Quantifying Audience Experience in the Wild

Heuristics for Developing and Deploying a Biosensor Infrastructure in Theaters

Chen Wang[1], Jacqueline Wong[1], Xintong Zhu[1], Thomas Röggla[1], Jack Jansen[1] and Pablo Cesar[1,2]
[1]CWI: Centrum Wiskunde Informatica, Amsterdam, the Netherlands
[2]Delft University of Technology, Delft, the Netherlands
c.wang@cwi.nl, p.s.cesar@cwi.nl

Abstract**—***Measuring the experience of audience of arts events is essential in the "experience economy" of this day and age, but it is a difficult task. The value of such information goes beyond evaluating the impact of the arts, as it can provide insights and feedback to enhance the work of artists and the experiences of other audience members. Through in-depth understanding of the needs of the providers and consumers of the arts, we progressively developed a biosensor infrastructure that was deployed in theaters. Over the years, we identified the challenges and issues related to developing and deploying a biosensor infrastructure in theaters. These collective experiences and identified issues were categorized into three main areas: processes, data, and system. A total of seven heuristics are developed across the three main areas. Processes place the stakeholders and audiences at the core of the research; data provides guidelines for data validity, collecting a variety of data, and supporting real-time data gathering; and systems covers the concurrency, scalability, deployment and feedback of the infrastructure. We believe that this set of heuristics forms the foundation for an adequate infrastructure to measure audience experience in the wild and it is a valuable source of guideline for future work.*

Keywords—wearable sensors; galvanic skin response; audience experience; theater performances

I. Introduction

The impact of the arts is valuable and at the same time hard to quantify. Hence, research about it is complex and not well understood [1]. Given the heavy investment in time, money and effort needed to put together arts and cultural events like theater productions, dances, and art exhibitions, it is important to understand the audience experience in order to measure the success of the event. The main contribution of this paper is a set of heuristics for developing and deploying in the wild a biosensor infrastructure that captures the fluctuations of audience experience throughout the entire event. The aim is to guide future work and provide a foundation for an adequate infrastructure to evaluate audience experience in realistic conditions.

While arts and cultural activities are often viewed as "nice to have" and non-essential, several studies have shown that exposure to the arts can bring about benefits to individuals and communities on the whole [2]. "Gifts of the Muse" [3] gives a comprehensive account of the numerous intrinsic benefits of the arts. Developing one's confidence and provoking positive thoughts and feelings through engagement in arts and cultural events help to build socially cohesive communities [4, 5]. Studies have also demonstrated that arts and cultural events have an effect on perception of quality of life [6], health and wellbeing [7], and education [8]. Other than the societal benefits, arts and cultural events also contribute to the economy. According to a 2014 study conducted to analyze the cultural and creative markets in the European Union (EU), the creative industries' revenue was €535.9 billion and more than 7 million people were directly or indirectly employed in arts and cultural activities [9]. The arts sector also generates spending in other sectors, like tourism [5]. In view of the potential benefits for the wider society, research in the arts is important to help realize these positive impacts.

Within the arts community, the players involved also seek tangible feedback to enhance their offer. A core mission of arts organizations is to provide arts and cultural experiences to more people, and to deepen and strengthen the quality of their experiences [2]. Our research involves working closely with people from arts organizations like Holland Dance [10], ByBorre [11], and National Theater of China [12] to understand the needs of the community. Conventionally, theaters have used a variety of methods to measure audience experience: questionnaires [13], interviews [14], text messages (through mobile phones) [15] and portable hand-held 'clicker' systems [16]. However, people from the arts community felt that these methods have not been optimal. First, data collected from questionnaires and interviews are fragmented. Arts organizers are unable to make improvements that are targeted. Second, the constant operation of a 'clicker' system or mobile phone during a performance may disrupt the theater experience. In view of the limitations of the research methods being used, there is a need to explore novel methods that can provide a less biased and more complete understanding of the audience experience.

In the same line of research, Latulipe et al. conducted lab studies to measure audience response for recorded dance performance by using questionnaires and Galvanic Skin Response (GSR) sensors [17]. Theater directors and dance choreographers who were shown the data were positive of how the information can be valuable to the production of performances. The study supports the view that GSR is a valid representation of audience engagement during performances [18]. GSR refers to the changes in conductance on the skin surface, reflecting activity within the sympathetic axis of the autonomic nervous system (ANS) [19]. Autonomic responses in the skin, e.g., sweating, piloerection, and vasomotor changes, can thus be elicited by various emotional states via the Papez circuit in the limbic system [20]. It has been recognized that increased GSR can be provoked by attention-related stimuli or tasks [21]. GSR includes two variables. The first one is skin conductance level (SCL), indicating the slow and tonic changes measured across many discrete stimuli. The second one is skin conductance responses (SCR) related to specific stimuli, representing the quick and phasic changes imposed on shifts in tonic level in conductivity [22]. However, the study was conducted in a lab. The environmental settings in lab conditions are rather different from theaters, where the audience is physically immersed in the performing environment with the actors, and they share the theater experience with other people.

Li Dong, Head of the National Theater of China's Centre for foreign collaboration and executive producer for Chinese adaptation of War Horse, foresees the use of physiological data from user experience as a form of feedback mechanism to increase audience satisfaction. Such fine-grained data on individual experiences provides a basis for the different artistic teams to objectively analyze their creative work and make improvements to further enhance their work.

Live visualization of audience experience can also enhance the experience itself. We conducted an experiment where we visualized the audience experience during the performance. The creative director, Borre Akkersdijk, stated, "[…] you can play with the fact that also the audience sees when the rest is losing focus or not… sort of group control." The ability to visualize experiences of other people attending the event can have an effect on their personal experiences. One can share their experiences with others and also feel part of the group. Such effect can enhance the shared experiences.

Previous studies have used physiological sensors and acceleration sensors to investigate audience response in movies [23] and theater performances [24], respectively. The studies effectively used novel methods to evaluate audience experiences. However, other uses of the data have not been explored. Visualizing audience experience can function as a feedback mechanism to performers, effectively providing them with information of how audience members are experiencing their work, especially when audience members are not in view (i.e. audience sitting at upper circles). As mentioned before, the real-time stream audience biofeedback can also enhance the feeling of shared experience.

Motivated by the extensive benefits of the arts for the public and the lack of befitting ways to measure audience response, we envisioned an infrastructure for establishing an efficient participatory biosensor network that can be deployed in theaters. The GSR sensors can simultaneously and independently deliver the data from multiple anonymous audience members directly to a central server that processes the data in real-time. The accompanying system enables not only offline processing of the GSR data, but also real-time analysis of the data allowing for visualization or interactive installations.

To achieve this vision, methodological approaches were followed. Adapting the Convergent-Divergent Model as the first step in the Waterfall approach [25], discussions were initiated with people from the arts community and related work was reviewed to establish a set of requirements that are to be met. Consequently, prototypes were developed and tested using an iterative approach. Emphasis was placed on evaluating prototypes in situ, addressing interdependencies among users, design, technology, and environment [26]. Each experiment built on the lessons learned from previous experiments, expanding our knowledge on conducting audience research in the wild. As a consolidation of our learning, we followed the basic steps similar to [27] in building heuristics. All the issues related to establishing an efficient participatory biosensor network that can be deployed in theaters are identified before

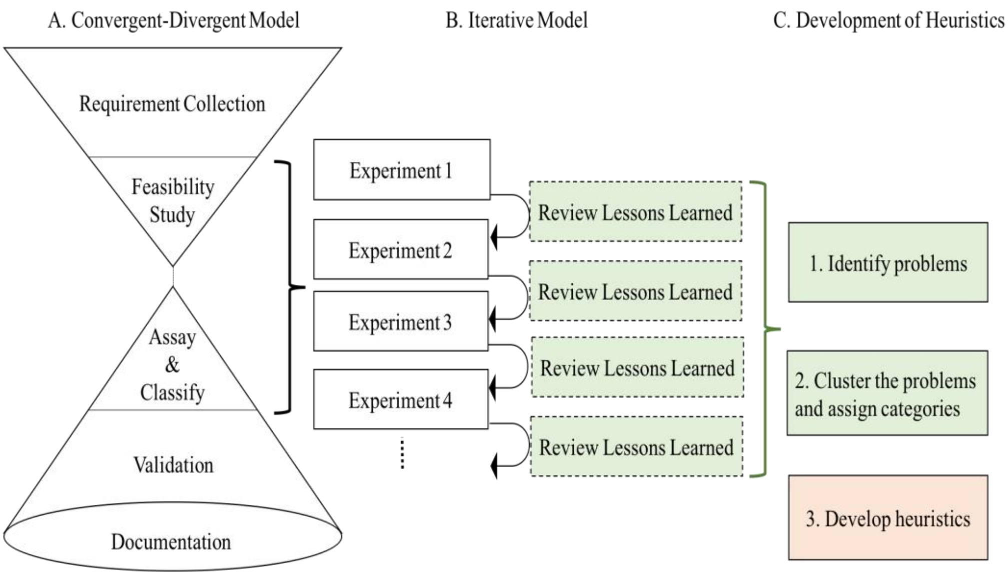

Figure 1. Methodologies adopted in our research

categorizing them, resulting in a set of heuristics. The described design principles are aimed at researchers and practitioners who are interested in novel methods to measure audience experience, serving as a starting point for future work.

II. Methodology

Conducting audience research in the wild is a complex task with many uncontrollable variables like the types of performance, the sizes and types of venue, different populations of audience. All of these may affect the design and deployment of the biosensor infrastructure. Therefore we adopted a hybrid methodology as shown in Figure 1.

For the first step, we adopted a similar process as the one presented by Madgunda and colleagues, the Convergent-Divergent model, consisting of five steps. The first step was collecting requirements by discussing with various users including theater producers, performers, and audiences to understand the requirements to meet for deploying a biosensor infrastructure to measure audience response in a theater. The second step was the feasibility study in which we analyzed the requirements and took steps to develop a prototype. Consequently, the prototype was tested in the wild and evaluations were made to improve the next prototype. Due to the vast differences in performances and environments, the second and third steps were repeated in an iterative process following an agile model. The fourth and fifth steps were validating the requirements and documenting them for later developmental stages.

The iterative methodology similar to an agile model was employed to accomplish steps two and three of the Convergent-Divergent model that resulted in four consecutive experiments. The experiments are described in the next section as the developmental path. Finally, putting together the lessons learned from the four experiments and our personal experiences as experimenters, we used the heuristics methodology to develop valuable knowledge to guide future work. Each of the experimenters identified problems encountered during the experiments individually. Next, the experimenters discussed and eliminated problems that were already mentioned. The remaining problems were clustered and categories were assigned to each cluster. Then, heuristics were developed to address problems at the cluster level.

III. Developmental Path

In this section, we describe the developmental path of building and deploying the biosensor infrastructure in the theaters as shown in Figure 2. An iterative method was used where each experiment resulted in a set of lessons learned that were used to advance the design of the biosensor infrastructure in the following experiment. Concurrently, this methodology completes the two sequential steps two and three in the Convergent-Divergent methodology. The requirements and constraints that were initially gathered through discussions with end users (i.e., theater companies, producers, artists, audience) as well as issues from related work were analyzed and evaluated for feasibility (i.e., Step 2: Feasibility Study). The first prototype was developed to test whether the expected outcomes are achieved; problems encountered were later identified and categorized (i.e., Step 3: Assay and Classify). Due to the many requirements of different nature and the complexity of conducting experiments in realistic environments, where constraints differ across theaters, four subsequent experiments were carried out to test whether the biosensor infrastructure deployed meets the needs of the different end users. Through the four experiments, we gained invaluable experiences that are translated into heuristics.

The initial requirements were to develop a biosensors infrastructure that can be used to measure audience response and to develop the mechanism to analyze the responses for understanding audience experience at different points in time

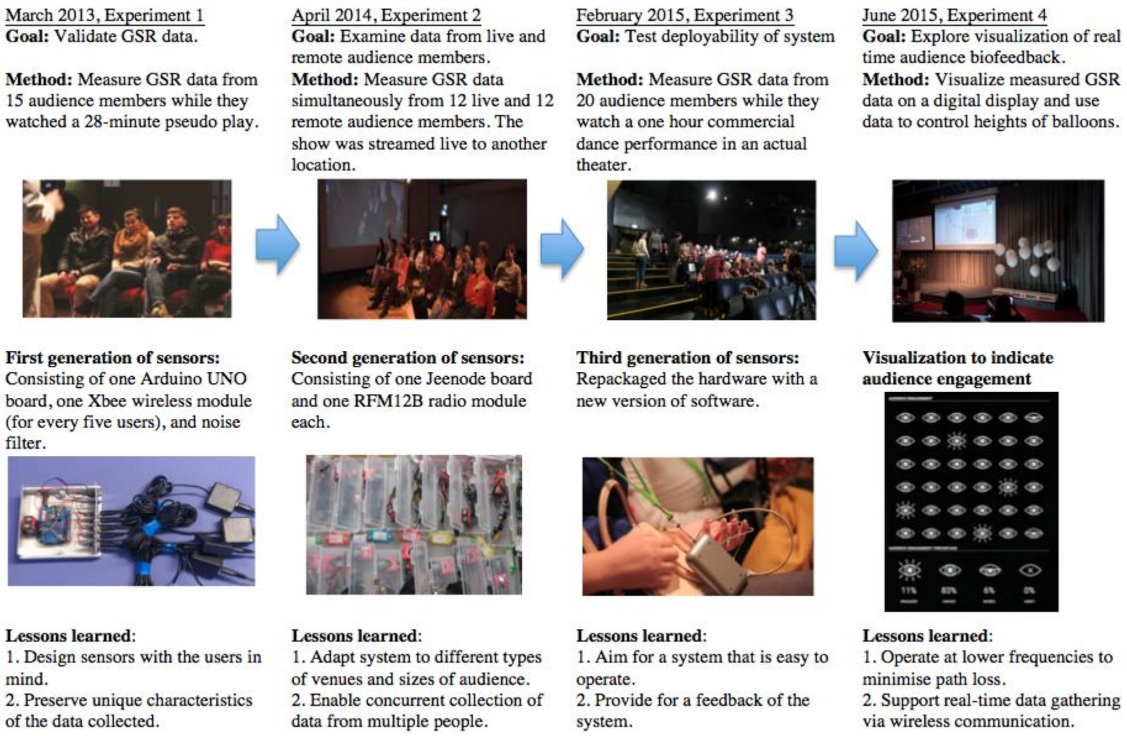

Figure 2. Developmental path of building and deploying the biosensor infrastructure for audience research in the wild

during a performance. The first generation of our sensors consisted of one Arduino UNO board and one Xbee wireless module (for every five users), and noise filter. Five sensors were connected to one main module. As such, we had to cluster audience members in groups of five. Biofeedback of 15 audience members was measured while they watched a play that was specially choreographed to elicit audience response. Based on the measurements that were synchronized with the video recordings of the performance, it was concluded that GSR highly reflect audience response [28].

From the first experiment, we learned that the data collected should preserve the unique characteristics of the signal we are collecting. There are many sources of artifacts that can influence the physiological signals (e.g., types of electrodes and placement of the electrodes). Being aware of these different sources of artifacts would help in processing the data to maintain its validity. Furthermore, priority of the show should come before the collection of data. The first generation of sensors was cumbersome to wear. It is important to respect the audience and the first priority would be their enjoyment of the theater-going experience.

A second generation of sensors was developed to address some of these issues. We worked on scaling up the system with small form-factor developments in wireless technology and GSR measurements. The Arduino UNO board was changed to a Jeenode board so that the infrastructure could support up to 250 different groups. In other words, a larger group could be measured simultaneously. The Jeenode board also works at different frequencies and uses the RFM12B radio module that can operate at long distances with rather low power consumption. By putting together our own biosensor infrastructure, we can experiment with different solutions to create an adequate infrastructure.

The improved sensors were tested in a second experiment intended to investigate the different responses of local and remote audience members attending a live theater play. At each location, each audience member wore an individual wireless sensor. The wireless sensors have the capability to simultaneously and independently send signals to a central server. The results showed a high correlation between the GSR measurements of the local and remote audience members.

Although the hardware design of the second-generation sensors made it more convenient for audience members to wear and we were able to simultaneously collect data from a larger group of audience and in different venues, we discovered that we had to improve the quality of the collected data.

Subsequently, a third generation of sensors was produced. The sensors were tested in a third experiment with 20 audience members watching a live one-hour commercial dance performance. A more advanced algorithm was adapted from Julien et al. [23] to process the data and to find significant SCR points in audience response. The network performance was also improved by setting up a transmission rate that maximized the capacity of the network.

With the third generation of sensor infrastructure developed, we identified new challenges. We had not explored the possibility of using the data to manipulate visuals, so for the next experiment, we wanted to investigate visualizing audience feedback in real-time.

For the fourth experiment, we developed software that processed the captured data in real-time. During this experiment, 36 people wore the third generation of sensors that collected their physiological responses while they listened to a commercial presentation. The physiological responses were visualized on a digital display. The information provided the presenter and other audience members with an indication of audience engagement.

IV. HEURISTICS

Heuristics are like mental shortcuts that enable people to make quick judgments and handle problems. The set of heuristics should provide a wide variety of perspectives on usability and be as good as possible at explaining usability issues that occur in an actual environment [29]. There are many approaches in creating heuristics. Molich and Nielsen [30, 31] based their heuristics on personal experiences and the years of knowledge they have gathered from teaching and consulting others on usability engineering.

A more structured approach was taken by Dykstra [32] to create a set of heuristics using five steps: 1) listing all usability problem for each program and each participant using competitive analysis, 2) consolidating all the problems for each program, 3) categorizing the problems, 4) deleting duplicate problems and combining problems into fewer categories and 5) developing the final heuristics. Similar to Dykstra but an even more simplified version was used by [27]. They used a three-step approach: 1) identify problems from reviews of different users, 2) assign categories to groups of problems, and 3) create the heuristics to address the problems.

Steps from these previous studies were adopted and similar to [27], we derived a set of heuristics in three steps: 1) identify problems through interviews with users, personal experiences, and lessons learnt from our experiments, 2) organize problems into categories by eliminating redundant problems and clustering similar problems into categories, and 3) develop heuristics to help avoid the problems. A total of seven heuristics had been derived and they are further categorized into three main areas: A. processes, B. data, and C. system.

A. Processes

Heuristic #1: Ascertain the goals of the stakeholder

- Recognise the complex structure of the organisations and how the experiment would influence their core business and their routine
- Maintain an open channel of communication with all stakeholders and respect their priorities
- Plan well

Heuristic #2: Respect the audience

- Prioritize the show and not the data gathering process
- Design the sensors with the user needs in mind
- Ensure adaptability for different population members
- Ensure user privacy and feeling of privacy

B. Data

Heuristic #3: Ensure data validity

- Ensure that the collected data reflects the variables you are interested in
- Maintain data timing characteristics across audience members
- Preserve the unique characteristics of the data (e.g., GSR signals may need a different treatment than heart rate signals)
- Be aware of the many sources of artifacts.

Heuristic #4: Create a complete data set

- Collect all data available during the experiments (i.e., video, annotations, interviews, but also system data such as network delay and loss)
- Ensure traceability and replicability
- Ensure that other researchers can use different algorithms to process the same data set for comparison purposes

Heuristic #5: Allow support for real-time data gathering
(Only applicable for certain experiments)

- Enable transmission and processing of a sufficient number of samples in real-time for accurate reconstruction of signals
- In practice, given heuristics #1 and #2, real-time support requires wireless communication
- Be aware that operating at a lower radio frequency reduces path loss

C. Systems

Heuristic #6: Enable concurrency and scalability

- Allow gathering of data from multiple people at the same time
- Allow adaptability to different sizes of venues and audiences
- Do trial runs to ensure that the expected scale is met

Heuristic #7: Aim at deployability and provide feedback

- Be aware of the constraints (e.g. different locations, audiences, performances…)
- Allow for easy installation and deployment of the system
- Allow for the system to provide feedback to the experimenters about its current operational state (e.g., faulty sensors), and the gathered data

V. Discussion

The aim of our heuristics is to guide future work and provide a foundation for an adequate infrastructure to measure audience experience in realistic conditions. We believe that this set of heuristics is generalizable to other types of biosensors and other applications. For example, measuring students' response in a lecture. In sum, an adequate infrastructure should cover at least three main areas: processes, data, and system.

Processes put the needs of the stakeholders and users before the needs of the audience research. It is pertinent to have a grasp of these needs so that the ultimate benefits of audience research can be realized. The underlying principle of Heuristic #1 is that the experiment does not cause inconvenience or becomes imposing because of a lack of understanding with the stakeholders. From our experiments, we learned that there are many professionals and each one in charge of a number of tasks (e.g., lighting, scripting…). While producers and directors may be interested in quantifying audience experience, they may not have the extra time and cycles for attending to researchers. To avoid unnecessary backlash, extra planning, improved communication, and good organograms would be advantageous apart from the basic guidelines for conducting experiments (e.g., consent forms, questionnaires, questions for the interviews). Other than stakeholders, audiences' needs should not be neglected which is the underlying principle of Heuristic #2. The implementation of a biosensor infrastructure should not affect theater-going experience in any way. We listed some benefits of enjoyable audience experience in our introduction and such benefits can only be obtained if audience needs are placed at the core of the design of the biosensor infrastructure. Regarding ethics, using sensors without personal data ensures privacy of audience members. However, we are aware that deployment of our biosensor infrastructure requires further understanding of ethical issues. All in all, prioritizing stakeholders and users will prompt researchers to closely examine the relationship between technology, design and behavior and think about concerns from multiple angles [26]. In this way, enhance the processes of conducting experiments in realistic conditions.

Heuristics #3, #4 and #5 are related to data. A working knowledge of the unique characteristics of the data measured (e.g., GSR) is important so that these characteristics are preserved from the collection to the processing and analysis. Be prepared to cater for the many sources of bad or missing sensor readings due to poor sensor leads, sensor gels, packet loss during radio transmission or other sources. Without data validity, the experiment would be conducted in vain. Therefore, Heuristic #3 serves as a guideline for achieving valid data sets while Heuristic #4 addresses the importance of keeping data. Other than validity, having a complete set of data is helpful. A complete set of data refers to all forms of data collected during the experiment (i.e., video, interviews…). These forms of data can help to verify findings and form a more complete picture of audience experience. However, researchers should be aware of the privacy rules of different countries. Note that collecting system data is also important: future experiments with different data processing algorithms may well depend on having recorded detailed timing or packet loss information during the real experiment.

Although visualization may not be a requirement for all experiments, Heuristic #5 was developed to guide the real-time of data gathering. As mentioned in the introduction, visualization can enhance the feeling of shared experience and visualization requires real-time. As it would be difficult to see how running wires to individual audience members would fit Heuristics #1 and #2, we need some form of wireless communication. From experience, wireless communication was effective and efficient in meeting the demands of real-time data gathering.

Design principles of a robust system are embodied in Heuristics #6 and #7. To effectively understand the effect of a performance on different audience members, concurrently measuring audience experience becomes a necessity. Having conducted the experiments in different venues with different sizes of audience, we appreciated a system that is adaptable to these differences. Trial runs would also help to ensure the scale of the actual experiment could be achieved. Last but not the least, experimenters need to be aware of the short time window for attaching the sensors as audience members usually arrive at a performance venue close to the start of a performance. As a result, a system that provides feedback for experimenters to act upon and a system that can be easily deployed are crucial to the success of the experiment.

VI. Conclusion

Through the process of developing our unique biosensor infrastructure and deploying it in the wild, we have consolidated our learning and derived a set of heuristics. We consider our heuristics as a starting point. Following them, we have recently conducted a study in China where we measured audience experience of 150 audience members while they watched the Chinese adaptation of the acclaimed theater performance “War Horse” presented by the National Theater of China. Thanks to our heuristics, the experiment has been a success and the results will be reported in an upcoming publication as soon as the data sets are fully analyzed.

References

[1] Latulipe, Celine. "The value of research in creativity and the arts." In *Proceedings of the 9th ACM Conference on Creativity & Cognition*, pp. 1-10. ACM, 2013.

[2] Carnwath, John D., and Alan S. Brown. "Understanding the value and impacts of cultural experiences." Manchester, United Kingdom: Arts Council England ,2014.

[3] McCarthy, Kevin F., Elizabeth H. Ondaatje, Laura Zakaras, and Arthur Brooks. *Gifts of the muse: Reframing the debate about the benefits of the arts*. Rand Corporation, 2001.

[4] Matarasso, François. "Use or ornament." The social impact of participation in the arts 4, no. 2, 1997.

[5] Reeves, Michelle. Measuring the economic and social impact of the arts: a review. London: Arts Council of England, 2002.

[6] Walmsley, Ben. "“A big part of my life”: a qualitative study of the impact of theatre." Arts Marketing: An International Journal 3, no. 1 pp.73-87, 2013.

[7] Wilson, Ceri, Hilary Bungay, Carol Munn-Giddings, and Melanie Boyce. "Healthcare professionals’ perceptions of the value and impact of the arts in healthcare settings: A critical review of the literature." International journal of nursing studies, pp. 90-101, 2015.

[8] Catterall, James S. "The Arts and Achievement in At-Risk Youth: Findings from Four Longitudinal Studies. Research Report# 55." *National Endowment for the Arts* (2012).

[9] http://www.ey.com/Publication/vwLUAssets/Measuring_cultural_and_creative_markets_in_the_EU/$FILE/Creating-Growth.pdf

[10] www.holland-dance.com

[11] https://byborre.com

[12] http://www.ntcc.com.cn

[13] Arttu Perttula, Pauliina Tuomi, Marko Suominen, Antti Koivisto, and Jari Multisilta. 2010. Users as sensors: creating shared experiences in co-creational spaces by collective heart rate. In Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '10). ACM, New York, USA, 41-48.

[14] Livingstone, S. (2010), Giving People a Voice: On the Critical Role of the Interview in the History of Audience Research. Communication, Culture & Critique, 3: 566–571.

[15] Marchesi, Marco, Elisabetta Farella, Bruno Riccò, and Antonella Guidazzoli. "Mobie: a movie brain interactive editor." In SIGGRAPH Asia 2011 Emerging Technologies, p. 16. ACM, 2011.

[16] C. J. Stevens, E. Schubert, R. H. Morris, M. Frear,J. Chen, S. Healey, C. Schoknecht, and S. Hansen. Cognition and the temporal arts: Investigating audience response to dance using PDAs that record continuous data during live performance. International Journal of HumanComputer Studies, 67(9), pp. 800 – 813, 2009.

[17] Latulipe, Celine, Erin A. Carroll, and Danielle Lottridge. "Love, hate, arousal and engagement: exploring audience responses to performing arts." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1845-1854. ACM, 2011.

[18] VaezMousavi, S. Mohammad, Robert J. Barry, Jacqueline A. Rushby, and Adam R. Clarke. "Arousal and activation effects on physiological and behavioral responding during a continuous performance task." Acta neurobiologiae experimentalis 67, no. 4, 461, 2007.

[19] Benedek, M., & Kaernbach. A continuous measure of phasic electrodermal activity. Journal of neuroscience methods, 190(1), pp. 80-91, 2010.

[20] Papez, James W. "A proposed mechanism of emotion. " The Journal of neuropsychiatry and clinical neurosciences 7, no. 1 (1994): 103-112, 1937.

[21] Edelberg, Robert. "Electrodermal Recovery Rate, Goal-Orientation, and Aversion." Psychophysiology 9, no. 5 (1972): 512-520.

[22] Fowles, D.C.. Psychophysiology and psychopathology: A motivational approach. Psychophysiology, 25(4), pp.373-391, 1988.

[23] Fleureau, Julien, Philippe Guillotel, and Izabela Orlac. "Affective benchmarking of movies based on the physiological responses of a real audience." In Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on, pp. 73-78. IEEE, 2013.

[24] Martella, Claudio, Ekin Gedik, Laura Cabrera-Quiros, Gwenn Englebienne, and Hayley Hung. "How Was It?: Exploiting Smartphone Sensing to Measure Implicit Audience Responses to Live Performances." In Proceedings of the 23rd Annual ACM Conference on Multimedia Conference, pp. 201-210. ACM, 2015.

[25] Madgunda, Sneha, Upasna Suman, and Sai Praneeth G. Raunak Kasera. "Steps in Requirement Stage of Waterfall Model." International journal of computer & mathematical sciences, pp. 86-87, 2015.

[26] Rogers, Yvonne. "Interaction design gone wild: striving for wild theory." Interactions 18, no. 4, pp. 58-62, 2011.

[27] Pinelle, David, Nelson Wong, and Tadeusz Stach. "Heuristic evaluation for games: usability principles for video game design." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1453-1462. ACM, 2008.

[28] Wang, Chen, Erik N. Geelhoed, Phil P. Stenton, and Pablo Cesar. "Sensing a live audience." In Proceedings of the 32nd annual ACM conference on Human factors in computing systems, pp. 1909-1912. ACM, 2014.

[29] Nielsen, Jakob. "Enhancing the explanatory power of usability heuristics." In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 152-158. ACM, 1994.

[30] Molich, Rolf, and Jakob Nielsen. "Improving a human-computer dialogue." Communications of the ACM 33, no. 3, pp. 338-348, 1990.

[31] Nielsen, Jakob, and Rolf Molich. "Heuristic evaluation of user interfaces." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 249-256. ACM, 1990.

[32] Dykstra, Dean Julian. A Comparison of Heuristic Evaluation and Usability Testing: The Efficacy of a Domain-Specific Heuristic Checklist. A & M University, Texas, 1993.