The Play Is a Hit – But How Can You Tell?

Measuring Audience Bio-responses Towards A Performance

Chen Wang[1,2], Pablo Cesar[1,3]

[1] Centrum Wiskunde & Informatica [2] Future Media Convergence Institute

[3] Delft University of Technology

ABSTRACT

Research has shown that physiological sensors provide a valuable mechanism for quantifying the experience of audiences attending cultural events. In the strength of the applause or questionnaires, bio-sensors provide fine-grained timed data that can be used to infer the quality of the experience of the audience members. Unfortunately, available commercial sensors are designed for lab or home usage and studies, focusing on the individual instead of the reactions of a crowd of people. In this study, we present our own designed physiological measurement system that overcomes the challenges of using and deploying sensors in a theatrical environment (anonymity and privacy, real-time gathering of data, support for large crowds of 30-100 people), and the wearable system has two unique features that distinguished between adult audience members and children ones. We report our experimental results which particularly answered the research questions proposed by the producer, director, and the artists.

Author Keywords

Physiological sensor network; GSR sensors; physiological computing; audience response; theatre performance

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Arts add value to the lives of individuals and to society as a whole. Attendance and participation in arts can help individuals and society develop internal cognitive and emotional process. As a result of a cultural experience, such processes can have impact on external outcomes, e.g., increased educational attainment, reduced crime rates, health and overall well-being [29]. Besides, arts are a communicative experience, a bridge from artist to audience and a bridge linking individual beholders to one another [28]. Arts provide a medium where artists express their unique experience, and make their inner reality public and communicable to others. Therefore, arts are claimed to have cognitive, attitudinal, and behavioral benefits for users to develop their personal skills, and thus increase social bonding through arts participation [42].

Figure 1: Our physiological sensors were used during the War Horse performance Chinese version.

Arts engage audience at the emotional and intellectual as well as the aesthetic level. Understanding how the communication from artists to audience works can help artists appreciate the value of arts. Artistic creation is one of the most complex and mysterious activities of human consciousness, and the effects of arts are intangible and difficult to define. In particular, audience experience is an inner one (intensely personal and private), which makes it difficult to be appropriately measured.

C&C '17, June 27-30, 2017, Singapore, Singapore

 ISBN 978-1-4503-4403-6/17/06….$15.00

DOI:http://dx.doi.org/10.1145/3059454.3059465

Figure 2 (drawn from [14]): three discernable types of impacts of arts defined by their temporal proximity to a cultural event.

The literature defines the impacts of arts on an audience as a progression of three stages based on their temporal proximity to the cultural event: concurrent impacts, experienced impacts, and extended impacts [14]. As shown in Figure 2, concurrent impacts occur during the event, and can be measured through biometric research. Experienced impacts can happen before, during and after the event, and are typically measured through post-event surveys and interviews. With respect to extended impacts, they are measured through retrospective interviewing and longitudinal tracking studies.

Measuring audience concurrent reaction plays an important role to evaluate the effectiveness of an artistic production. Because the concurrent reaction reflects how an audience 'look, listen, and enjoy' a performance, and it can be used to estimate the reception of intended effects. For instance, audience may feel challenged, moved, engaged by the show, and they may not notice the time. Such emotional response is then brought together to give an overall measurement of audience reaction.

Various stakeholders, including artists, producers, and directors, have sought out different methods to measure the impact of a performance. In general, it takes the form of a questionnaire [10, 21] handed out to audiences after a production. Some other methods rely on box office figures, social monitoring [9] or critics' reviews [12] to assess the quality of a show. However, such solutions are mainly the experienced impacts of plays, other than concurrent impacts. In addition, individual's concurrent impact may not be always tracked by subjective measurements, as individual may respond to cultural events without being consciously aware of it, and any conscious reflection (e.g., using a labeling system [22, 34] during a performance) on the individual's state may interfere with the experience (i.e., interrupt their sense of flow or absorption).

Physiological and pre-cognitive psychological responses can be used to objectively measure the concurrent impacts of audiences [19]. Compared to the other methods, physiological responses can be measured non-intrusively at the very moments at which they occur, because intellectual and emotional reactions from the audience can take place at any time during the events. A number of studies have sought to capture biometrical data as a means of gauging audience responses to arts, by tracking eye movements [11], heart rate [39], skin conductance [43], emotional response [22] and aesthetic judgments [39] during performances. However, this method is not widely applied due to the complexity of setting up the biosensor infrastructure, the bulky form factors, and high prices of the commercial products.

In this paper, we propose an open-source solution to design, develop, and deploy a GSR sensor system, particularly designed for theatergoers. This new measurement system gives theatre experts an intriguing, alternative method by which we evaluate the quality of their output. By using our method, theater experts can achieve valid results, which are certainly an improvement on some of the more basic instrumental methods that have been used in the past. Thus, the main contribution of this paper are twofold:

C1: *The new GSR sensor system can be effectively deployed at theatrical environment;*

C2: *The obtained sensor data can be used to measure the concurrent audience response and have better results compared to other traditional methods (e.g., questionnaires)*

Motivated by the intrinsic and instrumental benefits from the arts, using video recordings as the stimulus and executing experiments in lab conditions, we run the studies in a real performance, and captured the realistic audience GSR readings.

Research Questions

We conducted a field study with the National Theater Company of China in particular for the Chinese version of War Horse. The Chinese version of War Horse is the first co-production between the National Theatre of Great Britain and the National Theatre of China (Figure 1). The original British version was first premiered in London in 2007, and the show has been seen by over 6 million people worldwide and been performed over four thousand times. In the year of 2015, War Horse took its biggest jump into China.

Challenges by experts

We interviewed (semi-structured) the producer, the director, and the stage manager from the National Theater Company. Li Dong, producer of the Chinese version of War Horse, received the interview in Shanghai, China (Figure 3). Li Dong stated that there is a need for quantified audience feedback to resolve ambiguity. For instance, he was unsure of the impact of the War Horse performance in China. In England, the production was popular for all ages. Even though the script was adapted from a children's novel, it dealt with history and complex emotions. Therefore, he was not sure how audiences in China would receive the show. In particular, he was extremely interested in knowing how child audience would react to the show. At the initial stage, they would like to prompt the performance to targeted at child market, but they lack quantified data to support the decision.

Figure 3: The interview with the producer Li Dong from China National Theater.

Furthermore, he felt that there is a need for concrete understanding of audiences' reception rather than an ambiguous perception of how popular the show is. For instance, mentioned their interest in how many performance scenes related to Joey (a horse, the main character of the play) would arouse the interest of the Chinese audience. With this understanding, they could make better plans for marketing strategies and promotional-related materials.

Besides, both producer and director felt that audience feedback could help them make informed choices. In particular, Li Dong mentioned that almost 70% of the theatregoers in China are females and they bring along their male friends and children to watch a show, as such, it will be useful to know what elements of the show are evoking for females and the difference in emotional intensities between female and male audience members.

In addition, by understanding the elements of a play that are stimulating the majority of the audience (e.g., the thrilling climax scene about the competition between Joey and Topthorn,), the director can make more astute choices in the later production and create more successful performances.

Last, all of them stated that a quantified audience experience should not limit artistic expression but help to form the basis of certain choices. For instance, selecting elements or narrating the story from which character's point of view, it may help generate audience interest. Besides, they foresee the possibilities to use audience response in real-time to explore new performing areas (e.g., an interactive play that is particularly designed to have the different endings based on the feedback of audience).

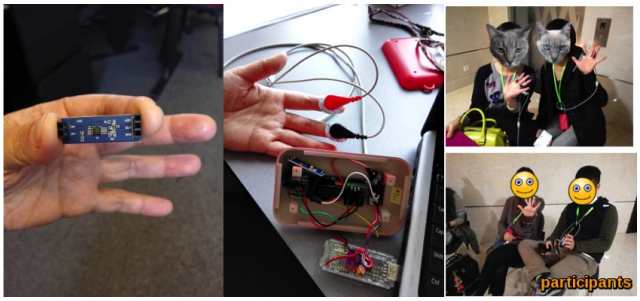

Figure 4: the GSR sensor (left). The Jeenode board and GSR sensor are connected and mounted in a customized 3d printed box(middle). The Chinese audience wore the sensors during the War Horse performance (right).

Figure 5: the child version GSR sensors. One of child participants wore the sensor during the performance (left). The five housing design of the sensors (middle). The sensors used in the experiments (right).

Research questions are from the interviews. The interviews revealed that audience, e.g., the event-related reactions, were the main interest of the National Theater. We believe that a bio-sensor system has significant advantages to measure the continuous audience experience, e.g., emotional intensities from the audience, because the sensors provide seamless and continuous data that map audience's reactions to the events. By summarizing the interviews, we identified the four research questions:

R1: How many performance scenes related to Joey aroused the imterest of audience?

R2: Which performance elements aroused the interest of female and male audience members respectively?

R3: How female audience members reacted differently to males during the thrilling climax scene about the competition between Joey and Topthorn.

R4: What were the different emotional intensities between parents and children?

In this paper, we first report on the related work, and then we present the methodology. After that, we show the hardware

development. Finally, we report on results from our work. Then, we conclude the paper with discussion and conclusion.

RELATED WORK

GSR Sensors

Electrodermal activity (GSR) is also known as skin conductance, GSR, electrodermal response, skin conductance response (SCR), and skin conductance level (SCL). GSR refers to the changes in skin conductance at the surface, reflecting activity within the sympathetic axis of the autonomic nervous system (ANS). 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 [32]. Furthermore, it is widely recognized that increased GSR responses can be provoked by attention-related stimuli or tasks [16, 38].

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 electrodermal responses (EDR) related to specific stimuli, representing the quick and phasic changes imposed on shifts in tonic level in conductivity [18, 37]. In general, there is a delay of 1-3 seconds between stimulus and SCR onset. Hands and feet can be used to measure GSR, where the density of the sweat glands is the highest.

There are some commercial GSR sensors available on the market, but they are not suitable for theaters. First, they are generally designed for lab studies, using technologies like Bluetooth [20, 27, 40] which are intended for individual rather than for measuring audience at a large scale. Second, commercial sensors are generally uniformly designed for a controlled laboratory experiment [4, 5, 24, 25], without considering user experience in a specific application domain. Third, some commercial smart watches have sensors integrated. In this case, the anonymity of the sensor signals is not preserved (since the mobile phone may have to sign on for using the web server). In addition, we foresee that theaters companies will provide the audience with their own lightweight sensors and deploy a specific network for the purpose of better understanding and quantifying the audience experience.

Physiological Sensor Network

The current available personal sensor networks (PSNs) are mainly developed and designed for lab [1, 2, 3] and home studies (e.g., monitoring elderly people) [20, 27]. In the majority of the cases, they communicate through Bluetooth, which is an excellent solution for measuring individuals [26, 33]. Besides, some studies [17, 35, 43] developed a PSN by using open source alternatives (e.g., Arduino, connected with Zigbee/Xbee or WiFi), which have enabled us to study several use cases [20, 23]. Recently, with the advent of miniature electronics, radio technology provides alternative frequency bands to construct a wireless network [40, 36, 41]. Interestingly, most of these products were used on healthcare applications (i.e., monitoring the heart rate of people [30, 45]).

Audience Research in Performing Arts

Researchers have developed numerous ways to measure, interrogate and assess audience's response to arts and culture, including biometric measurements, post-event surveying, qualitative post-event research, and longitudinal, or retrospective studies. Biometric measurements objectively manifest audience response to the aesthetic experience [43, 11, 39], even though interpreting such responses is difficult. Nonetheless, knowledge of audience bio-response will be of considerable value in advancing our conceptual understanding of impact on the individual[14]. Post-event surveys have been widely applied to evaluate the short-term effects of specific cultural events [10, 21], but surveying requires researchers to design the constructs in advance. Furthermore, surveys can only capture the conscious experience of respondents. There is limited comparability across events, and they fail to capture effects that unfold over time. Qualitative methods allow informants to reflect on the areas that are most significant to them. Unlike surveys, qualitative studies can help researchers contextualize numerical data. Longitudinal impacts on audience studies add durable value on cultural experiences [31, 8]. While the retrospective identification of cultural events may not be helpful to assess the impacts of specific artistic works, it can inform us how cultural participation plays an important role within the large scope of people's lives [14].

METHODOLOGY

We used the interviews to conduct this research. The pre - interviews were conducted to collect the research questions, and the post-interviews were used to evaluate the performance of our method. All these work were done with the experts.

The semi-conduct interview protocol was performed. The pre- interviews were taken in two rounds, and each round lasted 30 minutes, and the video and the audio were recorded. Five researchers took observational notes during interviews. After the interviews, researchers discussed the observational notes and identified the potentially interesting topics. More specific research questions were concluded after the discussion. The post-interviews were conducted two times.We first presented the results to the expert, and then we showed them to the theater company.

INFRASTRUCTURE

In this section, we present the GSR sensor infrastructure (Figure 6). The process to produce the newest factory-made production, which was in the War Horse experiment. The new sensors have two versions (Figure 4 & Figure 5), which aimed at two different audience categories: adult and children. We first introduced our own GSR sensors, and then we presented the development of the software and the mobile app.

Sensors

The knowledge for building our GSR sensors was based on previous work [15]. We built our GSR sensors by using Jeenode [1]. Based on our previous experience, we know that the factory PCBA service is much better than the extensive self -soldering work. Besides, for certain quantity, e.g., 100-200 sensors, the factory solution is much cheaper and quicker. Once the prototype was available, we sent the circuit to our supplier for production. Another advantage is that such manner can control the size and the interface of the sensors, so that we can easily intergrade the sensors to the development board and make small and comfortable wearables. At last, the whole system was put into a 3D printed casing. In order to guarantee the most unobtrusive and comfortable experience for theatergoers, we sent the sample product to our project partner in China for evaluation with a Chinese audience. After several versions and modification, we came to the final design as shown in Figure 4 and Figure 5, each sensor only weighing 76g.

Network topology and Media Access Layer (MAC)

We used 433MHz (free Industrial, Scientific, and Medical Band) to establish the wireless sensor network. Once the frequency band was set, we had to decide a media access mechanism and a network topology. The network is formed by a group of sensors and a sink node connected to a central server. Since the sink node communicates with a group of sensors, in order to minimize the packets collision, we implemented a polling method in the MAC layer in the C programming language. By this method, all the sensor nodes wait for a pull message sent by the sink node. Each time, the sink node broadcasts a poll message with a specific sensor ID, only the sensor with that ID can send packets, while the other sensors go to the standby state. In this way, we try to minimize the collisions in the wireless channel. However, the packet loss caused by the crowd presence was unlikely to be avoided in theatrical environment. Finally, the sensor system reached the 4Hz sampling rate. The architecture of the GSR wireless sensor network is shown in Figure 6.

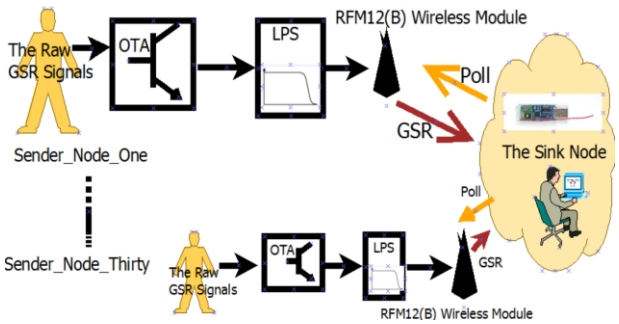

Figure 6: The schematic of the wireless GSR sensor network

Software

For Warhorse experiment, we adapted the method developed by Fleureau et. al [25] by observing SCR signals invoked by stimulus. The method first extracts the SCR signals based on the timeline, and then defines which SCR responses were statistically significant when compared to the background noise. In such a manner, we can provide links between the significant audience bio- responses and the related performing events (e.g., the appearance of Joey), so our customers can have a better understanding in terms of which events substantially affected the audience and which ones did not.

Extracting the SCR signals from the sensors requires signal processing, filtering, smoothing, and derivative procedures, as shown in Figure 7. First the raw signals are scaled and filtered by a 2Hz low-pass filter ($G(t)$). After that, the numerical derivation performed, but only the positive phasic changes are kept while the negative ones are ignored (from $G'(t)$ to $G'_{+}(t)$). The reason is that the negative phasic changes are not of our interest, as they only reflect the physical recovering from a stimulus. In the next step, we apply a moving window (W_i, W_j :i, j = 1……k, where k is the number of windows) with the window size of 120 samples (30 seconds), and the overlapping size of 60 (15 seconds) to smooth and compute the mean values of the derivative signal $G(i)$. Finally (since each individual may have a different amplitude for the derivative GSR signals), we normalized the computed signals by using the sum of subsampled skin response values as the denominator to calculate the individual mean arousal value ($G_n(i)$: n=1……N, where N is the number of users) (2).

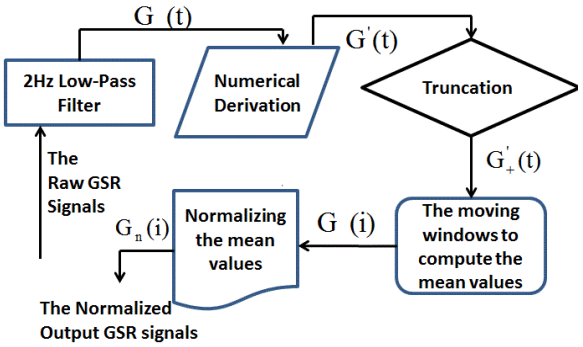

Figure 7: The description of the different steps of the algorithm for processing the raw GSR signals.

[1] http://jeelabs.net/projects/hardware/wiki/JeeNode

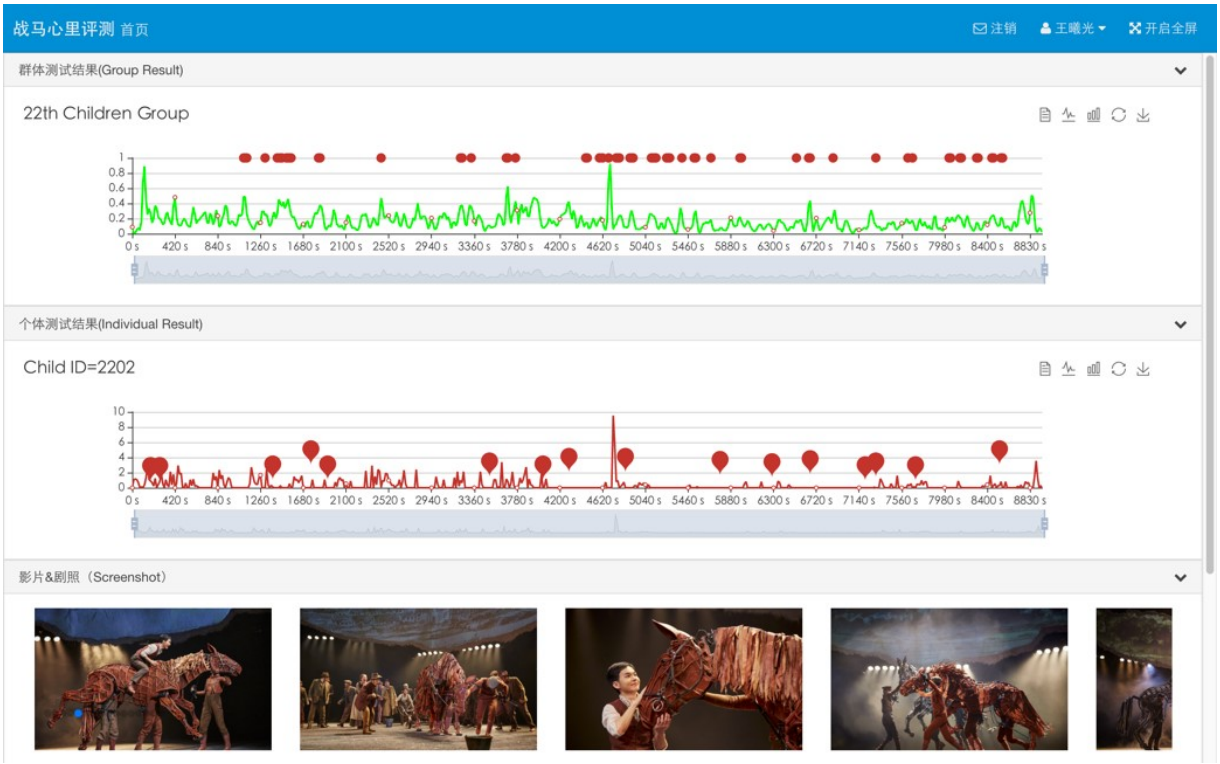

Figure 8: the visualization results based on the performance of 22th, November. The first row displays the significant arousal moments from the group children. The second row shows the significant performing events evoked from the child id (2202, his parental id: 2206 in Figure 10). By clicking the red dots in the second row, the related performing scenes (the screenshot) will be displayed in the below.

$$G_n(i) = \frac{G(i)}{\sum_{j=1}^{k} G(j)} = \frac{\int_{W_i} G'_+(t)dt}{\sum_{j=1}^{k} \int_{W_j} G'_+(t)dt} \quad (1)$$

The bilateral Mann-Whitney-Wilcoxon (MWW) test is performed to define the significant SCR responses. The mean p values are computed by averaging the p-values of the bilateral MWW test performed between the latent unknown distribution of $G_n(i)$ and the background noise. We define 10% of the computation results with the lowest mean as background noise [25]. Only an associated p value lower t han 5% is considered as significantly different from the background noise.

Trigger App

We created a trigger app in a mobile phone in order to annotate significant events in the timeline of the play, and to align them to the sensor recordings (Figure 9). Using the trigger interface, we can manually type the description about the performing events, and the system automatically records the time stamps of each description into the excel file while showings the local time on the app, as shown in Figure 9. The reason for developing such app was that the theater company did not allow us to record the performance due to copyright issues. In order to time align the sensor readings with the performing events, we needed this handy tool to manually type the events into the app, so that later on we could synchronize the sensor data with the events.

Development on the visualization platform

The psychology data analysis and viewer platform specialized for War Horse is a customizable end to end solution for the audience emotional intensity analysis and result visualization (as shown in Figure 8,9,11,12). The platform is a Web based application which was developed with Nodejs and MongoDB. It could dynamically load the raw data and generate the interactive charts in real-time. It can be divided into several major models, including raw data processing, data standardization, data analysis, result showing and the system configuration. Once the raw data was imported or updated into the database, they will be processed automatically, and the data will be stored and transferred as Json file. When an individual login the system, the username will be recognized and automatically associated with its ID number and his or her profiles which have been input to the database previously. The system then loads the processed data and creates the charts with the mark points. The end users could use the interactive components on the Web UI to compare with various group results.

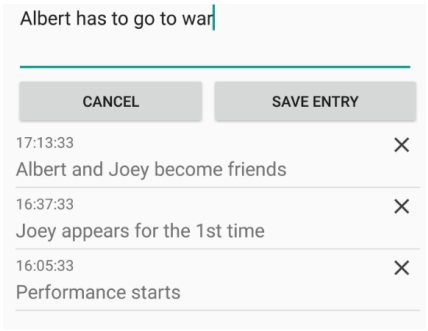

Figure 9: Trigger app to map events and sensor readings

EXPERIMENTS

There were in total 5 rounds experiments conducted. We recruited 90 adult participants (Male: Avg.: 33, STD: 12.4; Female: Avg.: 28.9, STD: 8.04) who joined 3 rounds of experiments (the same performance but at the different date). In addition, we also recruited 60 family participants (parents with their children (6-9 years old, Male: 57.1%; Female: 42.9%) to join the 2 rounds experiments respectively. However, since in the second round the child age scope was out of our recruitment requirements, so that we only report the results based on the first round experiment. Before the performance started, we explained the purpose of the experiment, and provided pre-questionnaires and consent forms. After this, the audience wore the sensors at their non-dominating hand, and they went inside the theater 15 minutes before the performance started. There was a 15 minutes break between the first part and the second part of the performance. When the performance finished, we helped the audience members take off the sensors. After that, there were post-performance questionnaires and interviews.

Before the real experiments, the project manager conducted all the communication work with the theater company and made sure that there was always an effective link between them and us. We critically evaluated the performance venue in order to understand the experimental limitations and how we could establish a biosensor infrastructure there.

All the seats for the participants were pre-reserved, and all the experimental procedures were prepared in advance and each task was assigned to the responsible person. The pre and post questionnaires were provided before and after the performance, and we also had interviews with the participants.

RESULTS

The experimental results show that our sensor system is robust for the theatrical environment. In addition, the developed algorithm successfully helped us answer all the research questions. The wearing experience from the audience feedback was positive. The deployment of the whole system took around 15-20 minutes from welcoming the participants, putting on the sensors, testing, filling the questionnaires, to answering the participants' questions.

Figure 10: the visualization results based on the performance of 22th, November. The first row displays the significant arousal moments from the adult group. The second row shows the significant performing events evoked from the parental id (2206). By clicking the red dots in the second row, the related performing scenes (the screenshot) will be displayed in the below.

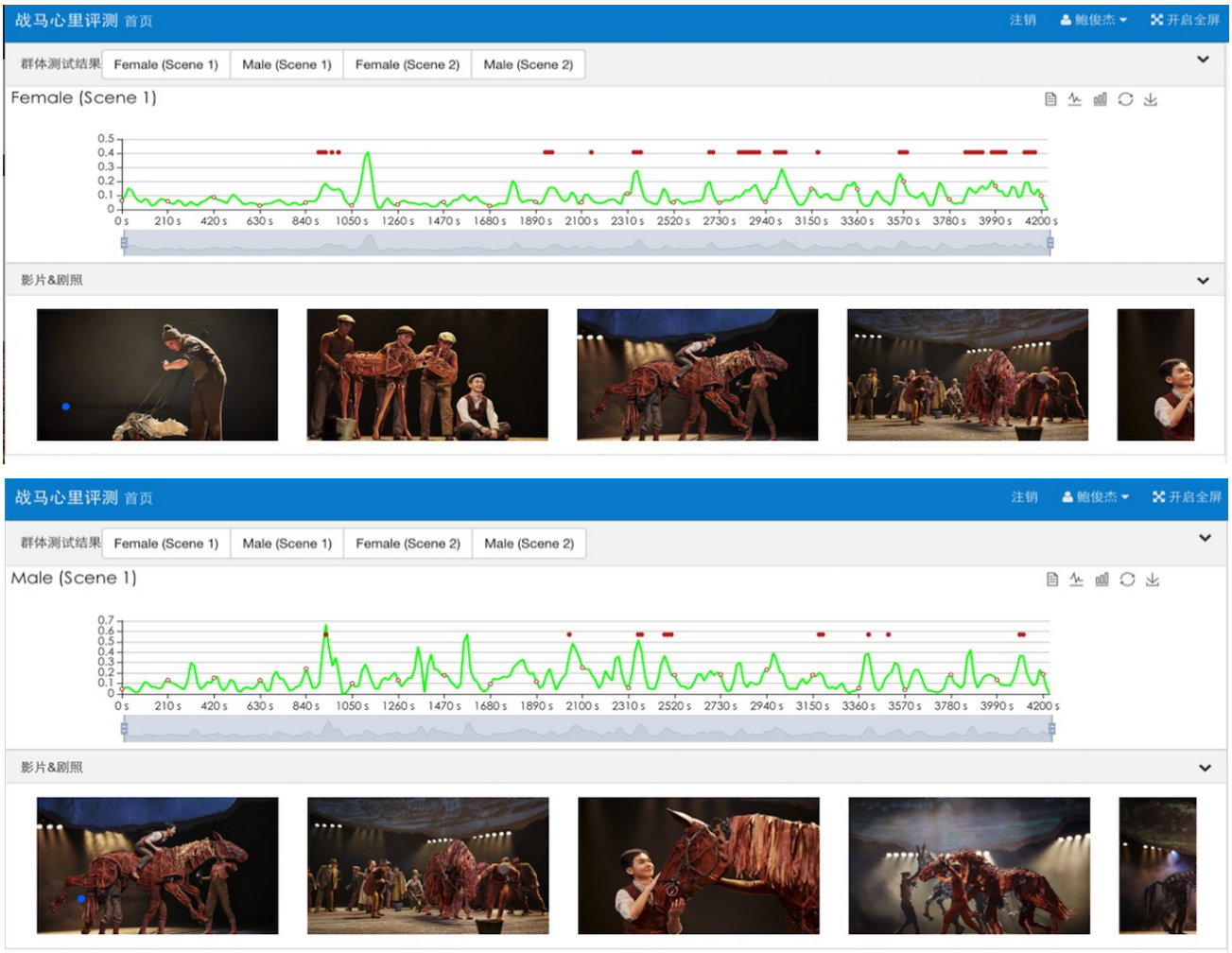

Figure 11: The significant arousal events compared between the female audience and the male audience in the scene 1.

Audience responses

Family audience

The results on the family audience were surprisingly interesting (R4). Based on the marketing investigation results from China theater company, the majority of the parents thought that the performance would be difficult to understand for children. However, based on our results, we found that the emotional intensities elicited from the children audiences were significantly stronger than their parents （Figure 8 & Figure 10）. Besides, from the subjective reports and the interviews with the children, they enjoyed the performance even though they might not fully understand. In comparison, the parental experience revealed that they had less emotional intensities compared to the children. According to their report, they were extremely tired during the day time, because they had to take care of their kids before they went to the theater. During the show, they were distracted by their children instead of enjoying a performance; some of them even slept during the play.

The sensor results also showed that during the whole performance parental audiences had less arousal moments compared to their children during the whole performance. For instance, there are in total 226 scenes where the Joey appeared: the children were significantly evoked by 73 ones, while their parents were only stimulated by 35 ones. Furthermore, compared to their parents, the children also had more substantial responses towards the emotional scenes when compared to their parents. For instance, there are in total 35 emotional scenes, where the children were statistical elicited by 13 ones, which was as twice as their parents.

The results also guide the theater company to adjust their marketing strategy. Before the experiment, the theater company arranged a visiting event to the back stage exhibition, where Joey and Thompton, and the family audience can have a direct contact with the horse puppet. One of the group participated in the visiting events before they watched the show, and their emotional intensities from the sensor readings were significantly reduced compared to the other. We assume that the familiarity has decreased the

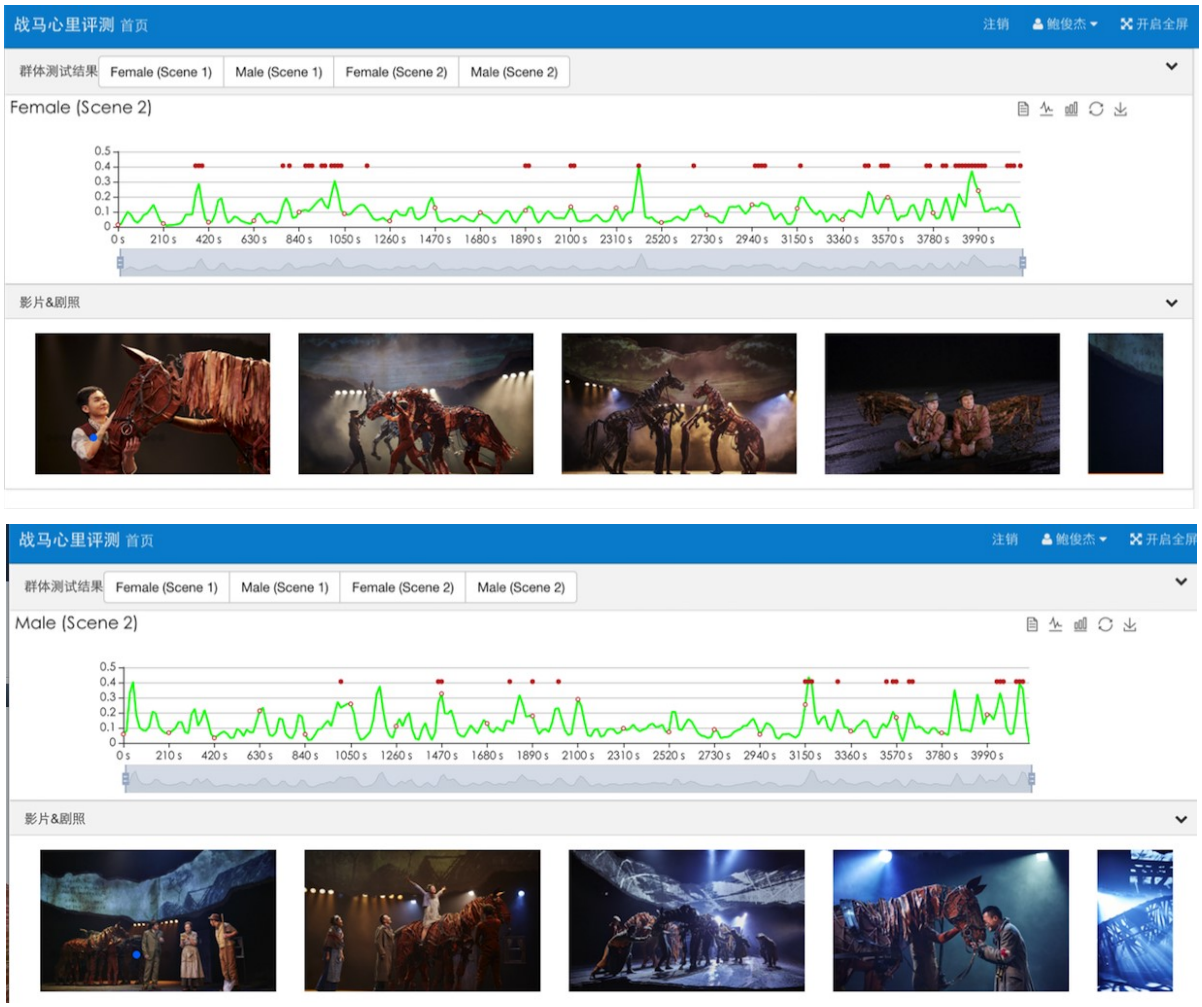

Figure 12: The significant arousal events compared between the female audience and the male audience in the scene 2.

emotional intensities during the watching experience after visiting the back stage exhibition. Based on our reports, the theatre company has adjusted their marketing strategy: inviting visit at the back stage only after the show.

Adult audience

As shown in Figure 11, we found that the interest of both male and female participants was significantly aroused during the thrilling scenes where the competition between the two main protagonists Joey and Topthorn happened, but the duration of this significance lasted longer for females than for males, as indicated by the length of red bars (R3). There were 226 scenes related to Joey. The interest of audience was significantly aroused in about 32 of those (R1). Interestingly, by observing the p values, we found that in some of the scenes both male and female audience was stimulated, but in some of the scenes females were more responsive than males. By comparing the results from the first and the second part (Figure 11, 12), we found that significant responses from females were more intense compared to the ones from males (R2). Besides, the duration of the significant responses of the female audiences was generally longer than those of the males (R2).

Feedback from experts

We invited theater experts to interpret the results. In general, they write journals on a performance based on their own feedback, or they interview several theatergoers in order to obtain the audience opinions. Even though sometime experts combine the internet information and generate a report, but such manner has a big delay. The traditional way of collecting data is based on small samples. If big samples are required, they have to sort out other methods, e.g., spreading out questionnaires. Still, for news reporting, the timing is extremely important. The experts think that our method can provide rich and timing data for them to write reports.

The results from the gender difference helped the experts prove their assumptions on audience watching experiences. In general, the female audience is easily touched by the emotional stories, while male is more linked to action scenes. The experts pointed out that the results revealed that the emotional scenes triggered the female audience the most,

while the male audience's interest was more aroused in the action scenes in the first part of the show (R2). While in the second part of the show, both genders were more connected to the ending in which the soldier Albert met Joey again, but the important scene describing the girl Emilie and Joey significantly elicited the female audiences (R2).

Feedback from China National Theater Company

The producer Li Dong received our interview again. He was extremely excited about the results. He told us that this method well solved the contradictions between him and the directors, as sometimes they may have different thoughts about how a performance should be constructed. The results also stimulated them to be more creative in the future artistic production, i.e., creating an interactive performance, where sensors can be used for visualization or interactions. The most important inspiring thing is that this method could help him define a marketing strategy. For instance, if he knew the children were intensively attached by the show, he would have invested the budget to promote on the children market.

DISCUSSION

This paper presents our new GSR sensor system, particularly suits for a theatrical environment. By using our method, theatrical experts can assess whether a show was more of a success, for instance it affected a handful of people profoundly, but left the majority nonplussed. Compared to the traditional methods, physiological measurements are more suitable to measure the concurrent audience response, and the sensor data can be effectively connected to the specific events during a play.

The new sensor system brings a new perspective on audience research in a theatrical environment. Based on the results, theater companies can develop a new business model and make better business decisions, i.e., promoting the show to the children market. In addition, our method also opens a new path for the future creative artistic production. The experiments revealed that the bio-sensor method has advantages to record the continuous data, and they can be easily mapped to the certain performing events. In such a manner, the theater experts can have a direct visualization platform to reflect their performance, and interpret them from different perspectives. In addition, due to the liveness of the sensors, the interactive artistic production can be vividly designed, which may bring a total new performing style in the future.

Collecting data related to emotional arousal is not new, as skin conductance has been studied for more than 100 years. For this study, the creative issues are how this technology can measure, communicate, and transform emotions from audience, and how we thinking about it. During working closely with theatrical expert, powerful new insights and changes can be achieved with the help of GSR sensors. In other words, objective data about audience emotions bring us more power than subjective reports for a performance. In our case, sensor data revealed more insights with audience affective experience, which enable theater experts to pinpoint what caused such reactions and work on in the future how to design a show aiming at a better affective experience.

Nevertheless, a number of issues still remain. First, the algorithm does not say anything about valence. If valence is required to be defined, additional information is necessary. Secondly, algorithms need to be more robust to some artifacts (e.g., body movements). If researchers want to use our sensor system on a real-time application, more sophisticated algorithms dependent on the experimental objectives should be developed. Third, the network should be more modular. Researchers should be aware that a theatrical environment is rather different to a laboratory condition. In theaters, especially when the crowd is present, the network performance can be rather critical. However, we found that packet loss is more tied to individuals rather than events that influence all sensors at the same time. Therefore, adding more sensors into the network may be a good strategy to cope with packet loss. For instance, we need to work out a solution that can automatically choose free frequency spacing and assign the sensor ID, so that we can avoid coincidently picking up a busy frequency channel

The operational procedure can be further improved. We have learned that a real-time monitoring system should be incorporated. When the experiment starts, we need to have a view of the states of the sensors. In case one of the sensors does not work, we can always replace it. Last, the hardware needs to be improved in the future, as the mechanical wires can cause the connection problems.

CONCLUSION

In this paper, we have reported how the creative GSR sensors measured audience concurrent response during the Chinese version of War Horse. In particular, we have detailed our design and development decisions for establishing an effective GSR sensor network in theaters. The resulting infrastructure, hardware and software, have shown that the system is robust, that it can be easily deployed in theaters, can simultaneously and accurately capture the GSR signals from the audience members.

The benefit of GSR technology enable theater experts to have a new understanding on audience emotional experience (as significantly different SCR responses) in a show. The sensor data provide insights with which gender and age impacts can be effectively visualized. The results motivated theatrical experts to be more innovative at both current and future artistic production.

In the future we will continue to improve the hardware, and develop the related software modules for the new GSR sensor system. Besides, since the Jeenode platform still has the capability of carrying more sensors, we will investigate how other types of sensors (e.g., ECG sensor and acceleration sensors), can be integrated into our sensor network in order to provide a more complete representation of audience emotional experience.

ACKNOWLEDGEMENTS

This work was supported in part by funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. ICT-2011-287760, and in part by funding from Xinhuanet.

REFERENCES

1. The Thought Technology Ltd.: http://thoughttechnology.com/
2. Biopac: http://www.biopac.com/electrodermal-finger-transducer-ss
3. Q sensor: http://qsensor-support.affectiva.com/
4. Shimmer sensor: http://www.shimmersensing.com/
5. Empatica: https://www.empatica.com/e4-wristband
6. https://en.wikipedia.org/wiki/Heuristic
7. Markusen, Ann, and Anne Gadwa. "Creative Placemaking: a white paper for the Mayors' Institute on City Design." National Endowment for the Arts. Online. Available at http://www. nea. gov/pub/CreativePlacemaking-Paper. pdf (accessed January 2010) (2010).
8. Walmsley, Ben. 2013. '"A big part of my life": a qualitative study of the impact of theatre.' Arts Marketing: An International Journal 3:1. 73 – 87.
9. Louise Barkhuus and Chiara Rossitto. 2016. Acting with Technology: Rehearsing for Mixed-Media Live Performances. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 864-875. DOI: https://doi.org/10.1145/2858036.2858344.
10. Bunting, Catherine, and John Knell. "Measuring quality in the cultural sector. The Manchester metrics pilot: Findings and lessons learned." Arts Council England (2014).
11. Stevens, Catherine, Renee Glass, Emery Schubert, Johnson Chen, and Heather Winskel. 2007. 'Methods for measuring audience reactions'. Proceedings of International Conference on Music Communications Science, Sydney.
12. David Goergen, Angelo Migliosi, Vijay K. Gurbani, Radu State, and Thomas Engel. 2014. Spatial and temporal analysis of Twitter: a tale of two countries. In *Proceedings of the Conference on Principles, Systems and Applications of IP Telecommunications* (IPTComm '14). ACM, New York, NY, USA, Article 5 , 6 pages. DOI=http://dx.doi.org/10.1145/2670386.2670392
13. Tom A. Garner and Mark N. Grimshaw. "The physiology of fear and sound: working with biometrics toward automated amotion recognition in adaptive gaming system." 2013. In proceedings of IADIS International Journal. Vol. 11. No. 2. Pp. 77-91. ISSN: 1645-7641.
14. Carnwath, John D., and Alan S. Brown. "Understanding the value and impacts of cultural experiences." Manchester, United Kingdom: Arts Council England (2014).
15. Guerreiro, José. A biosignal embedded system for physiological computing. Diss. Instituto Superior de Engenharia de Lisboa, 2013.
16. R. Edelberg 1972. "Electrodermal recovery rate, goal-orientation, and aversion," Psychophysiology, vol. 9, pp. 512–520, Sep.
17. Emre Ertin, Nathan Stohs, Santosh Kumar, Andrew Raij, Mustafa al'Absi, and Siddharth Shah. 2011. AutoSense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field. In Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems (SenSys '11). ACM New York, NY, USA, 274-287. DOI=10.1145/2070942.2070970
18. Don C. Fowles. 2007. Psychophysiology and Psychopathology: A Motivational Approach. Article first published online: 30 JAN 2007. Volume 25, Issue 4, pages 373–391. DOI: 10.1111/j.1469-8986.1988.tb01873.x
19. Lindgaard, Gitte, and TW Allan Whitfield. "Integrating aesthetics within an evolutionary and psychological framework." Theoretical Issues in Ergonomics Science 5.1 (2004): 73-90.
20. Charalampos Liolios, Charalampos Doukas, George Fourlas, and Ilias Maglogiannis. 2010. An overview of body sensor networks in enabling pervasive healthcare and assistive environments. In Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '10), Fillia Makedon, Ilias Maglogiannis, and Sarantos Kapidakis (Eds.). ACM, New York, NY, USA, , Article 43 , 10 pages. DOI=10.1145/1839294.1839346
21. Celine Latulipe, Erin A. Carroll, and Danielle Lottridge. 2011. Evaluating longitudinal projects combining technology with temporal arts. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). ACM, New York, NY, USA, 1835-1844. DOI=http://dx.doi.org/10.1145/1978942.1979209
22. Celine Latulipe, Erin A. Carroll, and Danielle Lottridge. 2011. Love, hate, arousal and engagement: exploring audience responses to performing arts. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). ACM (2011), New York, NY, USA, 1845-1854. DOI=10.1145/1978942.1979210
23. Gang Li, Haifeng Chen, and Jungtae Lee. 2010. A wireless EEG monitor system based on BSN node. In Proceedings of the 1st ACM International Health Informatics Symposium (IHI '10), Tiffany Veinot (Ed.).

ACM, New York, NY, USA, 758-764. DOI=10.1145/1882992.1883107

24. C. Martella*, E. Gedik*, L. Cabrera-Quiros*, G. Englebienne, and H. Hung, "How Was It? Exploiting Smartphone Sensing to Measure Implicit Audience Responses to Live Performances", ACM MM 2015 (oral) *These authors contributed equally to this article
25. Fleureau, Julien, Philippe Guillotel, and Izabela Orlac. 2013. "Affective benchmarking of movies based on the physiological responses of a real audience." Affective Computing and Intelligent Interaction (ACII), Humaine Association Conference on. IEEE.
26. Strauss, Marc, et al. 2005. "The handwave bluetooth skin conductance sensor." Affective computing and intelligent interaction. Springer Berlin Heidelberg, 699-706.
27. Sidharth Nabar, Ayan Banerjee, Sandeep K. S. Gupta, and Radha Poovendran. 2010. Evaluation of body sensor network platforms: a design space and benchmarking analysis. In Wireless Health 2010 (WH '10). ACM, New York, NY, USA, 118-127. DOI=10.1145/1921081.1921096
28. McCarthy, Kevin F., et al. Gifts of the muse: Reframing the debate about the benefits of the arts. Rand Corporation, 2001.
29. Matarasso, F. "Creative Progression: Reflections on quality in participatory arts." UNESCO Observatory Multi-Disciplinary Journal in the Arts 3.3 (2013): 1-15.
30. Fletcher, Richard Ribon, et al. 2010. "iCalm: Wearable sensor and network architecture for wirelessly communicating and logging autonomic activity." Information Technology in Biomedicine, IEEE Transactions on 14.2 : 215-223.
31. Radbourne, Jennifer, Katya Johanson, and Hilary Glow. 2010a. 'Empowering Audiences to Measure Quality'. Participations 7:2. 360-379.
32. J. W. Papez. 1937. "A proposed mechanism of emotion," J. Neuropsychiatry Clin. Neurosci., vol. 7, pp. 103–112.
33. Picard, Rosalind W. 2009. "Future affective technology for autism and emotion communication." Philosophical Transactions of the Royal Society B: Biological Sciences 364.1535: 3575-3584.
34. C. J. Stevens, E. Schubert, R. H. Morris, M. Frear,J. Chen, S. Healey, C. Schoknecht, and S. Hansen. 2009. Cognition and the temporal arts: Investigating audience response to dance using PDAs that record continuous data during live performance. International Journal of Human-Computer Studies, 67(9):800 – 813.
35. Yuta Sugiura, Calista Lee, Masayasu Ogata, Anusha Withana, Yasutoshi Makino, Daisuke Sakamoto, Masahiko Inami, and Takeo Igarashi. 2012. PINOKY: a ring that animates your plush toys. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12). ACM, New York, NY, USA, 725-734. DOI=10.1145/2207676.2207780
36. Patel, Shyamal, et al. 2012. "A review of wearable sensors and systems with application in rehabilitation." J Neuroeng Rehabil 9.12: 1-17.
37. Cacioppo, John T., Louis G. Tassinary, and Gary Berntson, eds. Handbook of psychophysiology. Cambridge University Press, 2007.
38. S.M. VaezMousavi, R. J. Barry, J. A. Rushby, and A. R. Clarke. 2007. "Arousal and activation effects on physiological and behavioral responding during a continuous performance task," Acta Neurobiol. Exp. (Wars), vol. 67, pp. 461–470.
39. Tschacher, Wolfgang, Steven Greenwood, Volker Kirchberg, Stéphanie Wintzerith, Karen van den Berg, and Martin Tröndle. 2012. 'Physiological correlates of aesthetic perception of artworks in a museum'. Psychology of Aesthetics, Creativity and the Arts 6:1. 96-103.
40. Rahman, Md Mahbubur, et al. 2014. "Are we there yet?: Feasibility of continuous stress assessment via wireless physiological sensors." Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics. ACM.
41. Woehrle, Matthias, Martin Bor, and Koen Langendoen. 2012. "868 MHz: a noiseless environment, but no free lunch for protocol design." Networked Sensing Systems (INSS), Ninth International Conference on. IEEE.
42. White, Mike. Arts development in community health: A social tonic. Radcliffe Publishing, 2009.
43. Chen Wang, Erik N. Geelhoed, Phil P. Stenton, and Pablo Cesar. 2014. Sensing a live audience. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14). ACM, New York, NY, USA, 1909-1912. DOI=http://dx.doi.org/10.1145/2556288.2557154
44. C. Wang, J. Wong, X. Zhu, T. Roggla, J. Jansen, and Pablo Cesar, "Quantifying Audience Experience in the Wild: Heuristics for Developing and Deploying a Biosensor Infrastructure in Theaters," in Proceedings of the *International Workshop on Quality of Multimedia Experience*, (QoMEX2016), Lisbon, Portugal, June 6-8, 2016.
45. Poh, M. Z., Kim, K., Goessling, A. D., Swenson, N. C., & Picard, R. W. 2009. Heartphones: Sensor earphones and mobile application for non-obtrusive health monitoring. In 2009 International Symposium on Wearable Computers (pp. 153-154). IEEE.