. Three of the scenes contained the highest levels of arousal. Two scenes contained the lowest levels of arousal, and the last two scenes were the ones that many respondents mentioned in a questionnaire. The questionnaire consisted of two open-ended questions. 1. Scenes respondents felt affected them the most and why. 2. Descriptions of events respondents had hard time concentrating on and why.

4.3 Results

The most common reflections in the interviews were within the episodic memory. This can be explained by the stimulus [11], since the first episode of *The Legacy* is characterized by being highly emotional in content. There are many examples where the respondents revealed how specific scenes were related to their own experiences, and this might be the reason why they were emotionally affected: E.g. 'I once had a girlfriend, who was adopted. I know what this means to her' (ID20, male, aged 44). However, there were also many examples of participants identifying with a personality trait of a character or a mood or emotion that a character exhibits, such as: 'I am just as quiet as Signe in this scene' (ID15, female, aged 25). 'I would be just as disappointed as Signe' (ID20, male, aged 44).

Respondents do not remember and cannot self-report on scenes with low levels of arousal (Figure 1). 'Signe's confrontation' achieves the highest levels of arousal and is also the scene about which most respondents made self-reports, which is similar to previous findings [11].

However, one scene 'Emil is notified (about the death of his mother') surprised us by not following this pattern. The scene achieved the second-highest levels of arousal (Figure 1), but no respondents recalled it in the self-reports.

Methodologically this scene is interesting because it shows that there is a need for integrating interviews with psychophysiological measurements to get the full picture. The scene notifying Emil about Veronika's death would not have been discussed in a traditional focus group setting because no respondents would have pointed out that it had made an impression.

Scene	Arousal level	Scene duration	No. recall
Signe's confrontation	High	1.40	8
Emil notified	High	0.50	0
Veronika's heart attack	Medium	0.30	5
The gameplay	Medium	2.20	3
Frederik notified	Low	0.35	0
Signe's Transport	Low	1.10	0

Table 1: Overview of the scenes with high, medium, and low levels of arousal and an overview of the scenes that most respondents mentioned in the self-reports.

However, with psychophysiological measurements, it turns out that the scene activated the respondents' emotions to a high degree. Based on the psychophysiological data, the scene was taken into the interviews, where many emotional comments were made on this specific scene.

5 VIEWER EVALUATION 2: SC AND WRITTEN REPORT

5.1 Stimulus

The first episode of the drama series *Rides Upon the Storm* (DR1, 2017) was shown in full-length. The series focuses on a family of priests in the Church of Denmark. The series deals with the concept of faith in a broader context.

5.2 Procedure

The SC measurements were conducted in a lab with three workstations. The respondents were seated at a workstation that consisted of a 17-inch laptop with high quality headphones, an external keyboard, and a mouse. The laptops were fitted with Tobii X2-30 Eye Tracker Compact Edition eye-trackers. The lab was windowless and air-conditioned, and the walls were unadorned to minimize interference from external stimuli. A small intersecting wall was set up between the workstations. Each respondent was fitted with a varioLAB-mini, a device to measure skin-conductivity levels (henceforth referred to as EDA - electrodermal activity), including a signal amplifier using standard disposable AAMI-electrodes (Covedien/Kendal H124SG, Ag/AgCl, ø24mm, pre-gelled adhesive tape). The SC measurement was an exosomatic measurement with a direct current. Prior to watching the stimulus, a short instruction was shown followed by a baseline sequence: a short series of pictures representing low and high arousal stimuli of positive and negative valences. A five-second black screen was inserted between the baseline and the actual episode to be tested, as previous experiments have shown that respondents are aroused at the beginning of a screening because of the unusual context. The electrodes were attached approximately one minute before the SC recording began. To uncover the conscious emotional experience of the drama series, the respondents were instructed to fill out a questionnaire after the screening ended. The questionnaire consisted of two open-ended questions. 1. Scenes respondents felt had affected them the most and why (recall elicited strong emotions/ self-reported high arousal). 2. Descriptions of events respondents had hard dime concentrating on and why (recall boredom/self-reported low arousal or flatliner sequences).

5.3 Results

Figure 1 shows the phasic and tonic component of the EDA signal for all respondents watching the first episode of *Rides upon the Storm*. The high peak in the phasic component at the beginning originates from a superficial elevation due to instructions prior to watching the show. We therefore do not draw any conclusions based on the first two minutes of the episode. The scenes with the highest levels of arousal occur at the beginning (400-second timestamp) and towards the end (3,500-second timestamp). In the first peak, the two sons of the priest, Johannes, are driving, and the brother behind the wheel decides to close his eyes, making his brother extremely excited and scared.

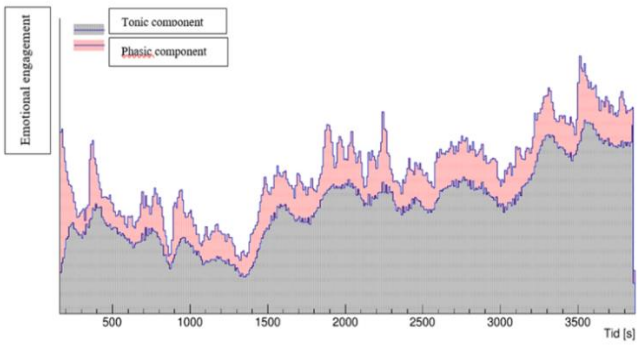

Figure 1: Phasic (red) and tonic (grey) component of the EDA signal for all respondents.

The last high peak contains a narrative, where Johannes conducts a funeral while he is drunk. At the grave, he says

a prayer, but before he is done he stumbles and falls into the grave on top of the coffin. In the middle, around the 2,000-second timestamp, four rather high peaks succeed each other. This is a build-up of narrative events which ends in a scene where the Johannes realizes he lost the bishop's election. The scene of Johannes' fall into the grave and the election defeat are two scenes many respondents self-reported on, but the scene where the sons of the priest drives madly does not appear in any of the self-reports. As we saw in viewer evaluation 1, some scenes were not captured by the respondent's self-reports.

6 CONCLUSION

Based on the outcome of a methodological reflection on the two empirical studies conducted in an applied research setting, we recommended a mixed-method design consisting of 1. Data collection with SC-electrodes performed while the respondents watching an entire episode of a television drama series; and 2. Written self-reports conducted after the screening of an episode of a drama series has ended to access the subjective experience of the first episode of a drama series.

Looking at the pros and cons of using EDA and EEG it turns out that SC measurements caused only little inconvenience to the respondents compared to EEG. There were fewer disturbances during the SC data collection because no electrodes went out of place, as was the case for the EEG electrodes. Further, SC measurements require less supervision of the participants compared to EEG. As the SC equipment is relatively inexpensive, it was possible to collect data from three respondents at a time, and, consequently, the data collection was finalized in fewer days than with EEG measurements. On the negative side, SC measures are only a proxy for an emotional experience (sweat) and are not a direct response like EEG. Because of the time delay in the SC reaction, SC measurements are less precise. Further, respondents must be seated relatively still during the data collection, and there is individual variation in the SC measurements as everyone has a different starting point or baseline. This means that some people sweat more than others, although they are not affected by a stimulus. To overcome this limitation, each respondent's measure is transformed statistically so that they can be compared. This is done by ranking each measurement's minimum and maximum from 0-1, following the statistical approach of [31]. Additionally, SC can measure attention, activation, and arousal. SC does not measure a single process. Therefore, isolating the aspect of interest has been recommended [32]. Figner and Murphy [32] recommended compensating by trying to limit the possible interpretations, for example, by allowing the respondents to self-report on their experiences. According to Figner and Murphy [32], it is necessary to remain within the framework that is created by the respondents in their self-reports in the interpretation of the fluctuations in the SC measurement. Specifically, this means that in the absence of an additional self-report, it would be impossible to determine whether the changes in arousal caused an emotional impact or whether one is engaged in a cognitively demanding task, unless the respondent is asked. To ensure that the arousal is generated by the fiction series, it is therefore important to combine an SC measurement with self-reports. Another point of attention in relation to SC is that the data may be affected by variations in moisture and temperature. To ensure data quality, it is advisable for all the measurements to be carried out in the same room with air conditioning, thus ensuring an even temperature and a consistent level of moisture [32]. Written self-reports in an online questionnaire provide a reflexive reaction to the emotional experience after the screening of a television drama series ends. Usually, the data are not as deep, detailed, and personal as the oral descriptions from the in-depth interviews. One reason for this is there is not an interviewer to ask follow-up questions about their emotional experiences. In the in-depth interviews, it was possible to focus on an individual emotional experience and to distinguish the differences between the respondent's emotional experiences. But since time restraints are an important issue in applied research, the written self-reports are preferred, since they are able to provide a glimpse into the respondent's subjective experience of a drama series.

REFERENCES

[1] Tobias Hochscherf and Heidi Philipsen. 2017. *Beyond the Bridge: Contemporary Danish Television Drama*. London: IB Tauris.

[2] Daniel Kahneman. 1999. Well-being: The Foundations of Hedonic Psychology. NY: Russell Sage Foundation.

[3] Joseph LeDoux. 1996. *The Emotional Brain*. NY: Simon and Schuster.

[4] Michael L. Rothschild, Esther Thorson, Byron Reeves, Judith E. Hirsch and Robert Goldstein. 1986. EEG activity and the processing of television commercials. *Communication Research*, 13,2, 182-220.

[5] Victor Cline, Roger G. Croft and Steven Courrier. 1973. Desensitization of Children to Television Violence. *Journal of Personality and Social Psychology* 27, 3, 360–365.

[6] D. K. Osborn and R. C. Endsley. 1971. Emotional Reactions of Young Children to TV Violence. *Child Development* 42, 1 (March 1971), 321-331.

[7] Michael L. Rothschild, Esther Thorson, Byron Reeves, Judith E. Hirsch and Robert Goldstein. 1986. EEG activity and the processing of television commercials. *Communication Research*, 13,2, 182-220.

[8] Niklas Ravaja. 2004. Contributions of Psycho-physiology to Media Research: Review and Recommendations. *Media Psychology* 6(2), 193–235

[9] Monika Suckfüll. 2013. Emotion Regulation by Switching Between Modes of Reception. *Psychocinematics* (March 21, 2013), 314–336.

[10] Herbert E. Krugman. 1971. Brain wave measures of media involvement. *J. of Advertising Research*, 11, 1, 3-9.

[11] Alisha C. Holland and Elizabeth A. Kensinger. 2013. Emotion in Episodic Memory. In: *The Cambridge Handbook of Human Affective Neuroscience*, Jorge Armony and Patrik Vuilleumier (eds), 465-491. Cambridge University Press, Cambridge.